

Article

Estimation of Blast-Induced Peak Particle Velocity through the Improved Weighted Random Forest Technique

Biao He ¹, Sai Hin Lai ^{1,*}, Ahmed Salih Mohammed ², Mohanad Muayad Sabri Sabri ^{3,*}
and Dmitrii Vladimirovich Ulrikh ⁴

¹ Department of Civil Engineering, Faculty of Engineering, Universiti Malaya, Kuala Lumpur 50603, Malaysia; s2005282@siswa.um.edu.my

² Civil Engineering Department, College of Engineering, University of Sulaimani, Sulaymaniyah 46001, Iraq; ahmed.mohammed@univsul.edu.iq

³ Peter the Great St. Petersburg Polytechnic University, St. Petersburg 195251, Russia

⁴ Department of Urban Planning, Engineering Networks and Systems, Institute of Architecture and Construction, South Ural State University, 76, Lenin Prospect, Chelyabinsk 454080, Russia; ulrikhdv@susu.ru

* Correspondence: laish@um.edu.my (S.H.L.); mohanad.m.sabri@gmail.com (M.M.S.S.)

Abstract: Blasting is one of the primary aspects of the mining operations, and its environmental effects interfere with the safety of lives and property. Therefore, it is essential to accurately estimate the environmental impact of blasting, i.e., peak particle velocity (PPV). In this study, a regular random forest (RF) model was developed using 102 blasting samples that were collected from an open granite mine. The model inputs included six parameters, while the output is PPV. Then, to improve the performance of the regular RF model, five techniques, i.e., refined weights based on the accuracy of decision trees and the optimization of three metaheuristic algorithms, were proposed to enhance the predictive capability of the regular RF model. The results showed that all refined weighted RF models have better performance than the regular RF model. In particular, the refined weighted RF model using the whale optimization algorithm (WOA) showed the best performance. Moreover, the sensitivity analysis results revealed that the powder factor (PF) has the most significant impact on the prediction of the PPV in this project case, which means that the magnitude of the PPV can be managed by controlling the size of the PF.

Keywords: blasting; ground vibration; PPV prediction; random forest; whale optimization algorithm



Citation: He, B.; Lai, S.H.; Mohammed, A.S.; Sabri, M.M.S.; Ulrikh, D.V. Estimation of Blast-Induced Peak Particle Velocity through the Improved Weighted Random Forest Technique. *Appl. Sci.* **2022**, *12*, 5019. <https://doi.org/10.3390/app12105019>

Academic Editor: Ricardo Castedo

Received: 12 April 2022

Accepted: 11 May 2022

Published: 16 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Blasting is an economical method of rock excavation in mining and civil engineering, and produces a series of adverse environmental effects such as blasting vibration [1–3], flying rocks [4–6], and back break [7–9]. Among these adverse effects, the harm caused by blasting vibration is quite serious. For example, the surrounding structures can be damaged or fail because of excessive structural vibration produced by ground vibration during blasting [10,11]. Therefore, it is indispensable to predict the magnitude of blast vibration accurately.

The standard base parameter for assessing the magnitude of the blast-induced ground vibration is the peak particle velocity (PPV) [12]. Many studies on PPV prediction have been implemented, such as the proposed empirical formulas, multiple linear and nonlinear regression methods, and machine learning (ML) methods. Among these methods, the empirical formulas for the PPV prediction are easy to construct, but these formulas do not have enough accuracy since few factors are considered to implement the prediction task [13,14]. Moreover, some empirical formulas are usually designed for a given location with distinct geological parameters and field morphology, indicating that these formulas are limited and unable to predict the PPV at other blasting locations [15] accurately. Concerning the multiple linear and nonlinear regression methods, several scholars have demonstrated

that these methods are capable of handling high-dimensional problems, which means that the effects of multiple factors can be considered simultaneously when predicting PPV [15,16]. However, many studies prove that the accuracy of statistical models is inferior compared to the ML methods [17–20].

In recent years, a large number of ML techniques have been applied in different areas of geotechnics, such as tunnel construction and risk assessment [21–28], soil classification [29], pile technology [30,31], materials properties [32–35], slope stability [36,37], blasting environmental issues [38,39], pillar stability prediction [40], and rock material properties [41–44], which reveal the favorable application prospects of the ML techniques. Similarly, as for the research on PPV prediction, to compensate for the shortcomings of empirical formulas and statistical models, there has been a strong inclination toward using the ML techniques in predicting the PPV because of their feasible ability to handle multidimensional nonlinear problems. For instance, Zhang et al. [1] used five machine learning techniques (i.e., classification and regression trees (CART), chi-squared automatic interaction detection (CHAID), random forest (RF), artificial neural network (ANN), and support vector machine (SVM) to predict the PPV caused by mine blasting. Their research utilized five parameters, including maximum charge per delay (MC), stemming (ST), distance from the measuring station to the blast face (DI), powder factor (PF), and hole depth (HD), to develop the ML models. The results showed that the RF had a superior capability in predicting PPV compared with the other four techniques. Lawal [45] developed an ANN-based formula to predict the PPV using two inputs, i.e., the distance from the monitoring point to the blast face and the explosive charge per delay. This study attempts to construct an interpretable formula through the ANN to identify the effect of inputs on the PPV. Rana et al. [46] compared the performance of two AI techniques (e.g., ANN and decision tree (DT)) for forecasting the blast-induced PPV. In the study, eight input parameters, i.e., total charge, number of holes, hole diameter, distance from blasting face, hole depth, tunnel cross section, the maximum charge per delay, and charge per hole, were used to design the ML models. The results indicated that the precision of the DT model performed better than the ANN model. Some relevant studies for predicting the PPV are shown in Table 1.

Table 1. Research on blast-induced PPV prediction.

Techniques	Input Variables	Number of Samples	Studies
RF, CART, CHAID, SVM, ANN	MC, HD, ST, PF, DI	102	Zhang et al. [1]
ANN	DI, MC	100	Lawal, A.I. [45]
GEP, ANFIS, SCA-ANN	DI, MC, ρ , SRH	100	Lawal, A.I. et al. [47]
MPSO-ANN	DI, MC	137	BUI Xuan-Nam et al. [48]
ANN, DT	TC, A, MC, NH, H, DI, HD, CPH	137	Rana et al. [46]
ANFIS	DI, MC	44	Iphar et al. [49]
ANN	HD, ST, DI, MC	182	Monjezi et al. [50]
FIS	B, S, ST, N, MC, DI	120	Ghasemi et al. [51]
ICA-ANN	BS, ST, MC, DI, V_p , E	95	Hajmassani et al. [52]
ICA	MC, DI	73	Jahed Armaghani et al. [53]
ANN, FIS	MC, DI	162	Mohamed [54]

Note: MC—maximum charge per delay; HD—hole depth; ST—stemming; PF—powder factor; DI—distance from the measuring station to the blast face; ρ —rock density; SRH—Schmidt rebound hardness; TC—total charge; A—tunnel cross section area; NH—number of holes; H—hole diameter; CPH—charge per hole; B—burden; S—spacing; N—number of raw; BS—burden to spacing; E—Young modulus; V_p —p-wave velocity; CART—classification and regression trees; CHAID—chi-squared automatic interaction detection; RF—random forest; ANN—artificial neural network; SVM—support vector machine; MPSO—modified particle swarm optimization; GEP—gene expression programming; ANFIS—adaptive neuro-fuzzy inference system; SCA—sine cosine algorithm; DT—decision tree; ICA—imperialist competitive algorithm; FIS—fuzzy logic.

In light of the above, the good performance and adaptability of the ML models in predicting the PPV have been proven gradually, and hence a large number of studies of ML models for predicting the PPV emerged. Furthermore, the ML models with better

adaptability and higher prediction accuracy used for the prediction of blast-induced PPV need to be further developed and improved.

In the present study, we utilized a classic ML model termed the RF to conduct the PPV prediction task. After designing the regular RF model using the datasets collected from a quarry mine, we proposed five categories for modifying the regular RF model's weights and, as a result, making better predictions. The designed weighting frameworks include the improved weighted RF model based on the prediction accuracy of decision trees, as well as the optimized weights of decision trees obtained by three metaheuristic algorithms, i.e., whale optimization algorithm (WOA), gray wolf optimization (GWO), and tunicate swarm algorithm (TSA). Subsequently, four evaluation metrics were used to validate the performance of the developed models. Finally, a sensitivity analysis was conducted to identify the predominant factors for PPV prediction in this engineering case.

2. Project Description and Data Collection

In the present study, a granite mine in Penang state, Malaysia, was selected as the subject to research PPV prediction, and thus its blasting operations were investigated. Granite is the most common rock type found in the study area. The top layer is generally less than three feet thick and consists mostly of sandy clay with humus and tree roots. Explosive operations are rather common in this mine, and these operations are repeated at various intervals. Blasting at this location aims to create aggregates for various building projects with annual capacity ranging from 500–700 thousand tons per year, and large quantities of explosives are used in explosive operations. For example, in holes with sizes ranging from 76 mm to 89 mm, explosives weighing 856 kg to 9420 kg are commonly utilized. Concerning the present study, 102 blasting operations were recorded, including the design details of the blast parameters and the PPV values obtained by the Vibra ZEB seismograph. The measured and recorded blast parameters involve the number of holes, hole diameter, hole depth, burden, spacing, stemming length, subdrilling, total charge, powder factor, the maximum charge per delay, and distance from the blast-face to the measuring points. Among them, six parameters were selected as the inputs that will be used to implement the ML modeling, which is in accord with previously published works [1,48,49]. The simple statistics of the six parameters are shown in Table 2, including their max/min values, mean values, and standard deviations. It can be seen from Table 2 that the distance from the seismograph to the explosion site was about 285 m–531 m, and the range of the PPV is between 0.13 and 11.05 mm/s.

Table 2. Statistic of the data collected from the study area.

Parameters	Symbol	Unit	Type	Max	Min	Mean	Std. Dev.
Burden to spacing	BS	-	Input	0.92	0.7	0.819	0.004
Hole depth	HD	m	Input	23.17	5.23	14.115	15.973
Stemming	ST	m	Input	3.6	1.9	2.630	0.157
Powder factor	PF	kg/m ³	Input	0.94	0.23	0.654	0.034
Max charge per delay	MC	kg	Input	305.6	45.8	179.623	4246.587
Distance	DI	m	Input	531	285	379.520	5100.269
Peak particle velocity	PPV	mm/s	Output	11.05	0.13	5.337	9.267

3. Methods

3.1. RF

The RF approach is a machine learning model proposed by Leo Breiman in 2001 [55]. Because of its special algorithmic mechanism and efficient performance, the RF model has attracted much attention from researchers in various research fields. The RF model is an approach that integrates many base learners (also called decision trees) through integrated learning. It uses bagging and bootstrap techniques to train the decision trees, which solves the problem of insufficient performance of individual decision trees when dealing with

complex data. At the same time, the RF is also a nonparametric classification or regression method, so it does not require prior knowledge when processing data.

