Natural Resources Research (© 2024)
https://doi.org/10.1007/s11053-023-10299-w

Original Paper

Enhanced Prediction Model for Blast-Induced Air Over-Pressure in Open-Pit Mines Using Data Enrichment and Random Walk-Based Grey Wolf Optimization–Two-Layer ANN Model

Hoang Nguyen,[1,3,5] Xuan-Nam Bui,[1,3] Carsten Drebenstedt,[2,3] and Yosoon Choi[4]

Received 1 September 2023; accepted 12 December 2023

In this study, two innovative techniques were introduced, including data enrichment and optimization, with the aim of significantly improving the accuracy of air over-pressure (AOP) prediction models in mine blasting. Firstly, the Extra Trees algorithm was applied to enrich the collected dataset with the goal of enhancing the understanding of the predictive models for AOP prediction. Then, a neural network model with two hidden layers (ANN) was designed to predict AOP using both the original and enriched datasets. Secondly, to further enhance the accuracy of the ANN model, a novel optimization algorithm based on a random walk strategy and the grey wolf optimization algorithm (RWGWO) was employed to optimize the weights of the ANN model. This optimized model, referred to as the RWGWO–ANN model, was developed and evaluated for predicting AOP using both the original and enriched datasets. To comprehensively assess the impact of data enrichment and the proposed RWGWO-ANN model, three other optimization algorithms—particle swarm optimization (PSO), fruit-fly optimization algorithm (FOA), and single-based genetic algorithm (SGA)—were also applied to optimize the ANN model for AOP prediction. These models were named PSO–ANN, FOA–ANN, and SGA–ANN, respectively. The tenfold cross-validation procedure was applied and repeated three times to ensure the objectivity and consistency of the models. Additionally, conventional ANN and the United States Bureau of Mines empirical model were developed for comparison, serving similar purposes to evaluate the efficiency of the optimization algorithms employed in this study. To demonstrate the advantages of the proposed method and models, a dataset comprising 312

[1]Department of Surface Mining, Mining Faculty, Hanoi University of Mining and Geology, Bac Tu Liem District, 18 Vien Str., Duc Thang Ward, Hanoi 100000, Vietnam.
[2]Institute of Mining and Special Engineering, Freiberg University of Mining and Technology, Gustav-Zeuner Strasse 1A, 09599 Freiberg, Sachsen, Germany.
[3]Innovations for Sustainable and Responsible Mining (ISRM) Research Group, Hanoi University of Mining and Geology, Bac Tu Liem District, 18 Vien Str., Duc Thang Ward, Hanoi 100000, Vietnam.
[4]Department of Energy Resources Engineering, Pukyong National University, Busan 48513, Korea.
[5]To whom correspondence should be addressed; e-mail: nguyenhoang@humg.edu.vn

Published online: 24 January 2024

blasting events and six input parameters at the Coc Sau open-pit coal mine in Vietnam was gathered and analyzed. These parameters included burden, spacing, rock hardness, powder factor, monitoring distance, and maximum explosive charge per delay. An additional input variable—Extra Trees—was introduced, making the total number of input variables seven in the enriched dataset. The proposed hybrid model, along with others, was developed based on both the original and enriched datasets. The results revealed that the Extra Trees algorithm is robust and effectively enriches the raw dataset, enhancing the understanding of predictive models and providing improved accuracy. Sensitivity analysis results also highlighted the robust contribution of the Extra Trees variable in the enriched dataset. Compared to the original dataset, the performance of AOP predictive models was improved by 7–24% using the enriched dataset enriched by the Extra Trees algorithm. Furthermore, the findings indicated that the RWGWO–ANN model exhibited the highest accuracy in predicting AOP in this study, achieving an accuracy of 96.2%. This marked a 16–20% improvement over the accuracy of the conventional ANN model.

KEY WORDS: Open-pit mine safety, Mine blasting, Air over-pressure, Sustainable and responsible mining, Metaheuristic algorithms, Data enrichment, Random walk grey wolf optimization, Extra Trees.