The schematic diagram of the design of the RF and improved RF models is depicted in Figure 1. In the process of building decision trees, the RF model generates more randomness as the number of decision trees in the forest increases. Instead of selecting the optimal value step by step like a decision tree, the RF uses its ability of random selection and the voting mechanism of the decision tree to find the optimal value quickly. This property allows the RF model to have better classification or regression performance and a strong generalization ability of the learning system. For the regression task, when predicting the unknown output value, every single decision tree yields a predicted value, and the final value is the average of all decision trees, as shown in Equation (1). In this process, each decision tree occupies the same weight:

$$y_{pre} = \sum_{i=1}^t weight_i \times Tree_i(x) \quad (1)$$

where x is the input variable, y_{pre} denotes the predicted value corresponding to the input x , t is the amount of the constructed decision trees, $\{Tree_1, Tree_2, \dots, Tree_t\}$ represent the set of decision trees in a forest, and $weight_i$ is the weight of each decision tree. In the current case (i.e., for the regular RF model), the value of weight is obtained as follows:

$$weight_i = \frac{1}{t} \quad (2)$$

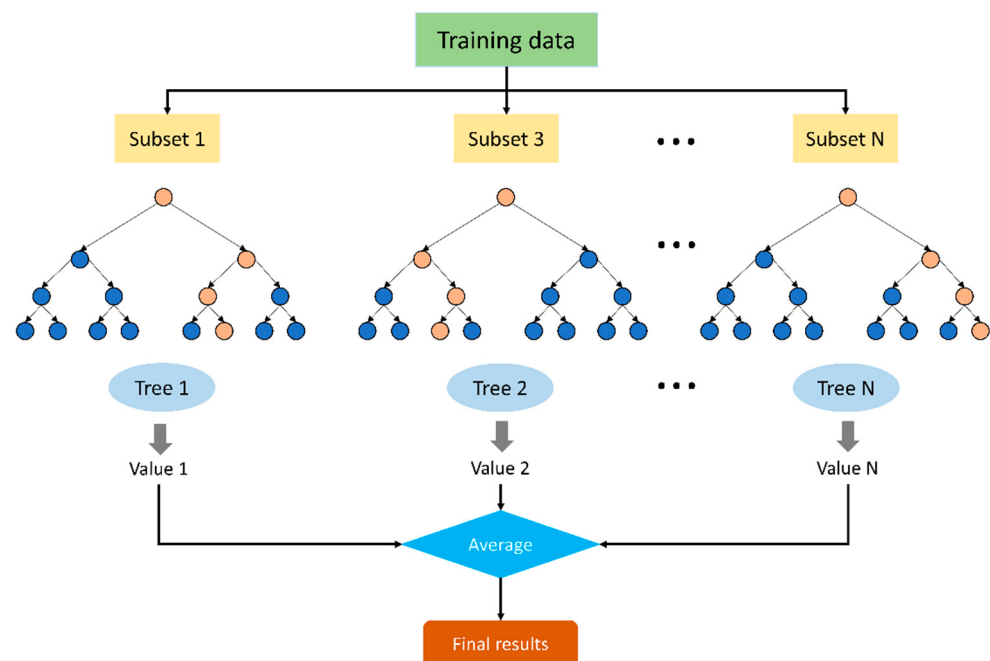


Figure 1. Schematic diagram of the RF model.

In light of the above, it can be inferred that the number of decision trees is a crucial hyperparameter that governs the predictive capability of the RF model. Therefore, a key task in this paper is to determine the optimal number of decision trees. Moreover, to avoid possible overfitting of the RF model, another hyperparameter termed the maximum depth of decision trees must also be tuned. These two critical steps will be discussed in more detail in a later section. In addition, the datasets used for constructing the RF model are randomly split into two parts; that is, 80% of datasets are used for training the RF model, and 20% of datasets are used for validating the performance of the built models. Simultaneously, a fourfold cross-validation procedure is applied in this work when implementing model

training procedures. Note that fourfold cross-validation means that the datasets are divided into four equal parts, one of which is used to test the model, and the rest are used to train the model in each modeling session. A total of four modeling sessions are performed, and the average of the four modeling sessions is used as the final result.

3.2. GWO

The gray wolf optimization (GWO) algorithm is a new swarm intelligence optimization algorithm proposed by Mirjalili et al. in 2014, which is based on the simulation of the hierarchical mechanism and predatory behavior of the gray wolf population in nature. Moreover, the optimization of the GWO algorithm is achieved through the process of wolf stalking, encircling, chasing, and attacking prey [56]. Gray wolves are top carnivores, and their lifestyles are mostly gregarious, thus constituting a hierarchical pyramid in the gray wolf population, with a strict hierarchical management system, as shown in Figure 2.

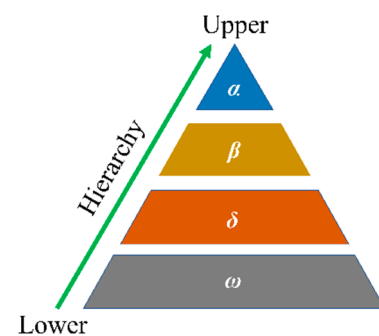


Figure 2. Hierarchy of wolf in GWO algorithm.

The first level of the pyramid is the head of the population, called α , which is mainly responsible for all the decision-making matters of the population. The second pyramid level is called β , which assists α in making management decisions. The third level of the pyramid is δ , which is mainly responsible for scouting, sentry, hunting, and guarding. Furthermore, the bottom level of the pyramid is called ω , which is mainly responsible for coordinating the relationship within the population. The hierarchy of gray wolves plays a crucial role in achieving prey hunting. The predation process is led by α . At first, the wolves search, track, and approach the prey in a team pattern; then, the wolves encircle the prey from all directions, and when the encirclement is small enough and perfect, the wolves will attack from β and δ , which are closest to the prey, under the command of α . The mathematical models of this process are explained below.

First, the encirclement of the prey by the gray wolves during predation can be characterized by the following equation:

$$D = |C \times X_p(t) - X(t)| \quad (3)$$

where $X_p(t)$ and $X(t)$ denote the position of the prey and the position of the wolves during the t -th iteration, respectively, and C is the coefficient, which is computed by the following equation:

$$C = 2r_1 \quad (4)$$

Here r_1 is the random value in the interval $[0, 1]$.

Then, the equation for updating the position of the gray wolf in search space is as follows:

$$X(t+1) = X(t) - A \times D \quad (5)$$

where A is the convergence factor. A can be computed by the following equation:

$$A = 2a \times r_2 - a \quad (6)$$

$$a = \left(2 - 2 \times \left(\frac{t}{t_{max}} \right) \right) \quad (7)$$

where r_2 is the random value in the interval $[0, 1]$, and t_{max} denotes the maximum iterations.

After that, when the gray wolf determines the position of the prey, the wolf α will lead β and δ to initiate the pursuit behavior. In the wolf pack, α , β , and δ are the closest to the prey. The positions of these three wolves can be used to determine the location of the prey. The mathematical description is as follows:

$$D_\alpha = |C_1 \times X_\alpha(t) - X(t)| \quad (8)$$

$$D_\beta = |C_2 \times X_\beta(t) - X(t)| \quad (9)$$

$$D_\delta = |C_3 \times X_\delta(t) - X(t)| \quad (10)$$

$$X_1 = X_\alpha - A_1 \times D_\alpha \quad (11)$$

$$X_2 = X_\beta - A_2 \times D_\beta \quad (12)$$

$$X_3 = X_\delta - A_3 \times D_\delta \quad (13)$$

$$X_p(t+1) = \frac{X_1 + X_2 + X_3}{3} \quad (14)$$

Using Equations (8)–(13) can obtain the distance between the prey and the α , β , and δ , and then using Equation (14) identifies the direction of movement of individual gray wolves toward the prey.

3.3. WOA

The whale optimization algorithm (WOA) is a novel swarm intelligence optimization algorithm proposed by Mirjalili and Lewis in 2016 [57]. The WOA is based on the simulation of the hunting behavior of humpback whales in nature, and it optimizes the searching process by mimicking the behavior of searching, encircling, pursuing, and attacking the prey by the humpback whales. Whales are considered the largest mammals globally, with adults growing up to 30 m long and weighing 180 tons. Whales have a unique feeding behavior, which is the bubble-net method, as shown in Figure 3. The method is divided into two stages: upward spiral and double circulation. Based on this special foraging behavior, the WOA algorithm is obtained.

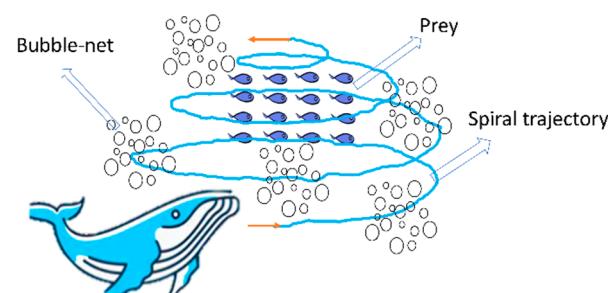


Figure 3. Bubble-net attacking method of whales.

In the WOA, it is assumed that the number of populations is N and the dimension of the search space is d , and the position of the i -th whale in the dimensional space can be expressed as $X_i = (X_1^d, X_2^d, X_3^d, \dots, X_N^d)$. The position of the prey in the search space corresponds to the optimal global solution. Whales can identify the position of the prey and then encircle it since there is no a priori knowledge of the position of the optimal global solution in the search space before solving the optimization problem. Therefore, in the WOA, assuming that the whale at the optimal position in the current population is

the prey and all other whale individuals in the population encircle the optimal whale, the mathematical model of the process is as follows:

$$X(t+1) = X_p(t) - A \times |C \times X_p(t) - X(t)| \quad (15)$$

where t is the current iteration, $X(t)$ denotes the position of whales, $X_p(t)$ denotes the position of prey, and A and C are the coefficients that can be defined as follows:

$$A = 2a \times rand_1 - a \quad (16)$$

$$C = 2 \times rand_2 \quad (17)$$

Here $rand_1$ and $rand_2$ represent the random values in the interval $[0, 1]$, and a is the convergence factor that decreases linearly from 2 to 0 as the number of iterations increases.

To describe the bubble-net attacking behavior of whales with a mathematical model, two different methods were designed in the WOA, namely, the shrinking encircling mechanism and the spiral updating position. Among them, the shrinking encircling mechanism is implemented by the linear reduction in the convergence factor a in Equation (16). In the method of spiral updating position, the spiral motion of the whale is simulated to capture the prey, and its mathematical model is shown as follows:

$$X(t+1) = D' \times e^{bl} \times \cos(2\pi l) + X_p(t) \quad (18)$$

where $D' = |X_p(t) - X(t)|$ denotes the distance between the prey and the i -th whale, b signifies a constant for controlling the shape of the logarithmic spiral, and l is a random value in the interval $[-1, 1]$. Noted that, the whales move within a constricted circle and simultaneously follow a spiral path towards their prey.

In addition to the bubble-net attacking method, whales will also randomly search for prey. Individual whales search randomly according to each other's position, and the mathematical model can be expressed as follows:

$$X(t+1) = X_{rand}(t) - A \times |C \times X_{rand}(t) - X(t)| \quad (19)$$

where X_{rand} represents a position vector of an individual search agent randomly chosen from the current population.

3.4. TSA

The tunicates are small, net-like organisms that are found throughout the sea. They live solitary or parasitic lives and can locate food in the ocean. Based on the biological inspiration of the predatory behavior of tunicates, Kaur et al. proposed a swarm intelligence optimization algorithm termed tunicate swarm algorithm (TSA), which is inspired by the jet-propelled migration mechanism and the intelligent foraging behavior of tunicates in the ocean [58]. The mathematical models of this algorithm are explained below.

(1) Initialization

Similar to most metaheuristic algorithms, the TSA starts to execute the optimization process by initializing the tunicate population, i.e., by initializing the positions of every tunicate ($\vec{A}_0$) in the search space, as shown in the following equation:

$$\vec{A}_0 = \vec{A}_{min} + rand \left(\vec{A}_{max} - \vec{A}_{min} \right) \quad (20)$$

where $\vec{A}_{min}$ and $\vec{A}_{max}$ denote the upper and lower limits of the search space, respectively, and $rand()$ signifies the random value in the interval $[0, 1]$.

(2) Avoid conflicts between search agents

To avoid conflicts between the individuals when implementing the searching task, the TSA utilizes vector $\vec{A}$ to calculate the position of new search agents, which can be illustrated by the following equations:

$$\vec{A} = \frac{\vec{G}}{\vec{M}} \quad (21)$$

$$\vec{G} = c_1 + c_2 - \vec{F} \quad (22)$$

$$\vec{M} = [P_{min} + c_3 \times P_{max} - P_{min}] \quad (23)$$

where $\vec{A}$ denotes the new positions of the search agents, $\vec{G}$ denotes the gravity, $\vec{M}$ denotes the interaction forces between the tunicates, $\vec{F}$ signifies the current advection in the deep sea, and $\vec{F} = 2c_3$, c_1 , c_2 , and c_3 are the random value in the interval $[0, 1]$.

(3) Move to the best neighbor

After avoiding conflicts between the individuals, the search agents will move towards the position of the best neighbor, which the following equation can interpret:

$$\vec{P}_d = \left| \vec{F}_s - r \times \vec{P}_p(t) \right| \quad (24)$$

where $\vec{P}_d$ denotes the distance between the food and the search agents, $\vec{F}_s$ represents the position of the food, $\vec{P}_p(t)$ signifies the positions of the search agents during the t -th iteration, and r is the random value in the interval $[0, 1]$.

(4) Move towards the best individual

The mathematical description of the movement of the tunicate population towards the position of the optimal search agent is as follows:

$$\vec{P}_p(t^*) = \begin{cases} \vec{F}_s + \vec{A} \times \vec{P}_d, & r \geq 0.5 \\ \vec{F}_s - \vec{A} \times \vec{P}_d, & r < 0.5 \end{cases} \quad (25)$$

where $\vec{P}_p(t^*)$ is the position of the search agent closest to the target food.

(5) Swarm behavior

TSA is implemented to mimic the tunicates' swarm behavior by saving the first two optimal solutions and updating the other tunicates' positions according to the first two optimal solutions. The swarm behavior of the tunicates is mathematically described as follows:

$$\vec{P}_p(t+1) = \frac{\vec{P}_p(t) + \vec{P}_p(t+1)}{2 + c_1} \quad (26)$$

where $\vec{P}_p(t+1)$ denotes the positions of the search agent when implementing the $(t+1)$ -th iteration.

3.5. Improved RF Models

Although the RF model has good performance and has been applied in many areas, there is still some room for improvement. For the random forest approach, in addition to sampling with the replacement for constructing training datasets, a random number of features are picked each time to lower the degree of correlation between individual produced decision trees. The end outcome is a simple average computed by each created decision tree of the forest [55]. Although the RF shows remarkable performance in some

regression or classification tasks, it seems that some improvements for the combined way of the base learners (i.e., decision trees) can be made to achieve a better prediction of the RF. Decision trees exhibit different predictive capabilities because of the bootstrap replicates and random feature picking. However, in the regular RF, each decision tree is assigned with the same weights, which seems unreasonable and can be further improved. Therefore, some techniques such as weighting the decision trees based on their predictive capabilities are proposed to optimize the regular RF. The accuracy of decision trees is considered the indicator that can amend the weights of decision trees instead of using the same weight for each tree [59]. In this paper, two indicators, i.e., the coefficient of determination (R^2) and the root mean squared error (RMSE) of decision trees, are used as benchmarks for improving the weights of decision trees. For example, the first category is that decision trees with higher R^2 will be assigned relatively higher weights, and vice versa. Another case is that decision trees with lower RMSE will be assigned relatively higher weights, and vice versa. Furthermore, we also put forward using three metaheuristic algorithms (i.e., GWO, WOA, and TSA) to search for the best weights of decision trees and then compare the effect of their improvements on the regular RF model.

To sum up, there are a total of five methods that are leveraged to refine the weights of decision trees in a forest. Before applying these methods, the first task that needs to be carried out is the parameter tuning of the RF model, which is to establish the optimal RF model based on the training dataset. Subsequently, the task is to use the five categories mentioned above to amend the weights of decision trees. This part will be elaborated hereinafter. Moreover, the schematic diagram of the design of the RF and improved RF models is depicted in Figure 4.

3.5.1. Improved Weights Based on the Accuracy

First, let $\{Tree_1(x), Tree_2(x), Tree_3(x), \dots, Tree_t(x)\}$ represent the set of decision trees. For the regular RF model, the final predicted values can be obtained through Equation (27), which indicates that the weight of every single decision tree is equal to $1/t$.

$$y_{pre} = \sum_{i=1}^t \frac{1}{t} \times Tree_i(x) \quad (27)$$

where t is the number of decision trees in a forest.

For the improved weights based on R^2 , the weight of every single decision is obtained through Equation (28), and the final predicted values can be obtained using Equation (29). The key step of this approach is to compute the weights of decision trees. After constructing the optimal RF model based on the training sets, the accuracy (i.e., R^2) of each decision tree can be obtained. Then, according to Equation (28), all decision trees will be assigned a normalized weight, and the larger the term R^2 , the larger the weight assigned, which means that these decision trees with large weights will have a greater role in the final prediction. Equations (28) and (29) are:

$$\text{weight_}R^2_i = \frac{R^2_i}{\sum_{i=1}^t R^2_i} \quad (28)$$

$$y_{pre} = \sum_{i=1}^t \text{weight_}R^2_i \times Tree_i(x) \quad (29)$$

Similarly, for the improved weights based on the RMSE, the weight of every single decision is obtained through Equation (30), and the final predicted values can be obtained using Equation (31). The weights of the decision trees are obtained by first taking the inverse of the RMSE and then performing a normalized calculation, which means that the

larger the RMSE of a decision tree, the smaller the weight it is assigned to, and vice versa. Equations (30) and (31) are:

$$\text{weight_RMSE}_i = \frac{\frac{1}{\text{RMSE}_i}}{\sum_{i=1}^t \left(\frac{1}{\text{RMSE}_i} \right)} \quad (30)$$

$$y_{pre} = \sum_{i=1}^t \text{weight_RMSE}_i \times \text{Tree}_i(x) \quad (31)$$

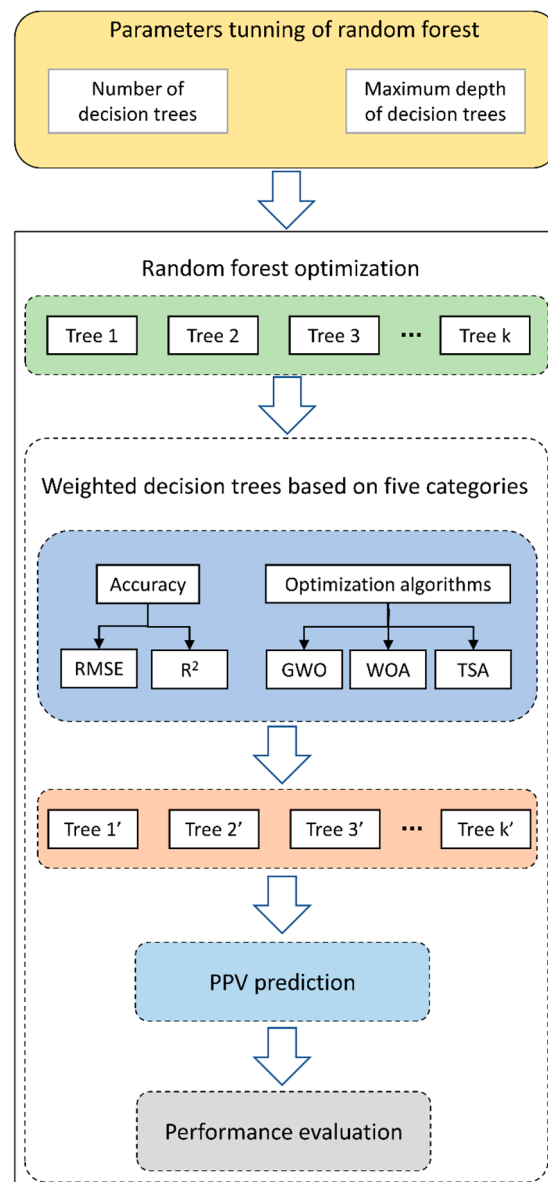


Figure 4. Flowchart of the design of the RF and improved RF models.

3.5.2. Improved Weights Based on Three Metaheuristic Algorithms

In this section, we consider the weights of decision trees in a forest as an unknown space of high dimensionality to be solved. After determining the number of trees in a forest, the dimension of the space to be solved is equal to the number of trees. Then, the GWO, WOA, and TSA algorithms are used to capture the optimal weights in the unknown space. To achieve this goal, it is assumed that the independent variables to be solved

are set to $\{x_1, x_2, x_3, \dots, x_t\}$, where t is the number of trees and the range of x is in the interval $[0, 1]$, and then the standardized weights are computed as follows:

$$\text{weight_optimization}_i = \frac{x_i}{\sum_{i=1}^t x_i} \quad (32)$$

where $\text{weight_optimization}_i$ represents the weights of decision trees optimized by the metaheuristic algorithms, and it is clear that $\sum_{i=1}^t \text{weight_optimization}_i = 1$.

After that, the RMSE is considered the criteria used to evaluate each result's performance. In other words, the fitness function that needs to be solved by the metaheuristic algorithms is the RMSE function. Moreover, the L2 regularization (also named penalty terms) is also incorporated in the RMSE function, which avoids overfitting. Therefore, the fitness function consists of two parts: the RMSE function and the L2 regularization, which is presented by Equation (34):

$$y_{pre} = \sum_{i=1}^t \text{weight_optimization}_i \times \text{Tree}_i(x) \quad (33)$$

$$\text{Fitness function} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - y_{pre})^2} + \gamma \times \sqrt{\sum_{i=1}^t \text{weight_optimization}_i^2} \quad (34)$$

where y_{pre} denotes the predicted PPV computed by the improved RF model based on optimization algorithms, m is the number of training samples, and γ is the coefficient of the L2 regularization, whose value is set to 0.08 in this paper according to the trial-and-error method.

For the GWO, WOA, and TSA, the parameters that need to be set are the swarm sizes and the number of iterations. The appropriate selection of these parameters can effectively and quickly lead to optimal results. Therefore, after constructing the model several times, the swarm sizes set to each optimization algorithm are 50, 100, 150, and 200, respectively, and the number of iterations is set at 1000.