INTRODUCTION

Rocks and ores in open-pit mines typically possess high degrees of hardness, necessitating their breaking prior to shoveling/loading, hauling/transporting, crushing, and other associated tasks. To achieve optimal rock fragmentation, drilling and blasting are widely acknowledged as the most efficient techniques. However, these methods have negative consequences, including air over-pressure (AOP), ground vibration (PPV), flyrock, and air pollution, which can endanger the nearby environment, especially neighboring areas and residential zones. Among these, AOP stands out as a significant contributor to environmental and human health degradation. The powerful shockwaves produced by explosive activities in open-pit mines can damage nearby structures like buildings, windows, and pipelines. This damage can be immediate or may occur over time due to repeated exposure. At a high-level of AOP, it can cause permanent hearing loss or damage to people who are exposed to it without adequate hearing protection (e.g., workers on the working sites of mines). In addition, AOP can cause disturbance to wildlife and disrupt their habitats. Thus, accurately predicting AOP poses a challenge in mitigating its negative effects on the surrounding environment and in promoting sustainable and responsible mining practices.

Numerous strategies have been explored by researchers to tackle the above-mentioned issue, including practical measures/techniques (such as adjusting blasting parameters, utilizing air-deck stemming, etc.), as well as the use of empirical equations to estimate AOP. In recent times, there has been a focus on soft computing approaches, such as artificial intelligence (AI) and machine learning (ML). These models have been created and studied, revealing encouraging outcomes in forecasting AOP. It is worth mentioning that these AI/ML models exhibited superior precision in AOP prediction compared to traditional empirical models (Armaghani et al., 2016).

Concerning the anticipation of AOP, Armaghani et al. (2015b) employed the adaptive neuro-fuzzy inference system (ANFIS) model to predict it. This involved taking into account blasting parameters and the distance (D) between monitoring positions and blast sites. The outcomes underscored the ANFIS model's supremacy, achieving a precision level of 96%. This surpassed the predictive prowess of both a pre-existing artificial neural network (ANN) model and a multiple linear regression model for AOP prediction. In another study, Ramesh Murlidhar et al. (2021) developed two tree-based ML models, including genetic programming (GP) and M5′ decision tree models for predicting AOP. The most accurate model found in their study had R^2 of 0.862.

By another approach based on the combination of multiple ML/AI models, Nguyen and Bui (2019) utilized the random forest (RF) algorithm to create a new model called ANNs–RF for predicting AOP by combining the outcome predictions of pre-developed ANN models. The model was tested at an open-pit coal mine in Vietnam and it outperformed other models, with MAE (mean absolute error) of 0.620, RMSE (root mean squared error) of 0.847, and R^2 of 0.985. The results showed significant improvements when combining multiple algorithms as such. Similarly, Armaghani et al. (2016) showed that the ANN model's performance can be improved by the imperialist competitive algorithm (ICA), named as ICA–ANN, and they developed this model to predict AOP in three quarries (Malaysia) with an accuracy of approximately 96%. Hasanipanah et al. (2017) utilized another optimization algorithm based on the swarm behaviors in nature (i.e., particle swarm optimization algorithm—PSO) to optimize the support vector machine algorithm for predicting AOP, resulting in the PSO–SVR model. Different kernel functions were experimented with while creating the PSO–SVR model for forecasting AOP, employing a comparable methodology. The suggested PSO–SVR model established a correlation coefficient of 0.997 within the testing dataset. In another work, Harandizadeh and Armaghani (2021) developed a model using a combination of an ANFIS, polynomial neural network (PNN), and optimization through the genetic algorithm (GA) to forecast AOP levels. The Fuzzy Delphi (FD) method was applied to select the potential features before developing the GA–ANFIS–PNN model. Finally, a correlation coefficient of 0.92 was defined for the proposed GA–ANFIS–PNN model in predicting AOP. In 2022, Ye et al. (2022) employed the stochastic fractal search (SFS) algorithm for optimizing the ANFIS model with a similar objective. They subsequently compared the performance of the SFS–ANFIS model to the PSO–ANFIS and GA–ANFIS models. The outcomes demonstrated the superiority of the SFS–ANFIS model, as it achieved the highest accuracy, with RMSE of 1.223 dB. Similar works for predicting AOP using ML/AI-based models are introduced in the literature (Armaghani et al., 2015a; Amiri et al., 2016; Nguyen et al., 2020; He et al., 2021; Zhang et al., 2022).