The GWO, WOA, and TSA optimization techniques can be used to improve the RF model in the following way:

- (1) Data preparation: randomly divide the raw data into a training set (80% of raw data) and a testing set (20% of raw data);
- (2) Initialization: Initialize the swarm size, iterations, as well as some necessary parameters of the three optimization algorithms;
- (3) Fitness evaluation: Calculate the fitness value of the population, evaluate its fitness, and then save the best fitness value before starting the next iteration;
- (4) Update parameters: Update the fitness value based on the outcome of each iteration, which aims to capture the ideal solutions;
- (5) Suspension conditions check: When the optimal fitness value no longer changes, or the maximum number of iterations is reached, the optimal solutions of the weights of decision trees are obtained.

3.6. Criteria for Evaluation

An important task to be conducted after the modeling is to evaluate each model's accuracy and generalization ability. As mentioned previously, in this work, 80% of the data samples were assigned as training that was used to train the RF and improve the RF models, while the rest 20% of the data samples were assigned as testing data to verify the performance of the RF and improved RF models. To evaluate the performance of the built models, four metrics were used in this work, i.e., the coefficient of determination (R^2), the root mean squared error (RMSE), the mean absolute error (MAE), and the variance account for (VAF). These evaluation indicators are defined below.

The square of the correlation between the predicted and measured values is represented by R^2 . The RMSE characterizes the standard deviation of the fitting error between the predicted values and the measured values. The MAE indicates the mean absolute error between the predicted values and measured values. The VAF describes the prediction performance by comparing the standard deviation of the fitting error with the standard deviation of the actual value. The following equations were utilized to calculate the R^2 , RMSE, MAE, and VAF values:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (35)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (36)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (37)$$

$$\text{VAF} = \left(1 - \frac{\text{var}(y_i - \hat{y}_i)}{\text{var}(y_i)} \right) \times 100 \quad (38)$$

where y_i , $\hat{y}_i$, and $\bar{y}$ denote the measured, predicted, and mean values of the PPV, respectively. When the predicted and measured PPV values are precisely the same, R^2 is 1, the RMSE is 0, the MAE is 0, and the VAF is 100 (%).

4. Results and Discussion

4.1. Parameter Tuning of the RF Model

The main purpose of this section is to determine the optimal hyperparameters of the RF model, including the number of decision trees and the maximum depth of decision trees. In addition, before conducting the model training, the original dataset was standardized to eliminate the negative effect of magnitude between the data and speed up the model's training. The formula for standardization of data is given as follows:

$$X_i^* = \frac{X_i - \mu_i}{\sigma_i}, \quad i = 1, 2, \dots, 6 \quad (39)$$

where X_i denotes the data samples belonging to feature i , μ_i and σ_i denote the mean and standard deviation of X_i , respectively, and X_i^* signifies the standardized datasets that are prepared for constructing the RF model.

The parameters tuning of RF consist of two steps. First, the scale of the RF is optimized to determine the number of decision trees in a forest. In this regard, the number of trees increases from 1 to 200 with 1 increment each time. Meanwhile, other parameters are set to default values. Figure 5 shows the performance of the RF model with respect to the number of trees. Intuitively, it can be seen that the mean squared error decreases when the number of trees increases from 1 to 50, and then with the increase in the number of trees, the mean squared error only shows small fluctuations. The results show that the minimum mean squared error is reached when the number of trees is 51. Consequently, the scale of the RF model is set to 51 decision trees. Furthermore, using the testing set to validate the performance of the RF model when the number of trees is 51, the results show that the R^2 , RMSE, MAE, and VAF values on the testing set are 0.915, 0.275, 0.224, and 93.504, respectively.

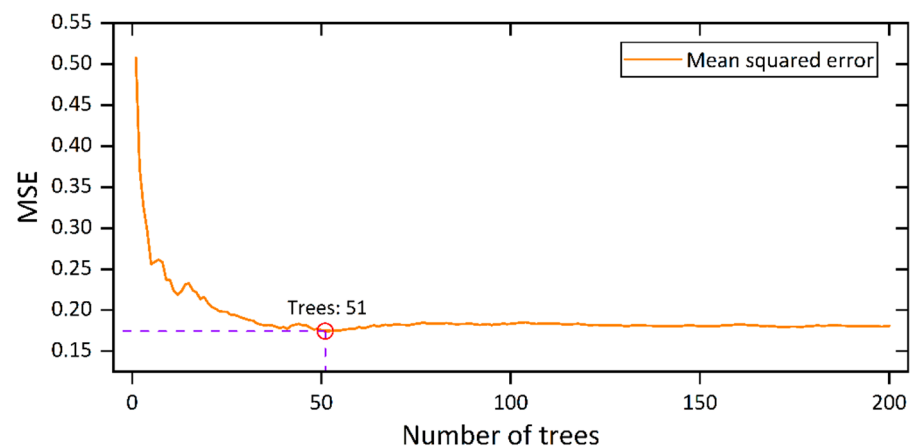


Figure 5. MSE with respect to the number of trees in a forest.

After that, the optimal maximum depth of decision trees was determined. The number of the maximum depth increases from 1 to 10 with 1 increment each time, while the number of decision trees is 51 currently. Figure 6 depicts the performance of the RF model concerning the maximum depth of decision trees. It can be found that the minimum mean squared error is reached when the maximum depth is set to 7. When the maximum depth values are larger than 7, there is no apparent decrease in the mean squared error. At the same time, the performance evaluation results of the current RF model on the testing set are R^2 of 0.923, the RMSE of 0.262, the MAE of 0.209, and the VAF of 93.994, respectively. Thus, according to the aforementioned results, the maximum depth is set to 7. Now, the RF model has been successfully constructed, and the later work uses five techniques to optimize the weights of decision trees, aiming to improve the performance of the current RF model.

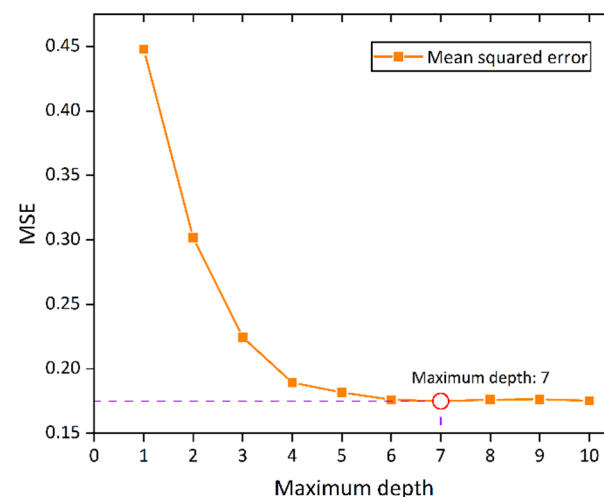


Figure 6. MSE with respect to the maximum depth of trees in a forest.

4.2. Improved RF-Based Models

In this section, five methods were used to optimize the regular RF model, namely the refined weights of decision trees based on their predictive capabilities that are characterized by R^2 and the RMSE, and the refined weights of decision trees based on the optimization solution of three metaheuristic algorithms (i.e., GWO, WOA, and TSA). The main goal of the five techniques is to obtain the optimal weight set of decision trees that can minimize the generalization error of the improved RF models. Moreover, to validate the models' performance, each model's accuracy and error on the training set and testing set are both utilized. In this way, the optimal improved RF model can be effectively obtained.

Figure 7 presents the calculation results of the three metaheuristic algorithms on the training set. It can be seen that there are some differences in the calculation results due to the swarm sizes. For example, for the RF-WOA model, the final fitness value is 200 swarm sizes < 150 swarm sizes < 100 swarm sizes < 50 swarm sizes; for the RF-GWO model, the final fitness value is 100 swarm sizes < 200 swarm sizes < 150 of swarm sizes < 50 swarm sizes; and for the RF-TSA model, the final fitness value is 150 of swarm sizes < 200 swarm sizes < 50 of swarm sizes < 100 swarm sizes. It should be noted that the above results are obtained based on the training set. For a clearer identification of the performance of each model, it should be combined with its performance on the testing set. For this purpose, a scoring evaluation method was employed to select the best model [60]. The principle of this method is that the higher the score, the better the performance, and vice versa. Consequently, the final scoring of the RF-WOA, RF-GWO, and RF-TSA models with different swarm sizes on both training and test sets is obtained in Tables 3–5. According to the results in Tables 3–5, it can be concluded that the best RF-WOA model with a total score of 26 is obtained when the swarm sizes are set to 100, the best RF-GWO model with a total score of 28 is obtained when the swarm sizes are set to 200, and the best RF-TSA model with a total score of 27 is obtained when the swarm sizes are set to 50.

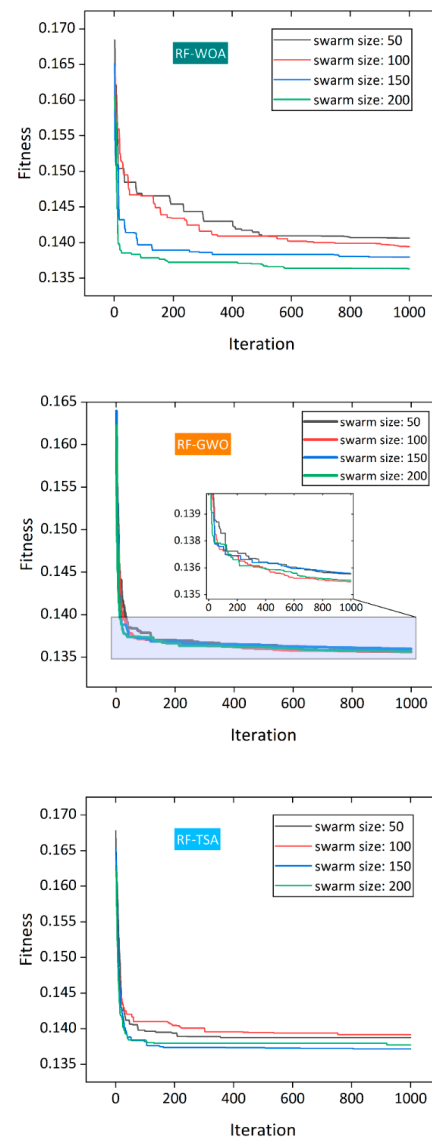


Figure 7. Various RF-based metaheuristic models based on different swarm sizes.

Table 3. Performance of different swarm sizes of the RF-WOA models on predicting the PPV.

Models	Training								Testing								Score Summation	
	Swarm Size	R ²	Score	RMSE	Score	MAE	Score	VAF	Score	R ²	Score	RMSE	Score	MAE	Score	VAF		Score
RF-WOA	50	0.986	4	0.119	1	0.094	2	98.584	1	0.932	4	0.246	4	0.187	4	94.723	3	23
	100	0.986	4	0.118	2	0.093	3	98.603	2	0.932	4	0.246	4	0.188	3	95.032	4	26
	150	0.986	4	0.117	3	0.092	4	98.612	3	0.927	2	0.254	2	0.191	2	94.430	2	22
	200	0.986	4	0.116	4	0.092	4	98.647	4	0.928	3	0.253	3	0.196	1	94.394	1	24

Table 4. Performance of different swarm sizes of the RF-GWO models on predicting the PPV.

Models	Training								Testing								Score Summation	
	Swarm Size	R ²	Score	RMSE	Score	MAE	Score	VAF	Score	R ²	Score	RMSE	Score	MAE	Score	VAF		Score
RF-GWO	50	0.986	4	0.116	4	0.092	4	98.632	1	0.930	4	0.250	4	0.192	3	94.516	3	27
	100	0.986	4	0.116	4	0.092	4	98.633	2	0.929	3	0.251	3	0.192	3	94.475	1	24
	150	0.986	4	0.116	4	0.092	4	98.637	4	0.929	3	0.252	2	0.193	2	94.488	2	25
	200	0.986	4	0.116	4	0.093	3	98.635	3	0.929	3	0.251	3	0.191	4	94.528	4	28

Table 5. Performance of different swarm sizes of the RF-TSA models on predicting the PPV.

Models	Training								Testing								Score Summation	
	Swarm Size	R ²	Score	RMSE	Score	MAE	Score	VAF	Score	R ²	Score	RMSE	Score	MAE	Score	VAF		Score
RF-TSA	50	0.986	4	0.118	2	0.093	3	98.598	2	0.931	4	0.248	4	0.191	4	94.743	4	27
	100	0.986	4	0.118	2	0.094	2	98.582	1	0.924	1	0.260	1	0.201	2	93.823	1	14
	150	0.986	4	0.116	4	0.093	3	98.635	4	0.927	2	0.255	2	0.191	4	94.483	2	25
	200	0.986	4	0.117	3	0.092	4	98.619	3	0.929	3	0.251	3	0.195	3	94.569	3	26

Table 6. Performance of the RF and improved RF models on predicting the PPV.

Models	Training							Testing							Score Summation		
	R ²	Score	RMSE	Score	MAE	Score	VAF	Score	R ²	Score	RMSE	Score	MAE	Score		VAF	Score
RF-WOA (100)	0.986	6	0.118	5	0.093	6	98.603	5	0.932	6	0.246	6	0.188	6	95.032	6	46
RF-GWO (200)	0.986	6	0.116	6	0.093	6	98.635	6	0.929	4	0.251	4	0.191	5	94.528	4	41
RF-TSA (50)	0.986	6	0.118	5	0.093	6	98.598	4	0.931	5	0.248	5	0.191	5	94.743	5	41
RF	0.973	3	0.164	2	0.123	3	97.298	1	0.923	1	0.262	1	0.209	2	93.994	1	14
RF-R ²	0.974	4	0.161	3	0.121	4	97.380	2	0.924	2	0.261	2	0.207	3	94.078	2	22
RF-RMSE	0.975	5	0.156	4	0.119	5	97.530	3	0.925	3	0.258	3	0.203	4	94.219	3	30

After obtaining the best model for individual metaheuristic algorithms in their respective model comparisons, the next step is to perform further comparative analysis of them and the two previously mentioned improved weighting methods based on the accuracy (i.e., R^2 and the RMSE) to identify the optimal improved RF model. Concerning the improved weights based on the R^2 and RMSE of every single decision tree, the weights of the decision trees are computed through Equations (28) and (30), respectively. For the metaheuristic algorithms, the RF-WOA with 100 swarm sizes, the RF-GWO with 200 swarm sizes, and the RF-TSA with 50 swarm sizes are chosen because of their best performance. At the same time, the regular RF model is also used for comparative analysis in this section. Then, the results of the performance of these six models are tabulated in Table 6. As can be seen from Table 6, the model with the highest score is the RF-WOA, whose scoring is 46, followed by the RF-GWO and RF-TSA, whose scorings are both 41. As for the RF-RMSE and RF- R^2 , their scores were 30 and 22, respectively, which indicates the performance of the RF-RMSE is slightly better than that of the RF- R^2 . In general, these five models all outperform the regular RF model with a scoring of 14. Moreover, the ranking of these five models is as follows: RF-WOA > RF-GWO = RF-TSA > RF-RMSE > RF- R^2 > RF.

To intuitively observe the differences among these models, we divide the four evaluation metrics into two groups, one for the VAF and $R^2 \times 100$, where larger values indicate better model performance, and the other one for the RMSE and MAE, where smaller values indicate better models. As depicted in Figure 8, for the group of the VAF and $R^2 \times 100$ (the left one in Figure 8), the accuracies of the RF-GWO, RF-WOA, and RF-TSA are quite close and are all better than the other three models (i.e., RF-RMSE, RF- R^2 , and RF), whereas, for the group of the RMSE and MAE (the right one in Figure 8), the error of the RF-GWO is slightly lower than those of the RF-WOA and RF-TSA. Likewise, the errors of the three metaheuristic-based RFs are all lower than those of the RF-RMSE, RF- R^2 , and RF. Moreover, it can be concluded that all of the refined weights RF models have a good performance on the training set compared with the regular RF model. As for the testing set, the results are shown in Figure 9. For the group of the VAF and $R^2 \times 100$ (the left one in Figure 9), the ranking of the accuracy of these six models is shown as RF-WOA > RF-TSA > RF-GWO > RF-RMSE > RF- R^2 > RF. The same situation occurs in the group of the RMSE and MAE (the right one in Figure 9), which indicates the ranking of error of these six models is RF-WOA < RF-TSA < RF-GWO < RF-RMSE < RF- R^2 < RF. Therefore, it can be inferred that the RF-WOA has a better generalization ability than other models, which can be verified as it has a better performance on the testing set. Additionally, we can also conclude that all of the refined weights RF models have a good performance on the testing set compared with the regular RF model, which means the refined techniques proposed in this paper can effectively enhance the performance of the regular RF model.

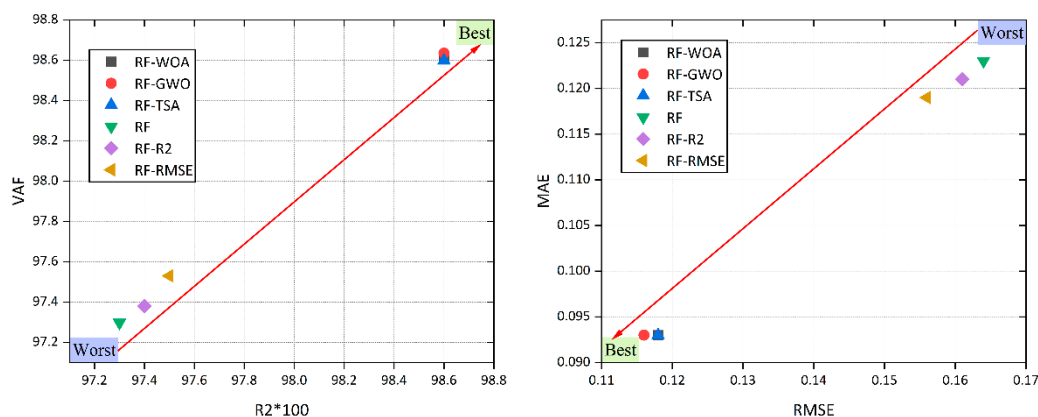


Figure 8. Results of the evaluation metrics of the developed models on the training set.

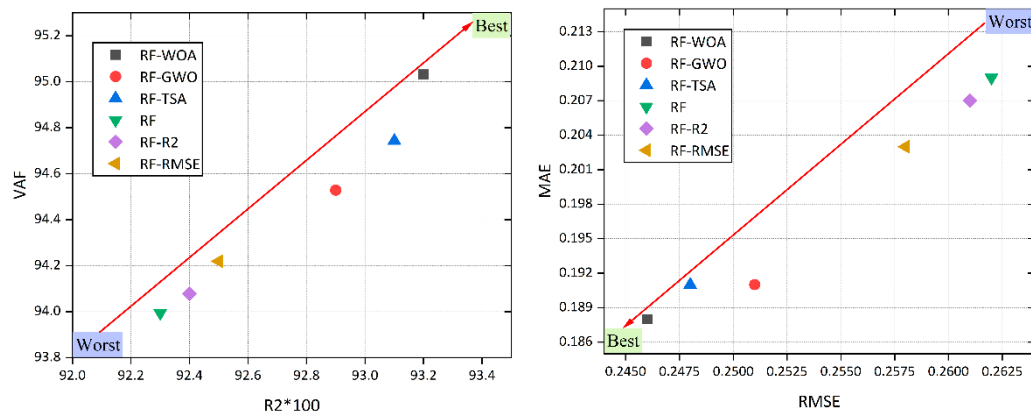


Figure 9. Results of the evaluation metrics of the developed models on the testing set.

In a different way, the Taylor diagram is presented in this work, which may be used to graphically summarize how closely a set of patterns fits observations [61]. The similarities between the patterns and observations are quantified by utilizing their correlation coefficients, centered root-mean-square errors, and standard deviations, as shown in Equation (40) [62]:

$$E'^2 = \sigma_p^2 + \sigma_a^2 - 2 \times \sigma_p \times \sigma_a \times R \quad (40)$$

where E' is the centered root-mean-square error between the predicted and measured values, σ_p^2 and σ_a^2 are the variances of the predicted and measured values, respectively, and R is the correlation coefficient between the predicted and measured values.

Figures 10 and 11 depict how closely the six developed models match the training and testing sets, respectively, as references. In the Taylor diagram, the standard deviation is shown by the distance between the circles representing the models and the origin point on the x -axis, and the ticks on the clockwise arc represent the correlation coefficient. The actual PPV values are shown by the star-shaped point 'REF' on the x -axis, and the distance between the other circles and the point 'REF' reflects the centered RMSE (i.e., the grey arc).

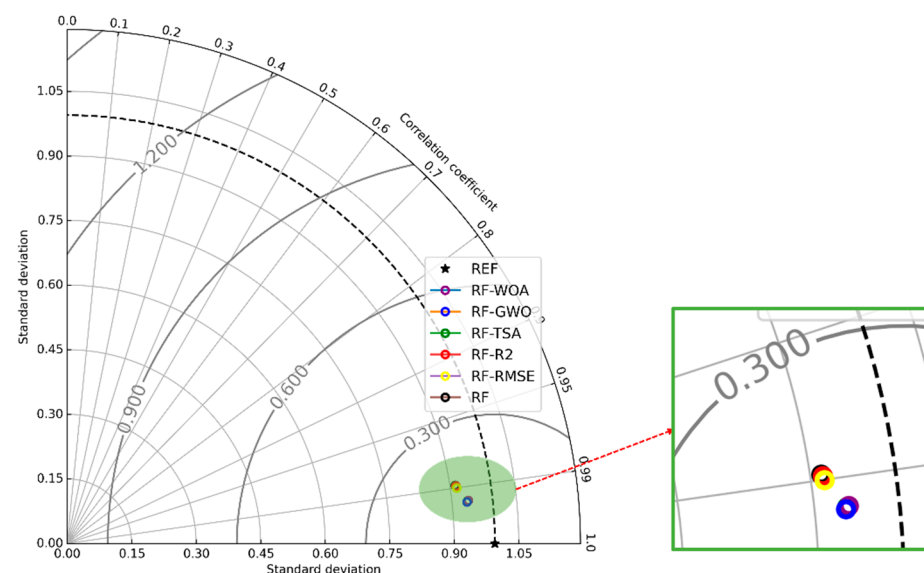


Figure 10. Taylor diagram of the training results for the six developed models.