After reviewing the literature, it is evident that ML/AI-based models have been used extensively in predicting AOP with promising results. Among these models, hybrid approaches that combine multiple algorithms tend to exhibit higher accuracy compared to individual models. On the other hand, in the mining industry, particularly concerning blasting effects, data represent one of the most significant challenges. This challenge can impact the accuracy of prediction and optimization models due to the influences of geological and geographical conditions and their inherent uncertainties. Typically, to enhance the accuracy of prediction models, researchers often apply various data analysis techniques, such as feature selection, data clustering, imputation of missing data, outlier handling, dataset scaling, and data transformation (Refaat, 2010; García et al., 2015; de-Paz-Centeno et al., 2023). It is advisable to utilize these methods as they have the potential to enhance the effectiveness of predictive models. Bagging and stacking techniques have also been applied to combine the outcome predictions of weak models into a new dataset to develop stronger models for predicting AOP, PPV, flyrock, and other parameters (Nguyen & Bui, 2018; Tran et al., 2023). However, there is still a need for the development of new hybrid models and handling the dataset to further enhance predictive accuracy and contribute to advancing knowledge in this field.

This research aimed to fill the existing voids by presenting an innovative method in data augmentation. This technique focuses on enriching data to offer supplementary information to the initial dataset, benefiting predictive models for AOP through the utilization of the Extra Tree (ExTree) algorithm. The enriched dataset was then used to develop a novel hybrid intelligence model called RWGWO–ANN for predicting AOP in open-pit mines. This model combines the innovative random walk grey wolf optimization algorithm (RWGWO) with an ANN model, resulting in the RWGWO–ANN model. Both of these techniques aim to significantly improve the accuracy of AOP predictions in mine blasting.

In addition to the RWGWO–ANN model, other hybrid models such as PSO–ANN, FOA–ANN (fruit-fly optimization algorithm), and SGA–ANN (single-based GA), as well as conventional ANN models and an empirical model (i.e., United States Bureau of Mines—USBM), were developed and compared to showcase the novelty and superiority of the RWGWO–ANN model, along with the effectiveness of the Extra Tree algorithm in data enrichment for predicting AOP.

ARTIFICIAL INTELLIGENCE APPROACHES

Extra Tree for Data Enrichment

The ExTree model, as introduced by Maier et al. (2015), is a type of ensemble machine learning method built on the foundations of supervised