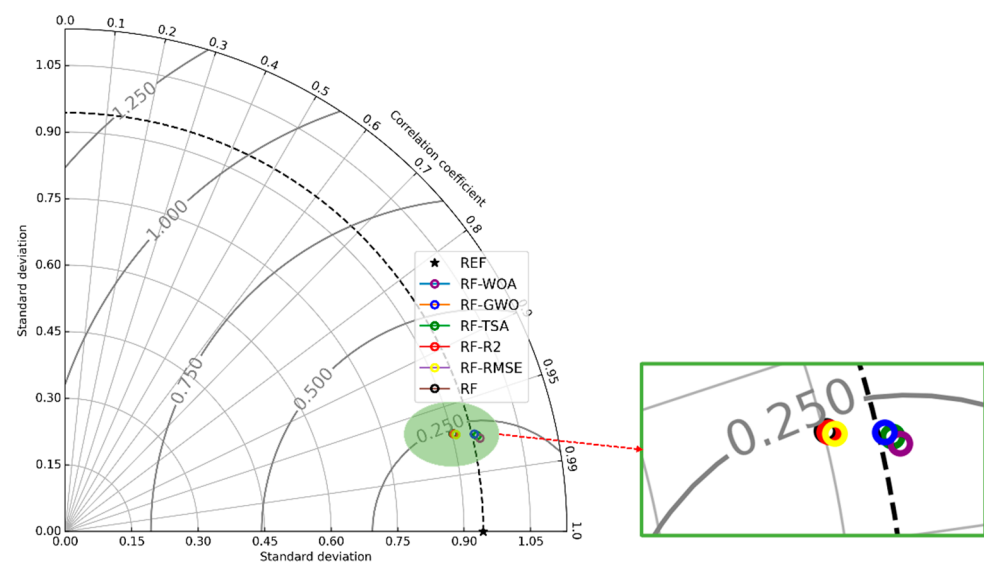


Figure 11. Taylor diagram of the testing results for the six developed models.

In Figure 10, it can be found that the standard deviations of all six models are less than the standard deviations of the measured values, and the standard deviations of the RF-WOA model, RF-GWO model, and RF-TSA model are closer to the standard deviation of the actual values compared with the models of the RF, RF-R², and RF-RMSE. Moreover, the same applies to the correlation coefficient, i.e., the correlation coefficients of the RF-WOA model, RF-GWO model, and RF-TSA model are closer to that of the actual values, but they differ little from each other. For the centered RMSE, the results indicate that the RF-WOA model, RF-GWO model, and RF-TSA model have smaller errors than those of the RF, RF-R² and RF-RMSE models. From a holistic perspective, the calculation results of the three metaheuristic algorithms on the training set are closer to the actual PPV values, while the RF-R² and RF-RMSE models perform slightly better on the training set than the RF model.

In Figure 11, it can be inferred that the standard deviations of the three metaheuristic algorithms on the testing set are close to the standard deviation of the actual PPV and slightly greater. With regard to the centered RMSE and correlation coefficient, the calculation results of the RF-WOA models are better than that of the RF-TSA and RF-GWO models. Overall, based on the distance between these circles and the point 'REF', it can be determined that the ranking of the model's superiority on the testing set is as RF-WOA > RF-TSA > RF-GWO > RF-RMSE > RF-R² > RF. Compared with the results on the training set, the results on the testing set can more clearly reflect the performance differences among these models, especially for the results obtained by the metaheuristic algorithms.

To sum up, the RF-WOA is the optimal model that is recommended for refining the weights of decision trees of the RF model in this study. Compared with the regular RF model, for the training set, the RF-WOA model increases the R² value of the RF from 0.973 to 0.986 and the VAF of the RF from 97.298 to 98.603 and simultaneously reduces the RMSE of the RF model from 0.164 to 0.118 and the MAE of the RF from 0.123 to 0.093. For the testing set, the RF-WOA model increases the R² value of the RF from 0.923 to 0.932 and the VAF value of the RF from 93.994 to 95.032 and simultaneously reduces the RMSE of the RF model from 0.262 to 0.246 and the MAE of the RF from 0.209 to 0.188. Finally, the measured and predicted PPV by the RF-WOA model on both training and test sets are depicted in Figures 12 and 13, respectively.

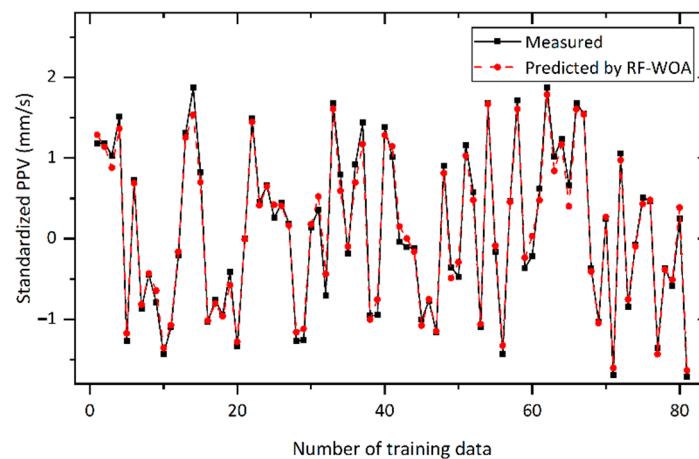


Figure 12. Measured and predicted PPV of the RF-WOA model on the training set.

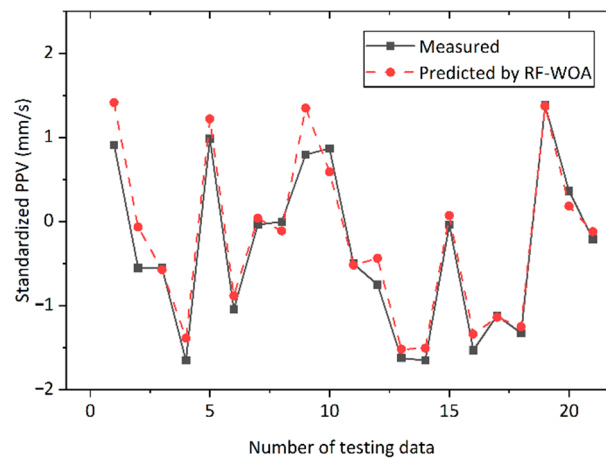


Figure 13. Measured and predicted PPV of the RF-WOA model on the testing set.

4.3. Sensitivity Analysis of Predictor Variables

Owing to the possible hazards of the PPV that are caused by blasting operations, it is indispensable to clarify the major factor affecting the PPV. Therefore, in this section, further analysis is conducted to identify the significant correlation of predictor variables that are used to predict the PPV. As previously stated, six variables, i.e., BS, HD, ST, PF, MC, and DI, were used to develop the RF models in this study. Based on the attribute of the RF model, a significant correlation of each input variable can be obtained. The (normalized) total decrease in the criterion brought by a feature is used to calculate the input variable's relevance with the target variable, which is also known as the Gini importance [56,57]; the higher it is, the more significant the input variable. In this way, the significant correlation of the input variables with PPV can be determined, as shown in Figure 14. It can be revealed that the most significant predictor variable is the PF, followed by the BS, DI, and HD. The least significant parameters include ST and MC. Accordingly, we will focus more on the relationship between the PF and PPV.

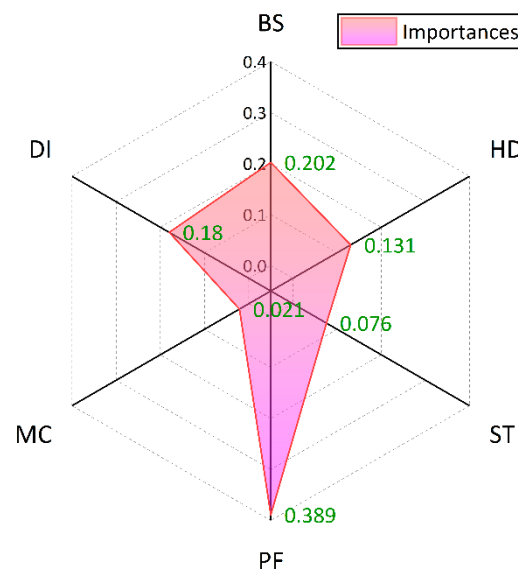


Figure 14. Significant correlation of the input variables with PPV.

To obtain a clearer picture of how the PF affects the PPV, partial dependency plots (PDPs) are utilized to achieve this goal. The PDPs illustrate the intuitive relationship between the target response and an input feature of interest and simultaneously marginalize the values of other input features. The PDPs visualize the average influence of the input feature on the target response, which can reveal a homogeneous connection between the feature and the target response [63]. Figure 15 shows the partial dependence between the PPV and PF. From an overall perspective, as the PF gradually increased from 0.23 to 0.94, the PPV gradually increased from 3.13 to 6.94. The process of the PF causing the PPV changes can be divided into three stages. The first stage is that when PF increases from 0.23 to 0.46, there is no significant change in the PPV, and the PPV reaches a minimum value when the PF equals 0.46. The second stage is that when the PF increases from 0.46 to 0.63, the PPV shows a dramatic increase from 3.01 to 6.27 in an exponential-like trend. The third stage is when the PF increases from 0.63 to 0.94, and the PPV increases from 6.27 to 6.94, but the amplification is significantly smaller compared with that in the second stage. From this, managing the PPV through curbing the size of the PF is a potentially effective means.

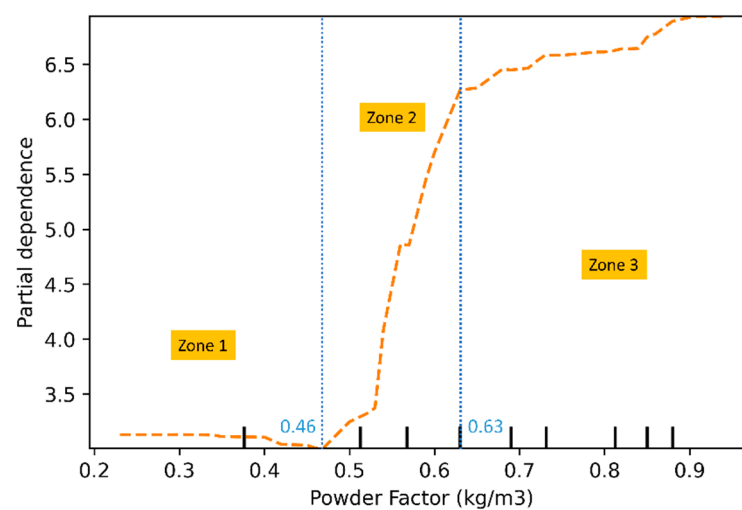


Figure 15. The partial dependency between the PPV and PF.

4.4. Comparison with the Published Works

As for the datasets used in the present paper, it has also been used in several published articles, e.g., references [1,64,65]. This section aims to compare the superiority of the model proposed in this paper with the existing models in the published papers.

Reference [64] proposed two models used for the PPV prediction, i.e., the nonlinear multiple regression (NLMR) and the gene expression programming (GEP). Likewise, 102 datasets of blasting operations were randomly divided into 80% for the training set and 20% for the testing set. The input parameters, including BS, HD, ST, PF, MC, and DI, are the same as the input parameters used in the present study. The results of the GEP and NLMR models are presented in Table 7. Reference [1] also used the same datasets to research the PPV prediction. In the research, a special case is that a feature selection technique was used to filter out unimportant input parameters before conducting the modeling. In this way, five variables except for BS, including HD, ST, PF, MC, and DI, were chosen as the input parameters, which is the main difference from the current paper. In their study, five ML models, i.e., CART, CHAID, RF, ANN, and SVM, were developed using the new datasets after feature selection. Furthermore, the datasets used for developing the models were randomly split into two parts, i.e., 70% of datasets for training the models and 30% of datasets for validating the performance of the models. The performance results of these five models are presented in Table 7. Reference [65] also utilized the same datasets to develop two ML models, i.e., RF and Bayesian Network (BN). Similar to Reference [1], feature selection was also used to filter features of low importance before building the ML model. After that, the input parameters that have been identified consist of DI, PF, HD, ST, and MC, in addition to BS.