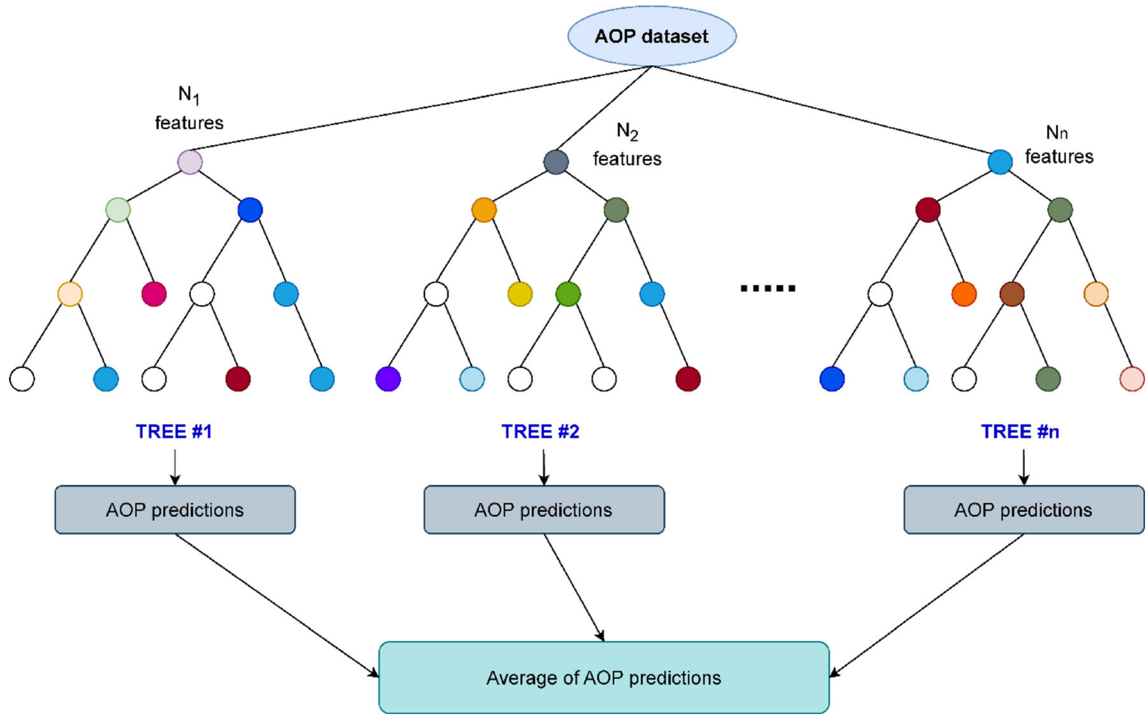


Figure 1. Configuration of the ExTree model aimed at predicting AOP.

learning. Renowned for its high level of randomization, this model consists of a collection of highly randomized trees, serving purposes in both regression and classification (Geurts et al., 2006). Acting as an extension of the random forest (RF) model, ExTree was formulated to tackle the challenge of overfitting.

Much akin to the RF algorithm, ExTree employs randomized subsets for training individual base models. These model predictions are then harmonized into a cohesive framework, yielding outcome predictions (John et al., 2015). Nonetheless, ExTree distinguishes itself by opting for optimal feature selection via randomized node splits. Structurally, ExTree encompasses numerous decision trees, each comprising a root node, split nodes, and leaf nodes, as illustrated in Figure 1. Given a dataset (e.g., AOP dataset in this study), ExTree initiates the process by segmenting data into randomized feature subsets at the root node. Each of these subsets evolves into a split or child node, with splitting progressing until a leaf node is reached. Within each tree, predictions are computed, eventually amalgamated across trees. The model's official outcome is then established as the average of these predictions, particularly for regression tasks.

When constructing the ExTree model, three vital parameters come into play: the quantity of trees, the selection of random features, and the minimum samples required for splitting. An inherent strength of ExTree lies in its ability to diminish variance and bias within the training dataset, accomplished through cut-point selection and the deliberate randomization of attribute subsets (Saeed et al., 2021). In a previous study (Tran et al., 2023), we introduced the ExTree model and its stacking model (BA–ExTree) for predicting PPV with promising results. The ExTree algorithm was applied in this research to enhance the initial dataset's content, aiming to predict AOP caused by blasting activities in open-pit mines.

The novelty of this work compared to the ensemble methods is the use of machine learning (i.e., ExTree algorithm) for data enrichment. In this approach, the ExTree algorithm was employed to predict blast-induced AOP in open-pit mines using the original dataset. Then, the AOP predictions from the ExTree model were utilized as an additional input variable for developing other predictive models. This process enhances the overall predictive capabilities and the insights of the original dataset generated by the ExTree algorithm. Meanwhile, in traditional ensemble methods, multiple models are combined to make predictions. These models can be combined by the same algorithms (e.g., random forest) or different algorithms (e.g., neural networks, random forest, support vector machine, etc.). The predictions from these individual models are then averaged or combined to create a final prediction.