Moreover, the datasets were randomly split into two parts, i.e., 70% of datasets for training the models and 30% of datasets for validating the performance of the models. The performance results of these two models are presented in Table 7. Overall, according to the results in Table 7, the improved RF model (i.e., the RF-WOA) proposed in this paper can significantly outperform the models proposed in the published papers, which proves the improved RF model is good enough and robust to predict the PPV caused by mining blasting. Likewise, this shows that it is reasonable and feasible to amend the weights of the decision trees of the RF model as proposed in this paper.

Table 7. Results of the evaluation metrics of the models that used the same datasets.

Reference	Models	Training				Testing			
		R ²	RMSE	MAE	VAF	R ²	RMSE	MAE	VAF
Present paper	RF-WOA	0.986	0.118	0.093	98.603	0.932	0.246	0.188	95.032
[64]	GEP	0.914	0.920	0.755	91.304	0.874	0.963	0.851	87.107
	NLMR	0.829	1.365	1.125	80.878	0.790	1.498	1.221	69.261
[1]	RF	0.940	0.770	0.620	92.970	0.830	1.460	1.190	82.170
	CART	0.670	1.670	1.320	67.030	0.560	2.390	1.840	54.600
	CHAID	0.910	0.860	0.540	91.300	0.680	1.900	1.470	67.790
	ANN	0.890	0.960	0.750	89.140	0.840	1.410	1.130	83.710
	SVM	0.880	1.020	0.770	88.480	0.850	1.500	1.170	84.540
[65]	RF	0.930	-	-	-	0.903	-	-	-
	BN	0.930	-	-	-	0.871	-	-	-

5. Conclusions

This paper utilized the RF and improved weighted RF models to predict the blast-induced PPV. Thus, a dataset of a total of 102 samples collected from an open granite mine was used to develop the target regular RF model. The input parameters used for modeling included BS, HD, ST, PF, MC, and DI, while the output was the PPV. Then, five techniques, i.e., refined weights based on the accuracy (R² and the RMSE) of decisions trees as well

as the optimization results of three metaheuristics algorithms (WOA, GWO, and TSA), were employed to enhance the performance of the regular RF model by reassigning the new weights of decision trees. The results concluded that the optimal hyperparameters of the regular RF model are 51 decision trees and 7 of maximum depth. Subsequently, the performance evaluation results of the five weighted RF models showed that all improved RF models outperformed the regular RF model. Moreover, the RF-WOA model shows the best performance among these five models, as evidenced by the fact that the RF-WOA model has the R^2 values of 0.986 and 0.932, the RMSE values of 0.118 and 0.246, the MAE values of 0.093 and 0.188, and the VAF values of 98.603 and 95.032 for the training and testing sets, respectively. Additionally, compared with the models developed in published articles that also used the same dataset, the RF-WOA model still shows the best accuracy, proving that the RF-WOA model developed in this study has better performance, adaptability, and robustness. Furthermore, the sensitivity analysis revealed that the PF model shows a great significant correlation with the PPV prediction, and the results of the PDPs indicated that when the PF is less than 0.46, the PPV maintains a small fluctuation. At the same time, when the PF exceeds 0.46, there is a significant increasing trend of PPV, and when the PF exceeds 0.63, the increasing trend of the PPV gradually shrinks, which suggests that governing the magnitude of the PPV by managing the PF is an effective and practical measure for the case of this project.

Author Contributions: Conceptualization, B.H. and A.S.M.; methodology, B.H., S.H.L.; software, B.H. and M.M.S.S.; validation, B.H., A.S.M.; formal analysis, B.H. and S.H.L.; investigation, B.H.; writing—original draft preparation, B.H., M.M.S.S., S.H.L., A.S.M. and D.V.U.; writing—review and editing, B.H., M.M.S.S., S.H.L., A.S.M. and D.V.U.; supervision, S.H.L., A.S.M. and M.M.S.S.; funding acquisition, M.M.S.S. All authors have read and agreed to the published version of the manuscript.

Funding: The research is partially funded by the Ministry of Science and Higher Education of the Russian Federation under the strategic academic leadership program ‘Priority 2030’ (Agreement 075-15-2021-1333 dated 30 September 2021).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are available upon request.

Acknowledgments: Authors of this study wish to express their appreciation to the Universiti Malaya for supporting this study and making it possible.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, H.; Zhou, J.; Armaghani, D.J.; Tahir, M.M.; Pham, B.T.; Huynh, V. Van A Combination of Feature Selection and Random Forest Techniques to Solve a Problem Related to Blast-Induced Ground Vibration. *Appl. Sci.* **2020**, *10*, 869. [\[CrossRef\]](#)
2. Hasanipanah, M.; Monjezi, M.; Shahnazar, A.; Armaghani, D.J.; Farazmand, A. Feasibility of indirect determination of blast induced ground vibration based on support vector machine. *Measurement* **2015**, *75*, 289–297. [\[CrossRef\]](#)
3. Yu, Z.; Shi, X.; Zhou, J.; Gou, Y.; Huo, X.; Zhang, J.; Armaghani, D.J. A new multikernel relevance vector machine based on the HPSOGWO algorithm for predicting and controlling blast-induced ground vibration. *Eng. Comput.* **2020**, *38*, 905–1920. [\[CrossRef\]](#)
4. Guo, H.; Zhou, J.; Koopialipoor, M.; Armaghani, D.J.; Tahir, M. Deep neural network and whale optimization algorithm to assess flyrock induced by blasting. *Eng. Comput.* **2021**, *37*, 173–186. [\[CrossRef\]](#)
5. Zhou, J.; Koopialipoor, M.; Murlidhar, B.R.; Fatemi, S.A.; Tahir, M.M.; Armaghani, D.J.; Li, C. Use of intelligent methods to design effective pattern parameters of mine blasting to minimize flyrock distance. *Nat. Resour. Res.* **2020**, *29*, 625–639. [\[CrossRef\]](#)
6. Guo, H.; Nguyen, H.; Bui, X.-N.; Armaghani, D.J. A new technique to predict fly-rock in bench blasting based on an ensemble of support vector regression and GLMNET. *Eng. Comput.* **2019**, *37*, 421–435. [\[CrossRef\]](#)
7. Yu, Q.; Monjezi, M.; Mohammed, A.S.; Dehghani, H.; Armaghani, D.J.; Ulrikh, D.V. Optimized support vector machines combined with evolutionary random forest for prediction of back-break caused by blasting operation. *Sustainability* **2021**, *13*, 12797. [\[CrossRef\]](#)
8. Sayadi, A.; Monjezi, M.; Talebi, N.; Khandelwal, M. A comparative study on the application of various artificial neural networks to simultaneous prediction of rock fragmentation and backbreak. *J. Rock Mech. Geotech. Eng.* **2013**, *5*, 318–324. [\[CrossRef\]](#)

9. Faramarzi, F.; Farsangi, M.A.E.; Mansouri, H. An RES-based model for risk assessment and prediction of backbreak in bench blasting. *Rock Mech. Rock Eng.* **2013**, *46*, 877–887. [\[CrossRef\]](#)
10. Afeni, T.B.; Osasan, S.K. Assessment of noise and ground vibration induced during blasting operations in an open pit mine—A case study on Ewekoro limestone quarry, Nigeria. *Min. Sci. Technol.* **2009**, *19*, 420–424. [\[CrossRef\]](#)
11. He, B.; Armaghani, D.J.; Lai, S.H. A Short Overview of Soft Computing Techniques in Tunnel Construction. *Open Constr. Build. Technol. J.* **2022**, *16*, 1–6. [\[CrossRef\]](#)
12. Lawal, A.I.; Kwon, S.; Kim, G.Y. Prediction of the blast-induced ground vibration in tunnel blasting using ANN, moth-flame optimized ANN, and gene expression programming. *Acta Geophys.* **2021**, *69*, 161–174. [\[CrossRef\]](#)
13. Rai, R.; Shrivastva, B.K.; Singh, T.N. Prediction of maximum safe charge per delay in surface mining. *Trans. Inst. Min. Metall. Sect. A Min. Technol.* **2005**, *114*, 227–232. [\[CrossRef\]](#)
14. Hasanipanah, M.; Bakhshandeh Amnieh, H.; Khamesi, H.; Jahed Armaghani, D.; Bagheri Golzar, S.; Shahnazar, A. Prediction of an environmental issue of mine blasting: An imperialistic competitive algorithm-based fuzzy system. *Int. J. Environ. Sci. Technol.* **2018**, *15*, 1–10. [\[CrossRef\]](#)
15. Hasanipanah, M.; Faradonbeh, R.S.; Amnieh, H.B.; Armaghani, D.J.; Monjezi, M. Forecasting blast-induced ground vibration developing a CART model. *Eng. Comput.* **2017**, *33*, 307–316. [\[CrossRef\]](#)
16. Ram Chandar, K.; Sastry, V.R.; Hegde, C.; Shreedharan, S. Prediction of peak particle velocity using multi regression analysis: Case studies. *Geomech. Geoengin.* **2017**, *12*, 207–214. [\[CrossRef\]](#)
17. Khandelwal, M.; Singh, T.N. Prediction of blast induced ground vibrations and frequency in opencast mine: A neural network approach. *J. Sound Vib.* **2006**, *289*, 711–725. [\[CrossRef\]](#)
18. Lawal, A.I.; Idris, M.A. An artificial neural network-based mathematical model for the prediction of blast-induced ground vibrations. *Int. J. Environ. Stud.* **2020**, *77*, 318–334. [\[CrossRef\]](#)
19. Parida, A.; Mishra, M.K. Blast Vibration Analysis by Different Predictor Approaches—A Comparison. *Procedia Earth Planet. Sci.* **2015**, *11*, 337–345. [\[CrossRef\]](#)
20. Xue, X.; Yang, X. Predicting blast-induced ground vibration using general regression neural network. *JVC/J. Vib. Control* **2014**, *20*, 1512–1519. [\[CrossRef\]](#)
21. Yang, H.; Wang, Z.; Song, K. A new hybrid grey wolf optimizer-feature weighted-multiple kernel-support vector regression technique to predict TBM performance. *Eng. Comput.* **2020**. [\[CrossRef\]](#)
22. Yang, H.; Song, K.; Zhou, J. Automated Recognition Model of Geomechanical Information Based on Operational Data of Tunneling Boring Machines. *Rock Mech. Rock Eng.* **2022**, *55*, 1499–1516. [\[CrossRef\]](#)
23. Armaghani, D.J.; Mohamad, E.T.; Narayanasamy, M.S.; Narita, N.; Yagiz, S. Development of hybrid intelligent models for predicting TBM penetration rate in hard rock condition. *Tunn. Undergr. Sp. Technol.* **2017**, *63*, 29–43. [\[CrossRef\]](#)
24. Armaghani, D.J.; Koopialipoor, M.; Marto, A.; Yagiz, S. Application of several optimization techniques for estimating TBM advance rate in granitic rocks. *J. Rock Mech. Geotech. Eng.* **2019**, *11*, 779–789. [\[CrossRef\]](#)
25. Zhou, J.; Yazdani Bejarbaneh, B.; Jahed Armaghani, D.; Tahir, M.M. Forecasting of TBM advance rate in hard rock condition based on artificial neural network and genetic programming techniques. *Bull. Eng. Geol. Environ.* **2020**, *79*, 2069–2084. [\[CrossRef\]](#)
26. Li, Z.; Yazdani Bejarbaneh, B.; Asteris, P.G.; Koopialipoor, M.; Armaghani, D.J.; Tahir, M.M. A hybrid GEP and WOA approach to estimate the optimal penetration rate of TBM in granitic rock mass. *Soft Comput.* **2021**, *25*, 11877–11895. [\[CrossRef\]](#)
27. Zhou, J.; Chen, C.; Wang, M.; Khandelwal, M. Proposing a novel comprehensive evaluation model for the coal burst liability in underground coal mines considering uncertainty factors. *Int. J. Min. Sci. Technol.* **2021**, *31*, 14. [\[CrossRef\]](#)
28. Zhou, J.; Li, X.; Mitri, H.S. Classification of rockburst in underground projects: Comparison of ten supervised learning methods. *J. Comput. Civ. Eng.* **2016**, *30*, 4016003. [\[CrossRef\]](#)
29. Pham, B.T.; Nguyen, M.D.; Nguyen-Thoi, T.; Ho, L.S.; Koopialipoor, M.; Kim Quoc, N.; Armaghani, D.J.; Le, H. Van A novel approach for classification of soils based on laboratory tests using Adaboost, Tree and ANN modeling. *Transp. Geotech.* **2021**, *27*, 100508. [\[CrossRef\]](#)
30. Armaghani, D.J.; Harandizadeh, H.; Momeni, E.; Maizir, H.; Zhou, J. *An Optimized System of GMDH-ANFIS Predictive Model by ICA for Estimating Pile Bearing Capacity*; Springer: Cham, The Netherlands, 2022; Volume 55, ISBN 0123456789.
31. Huat, C.Y.; Moosavi, S.M.H.; Mohammed, A.S.; Armaghani, D.J.; Ulrikh, D.V.; Monjezi, M.; Hin Lai, S. Factors Influencing Pile Friction Bearing Capacity: Proposing a Novel Procedure Based on Gradient Boosted Tree Technique. *Sustainability* **2021**, *13*, 11862. [\[CrossRef\]](#)
32. Huang, J.; Sun, Y.; Zhang, J. Reduction of computational error by optimizing SVR kernel coefficients to simulate concrete compressive strength through the use of a human learning optimization algorithm. *Eng. Comput.* **2021**. [\[CrossRef\]](#)
33. Huang, J.; Kumar, G.S.; Ren, J.; Zhang, J.; Sun, Y. Accurately predicting dynamic modulus of asphalt mixtures in low-temperature regions using hybrid artificial intelligence model. *Constr. Build. Mater.* **2021**, *297*, 123655. [\[CrossRef\]](#)
34. Asteris, P.G.; Lourenço, P.B.; Roussis, P.C.; Adami, C.E.; Armaghani, D.J.; Cavaleri, L.; Chalioris, C.E.; Hajihassani, M.; Lemonis, M.E.; Mohammed, A.S. Revealing the nature of metakaolin-based concrete materials using artificial intelligence techniques. *Constr. Build. Mater.* **2022**, *322*, 126500. [\[CrossRef\]](#)
35. Mahmood, W.; Mohammed, A.S.; Asteris, P.G.; Kurda, R.; Armaghani, D.J. Modeling Flexural and Compressive Strengths Behaviour of Cement-Grouted Sands Modified with Water Reducer Polymer. *Appl. Sci.* **2022**, *12*, 1016. [\[CrossRef\]](#)

36. Asteris, P.G.; Rizal, F.I.M.; Koopialipoor, M.; Roussis, P.C.; Ferentinou, M.; Armaghani, D.J.; Gordan, B. Slope Stability Classification under Seismic Conditions Using Several Tree-Based Intelligent Techniques. *Appl. Sci.* **2022**, *12*, 1753. [CrossRef]
37. Cai, M.; Koopialipoor, M.; Armaghani, D.J.; Thai Pham, B. Evaluating Slope Deformation of Earth Dams due to Earthquake Shaking using MARS and GMDH Techniques. *Appl. Sci.* **2020**, *10*, 1486. [CrossRef]
38. Zhou, J.; Dai, Y.; Khandelwal, M.; Monjezi, M.; Yu, Z.; Qiu, Y. Performance of Hybrid SCA-RF and HHO-RF Models for Predicting Backbreak in Open-Pit Mine Blasting Operations. *Nat. Resour. Res.* **2021**, *30*, 4753–4771. [CrossRef]
39. Zhou, J.; Qiu, Y.; Khandelwal, M.; Zhu, S.; Zhang, X. Developing a hybrid model of Jaya algorithm-based extreme gradient boosting machine to estimate blast-induced ground vibrations. *Int. J. Rock Mech. Min. Sci.* **2021**, *145*, 104856. [CrossRef]
40. Li, C.; Zhou, J.; Armaghani, D.J.; Li, X. Stability analysis of underground mine hard rock pillars via combination of finite difference methods, neural networks, and Monte Carlo simulation techniques. *Undergr. Sp.* **2021**, *6*, 379–395. [CrossRef]
41. Parsajoo, M.; Armaghani, D.J.; Mohammed, A.S.; Khari, M.; Jahandari, S. Tensile strength prediction of rock material using non-destructive tests: A comparative intelligent study. *Transp. Geotech.* **2021**, *31*, 100652. [CrossRef]
42. Asteris, P.G.; Mamou, A.; Hajihassani, M.; Hasanipanah, M.; Koopialipoor, M.; Le, T.-T.; Kardani, N.; Armaghani, D.J. Soft computing based closed form equations correlating L and N-type Schmidt hammer rebound numbers of rocks. *Transp. Geotech.* **2021**, *29*, 100588. [CrossRef]
43. Momeni, E.; Armaghani, D.J.; Hajihassani, M.; Amin, M.F.M. Prediction of uniaxial compressive strength of rock samples using hybrid particle swarm optimization-based artificial neural networks. *Measurement* **2015**, *60*, 50–63. [CrossRef]
44. Huang, J.; Zhang, J.; Gao, Y. Intelligently predict the rock joint shear strength using the support vector regression and Firefly Algorithm. *Lithosphere* **2021**, *2021*, 2467126. [CrossRef]
45. Lawal, A.I. An artificial neural network-based mathematical model for the prediction of blast-induced ground vibration in granite quarries in Ibadan, Oyo State, Nigeria. *Sci. African* **2020**, *8*, e00413. [CrossRef]
46. Rana, A.; Bhagat, N.K.; Jadaun, G.P.; Rukhaiyar, S.; Pain, A.; Singh, P.K. Predicting Blast-Induced Ground Vibrations in Some Indian Tunnels: A Comparison of Decision Tree, Artificial Neural Network and Multivariate Regression Methods. *Min. Metall. Explor.* **2020**, *37*, 1039–1053. [CrossRef]
47. Lawal, A.I.; Kwon, S.; Hammed, O.S.; Idris, M.A. Blast-induced ground vibration prediction in granite quarries: An application of gene expression programming, ANFIS, and sine cosine algorithm optimized ANN. *Int. J. Min. Sci. Technol.* **2021**, *31*, 265–277. [CrossRef]
48. Bui, X.N.; Nguyen, H.; Nguyen, T.A. Artificial Neural Network Optimized by Modified Particle Swarm Optimization for Predicting Peak Particle Velocity Induced by Blasting Operations in Open Pit Mines. *Inz. Miner.* **2021**, *1*, 79–90. [CrossRef]
49. Iphar, M.; Yavuz, M.; Ak, H. Prediction of ground vibrations resulting from the blasting operations in an open-pit mine by adaptive neuro-fuzzy inference system. *Environ. Geol.* **2008**, *56*, 97–107. [CrossRef]
50. Monjezi, M.; Ghafurikalajahi, M.; Bahrami, A. Prediction of blast-induced ground vibration using artificial neural networks. *Tunn. Undergr. Sp. Technol.* **2011**, *26*, 46–50. [CrossRef]
51. Ghasemi, E.; Ataei, M.; Hashemolhosseini, H. Development of a fuzzy model for predicting ground vibration caused by rock blasting in surface mining. *J. Vib. Control* **2013**, *19*, 755–770. [CrossRef]
52. Hajihassani, M.; Jahed Armaghani, D.; Marto, A.; Tonnizam Mohamad, E. Ground vibration prediction in quarry blasting through an artificial neural network optimized by imperialist competitive algorithm. *Bull. Eng. Geol. Environ.* **2014**, *74*, 873–886. [CrossRef]
53. Armaghani, D.J.; Hasanipanah, M.; Amnieh, H.B.; Mohamad, E.T. Feasibility of ICA in approximating ground vibration resulting from mine blasting. *Neural Comput. Appl.* **2018**, *29*, 457–465. [CrossRef]
54. Mohamed, M.T. Performance of fuzzy logic and artificial neural network in prediction of ground and air vibrations. *Int. J. Rock Mech. Min. Sci.* **2011**, *39*, 425–440.
55. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
56. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey wolf optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [CrossRef]
57. Mirjalili, S.; Lewis, A. The whale optimization algorithm. *Adv. Eng. Softw.* **2016**, *95*, 51–67. [CrossRef]
58. Kaur, S.; Awasthi, L.K.; Sangal, A.L.; Dhiman, G. Tunicate Swarm Algorithm: A new bio-inspired based metaheuristic paradigm for global optimization. *Eng. Appl. Artif. Intell.* **2020**, *90*, 103541. [CrossRef]
59. Li, H.B.; Wang, W.; Ding, H.W.; Dong, J. Trees Weighting Random Forest method for classifying high-dimensional noisy data. In Proceedings of the 2010 IEEE 7th International Conference on E-Business Engineering, Shanghai, China, 10–12 November 2010; pp. 160–163.
60. Zorlu, K.; Gokceoglu, C.; Ocakoglu, F.; Nefeslioglu, H.A.; Acikalin, S. Prediction of uniaxial compressive strength of sandstones using petrography-based models. *Eng. Geol.* **2008**, *96*, 141–158. [CrossRef]
61. Taylor, K.E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmos.* **2001**, *106*, 7183–7192. [CrossRef]
62. Taylor, K.E. Taylor Diagram Primer—Working Paper. 2005. Available online: http://www.atmos.albany.edu/daes/atmclasses/atm401/spring_2016/ppts_pdfs/Taylor_diagram_primer.pdf (accessed on 10 May 2022).
63. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
64. Shirani Faradonbeh, R.; Jahed Armaghani, D.; Abd Majid, M.Z.; MD Tahir, M.; Ramesh Murlidhar, B.; Monjezi, M.; Wong, H.M. 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.* **2016**, *13*, 1453–1464. [CrossRef]

-
65. Zhou, J.; Asteris, P.G.; Armaghani, D.J.; Pham, B.T. Prediction of ground vibration induced by blasting operations through the use of the Bayesian Network and random forest models. *Soil Dyn. Earthq. Eng.* **2020**, *139*, 106390. [[CrossRef](#)]