Prediction Model with Two-Layer ANN Model

With its robust processing and calculation abilities inspired by the human brain, ANN has gained renown as a promising tool for addressing a multitude of real-world problems in the fields of mining engineering (Tadeusiewicz, 2015; Kan, 2017; Bui et al., 2021), geotechnical engineering (Chao et al., 2018; Baghbani et al., 2022), geosciences (Li & Zheng, 2003; Dramsch, 2020; Zhang et al., 2021), environment and climate change (Liu et al., 2010; Han & Wang, 2021), to name a few. Numerous types of neural networks exist, featuring a range of architectures that span from straightforward to intricate topological structures, e.g., ANN, extreme learning machine neural network, recurrent neural network, convolutional neural network, long-short terms memory neural network, among others. Of those, ANN is well-known as one of the simple neural networks, and it can solve many problems with a simple topology network (i.e., containing a single hidden layer).

Many previous works applied a simple ANN model with a single hidden layer to predict AOP or other side effects of blasting operations in open-pit mines (Saadat et al., 2014; Mohamad et al., 2016; Hosseini et al., 2021; Murlidhar et al., 2021). This simple network was chosen primarily because it can prevent the overfitting problem that occurs with small datasets. Nonetheless, researchers face the challenge of improving network accuracy without encountering overfitting issues.

In order to enhance the accuracy of ANN models, several techniques can be employed. These techniques include increasing the model size by adding more neurons in each layer, using a deeper network with additional layers, implementing regularization techniques like Lasso regularization (L1) or Ridge regularization (L2), employing various activation functions such as ReLU, sigmoid, and tanh, applying batch normalization, utilizing different optimization algorithms like Adam, SGD, RMSprop, and metaheuristic algorithms, as well as implementing feature scaling techniques.

In this study, the following techniques were applied to enhance the accuracy of the ANN model for predicting AOP: (1) a topology network with two hidden layers (ANN); (2) feature scaling technique; (3) different activation functions; and (4) various optimization algorithms for training the ANN model. The general structure of an ANN model for predicting AOP is illustrated in Figure 2.

Principle of Optimization Algorithms

Random Walk Grey Wolf Optimization Algorithm

RWGWO is an enhanced iteration of the grey wolf optimization (GWO) algorithm, initially presented by Gupta and Deep (2019), for refining the search capability through the leadership within the GWO algorithm (i.e., α, β, δ). Accordingly, the leading wolves play a crucial role as responsible search agents in guiding the pack towards the optimal direction to approach the prey. It is crucial for these dominant wolves to have the highest level of fitness during each iteration, and so they can offer the best possible guidance to the rest of the wolves. However, a question arises regarding how the α wolf, being the dominant wolf, benefits from the guidance of the less fit (inferior) β, δ wolves to update its position. Likewise, what is the reason for the β wolf adjusting its location in collaboration with the subordinate δ wolf? This limitation in the GWO algorithm hinders the convergence of the pack towards the global optima. The selection of leading wolves becomes critical during each iteration as they update the position of every wolf in the pack. An enhancement is necessary in the selection of pack leaders to address the issue of early convergence due to getting stuck in local optimal points and to ensure the continuation of social interactions within the pack. To address these issues, Gupta and Deep (2019) proposed the use of a random walk technique as an improvization of the GWO algorithm. This technique overcomes the drawbacks mentioned above and enhance the algorithm's overall effectiveness, and it is expressed as:

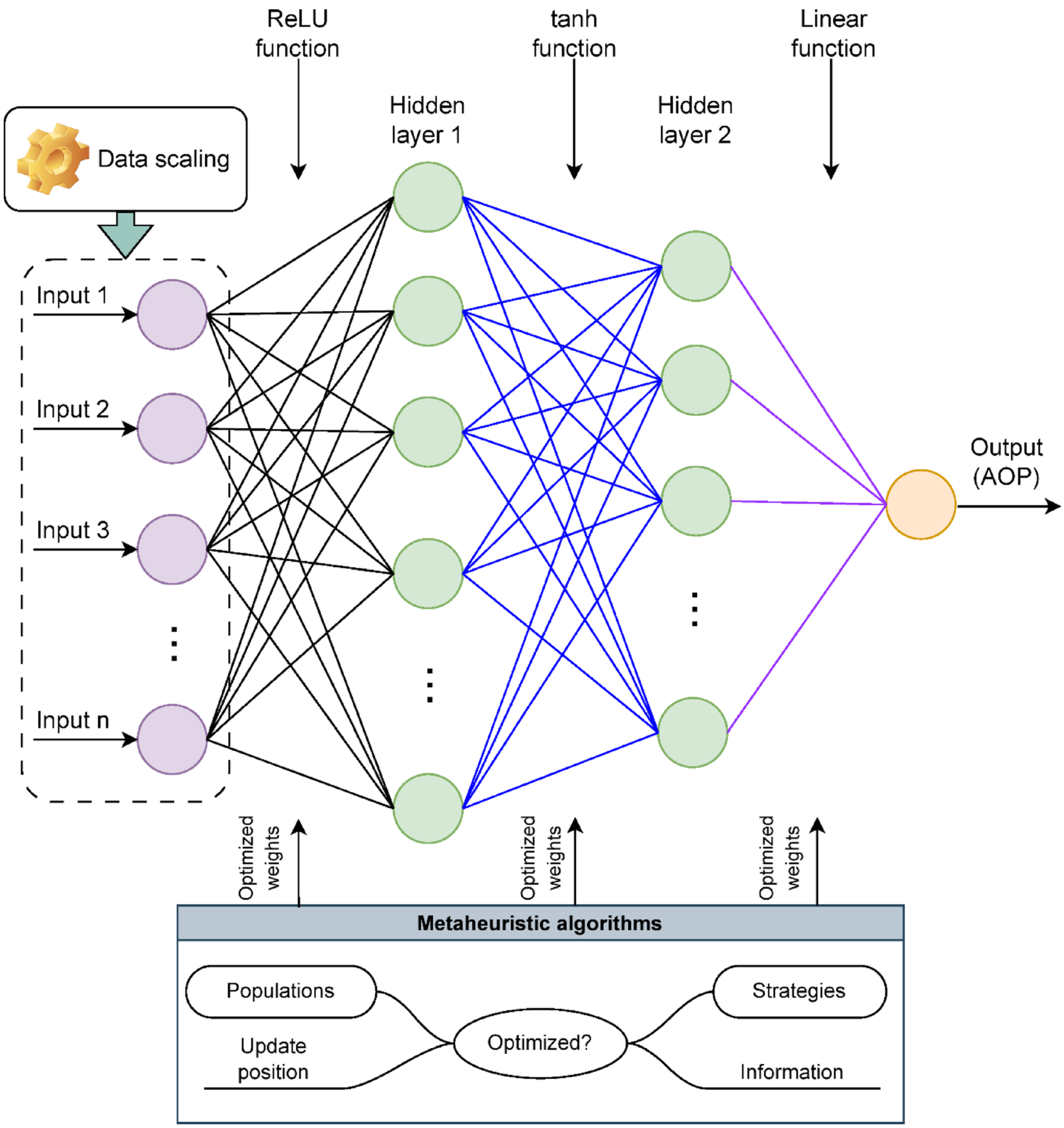


Figure 2. Structure of the AOP prediction model and techniques used for the network.

$$\mathrm{RW}_N = \sum_{i=1}^{N} \mathrm{Step}_i \tag{1}$$

where Step_i denotes the ith random step that is selectable from any random distribution, and it can be expressed as:

$$\mathrm{RW}_N = \sum_{i=1}^{N} \mathrm{Step}_i = \sum_{i=1}^{N-1} \mathrm{Step}_i + L_N = \mathrm{RW}_N - 1 + \mathrm{Step}_N \tag{2}$$

where the relationship between the current state ($\mathrm{RW}_N - 1$) and the next state (RW_N) demonstrates that the latter is solely influenced by the former and the step taken from the former to reach the latter.

The step size (Step_i) can either remain constant or vary.

Finally, a random walk can be defined as:

$$A_n = a_0 + \alpha_1\text{Step}_1 + \alpha_2\text{Step}_2 + \cdots + \alpha_N\text{Step}_N = a_0 + \sum_{i=1}^{N} \alpha_i\text{Step}_i \quad (3)$$

where we assumed that a wolf starting at a point a_0 and its final location as a_N. The α_i stands for the control parameter of Step_i and $\alpha_i>0$.

In order to implement the random walk in the GWO algorithm (RWGWO), step sizes drawn from a Cauchy distribution are performed. From the initial population of wolves, in each iteration, a random walk is applied exclusively to the leaders α, β, δ of the population. The parameter α_i is a vector that linearly decreases from 2 to 0 as the iterations progress. This allows for occasional large jumps, which can be highly effective during periods of stagnation. These larger jumps assist the leading wolves in exploring the search space to locate prey and provide valuable guidance to the other wolves. Importantly, it should be noted that no additional efforts are made to evaluate the objective function in the algorithm. Thus, the number of function evaluations remains unchanged in both algorithms. Figure 3 depicts the pseudo-code for the RWGWO algorithm.

Particle Swarm Optimization Algorithm

In this research, the PSO algorithm was utilized to optimize the ANN model for the prediction of AOP. This was then contrasted with the suggested RWGWO–ANN model. Comprehensive insights regarding the PSO algorithm can be found in Bansal (2019), Bensingh et al. (2019) and Guo et al. (2019). The PSO algorithm, introduced by Eberhart and Kennedy (1995), stands as a remarkably effective metaheuristic approach influenced by the actions of social animals or particles, like a swarm of birds in flight. Its application revolves around tackling optimization issues, wherein every conceivable solution is depicted as an individual particle. The sequential phases of the PSO algorithm were as delineated by Kennedy (2011):

Phase 1: Commence by initializing a population of particles along with their corresponding velocities. Evaluate the particles' fitness and ascertain both the most favorable local and global positions.

Phase 2: Every individual particle fine-tunes its position based on its velocity. During each cycle, determine the global paramount and local best positions to gauge the efficacy of the PSO–ANN model in prognosticating AOP. The global paramount denotes the unsurpassed position achieved by any particle, while the local best signifies the optimal solution within the ongoing cycle.

Phase 3: Revise the position of each particle. Subsequent to predicting particle velocity, tweak their positions within the exploration domain using the computed velocity. Update the new velocity as:

$$v_j^{i+1} = \varpi v_j^{(i)} + \left(C_1 \times r_3 \times \left(L_{\text{best\_}j} - x_j^{(i)}\right)\right) + \left(C_2 \times r_4 \times \left(G_{\text{best\_}j} - x_j^{(i)}\right)\right); \quad v_{\min} \le v_j^{(i)} \le v_{\max} \quad (8)$$

where particle position, velocity, inertial weight coefficient (ϖ), iteration number (i), and random numbers (r_3, r_4) are taken into account. When a new and improved solution is found, updates can be made to both the global best position and the local best positions. The position of each particle is then recalculated and modified as:

$$x_j^{i+1} = x_j^{(i)} + v_j^{(i+1)}; \ j = 1, 2, \ldots, n \quad (9)$$

Upon reaching the conclusion, the termination criteria are assessed. If these criteria are met, the optimized solution for the current problem is deemed to be the global best position. Refer to Figure 4 for the visual representation of the PSO algorithm's flowchart.

Fruit-Fly Optimization Algorithm

The FOA is an optimization algorithm inspired by the behavior of fruit flies that was proposed by Xing et al (2014). Mathematically, the FOA involves seven steps, illustrated in Figures 5 and 6. In this study, the FOA was employed to optimize the ANN model and predict AOP. The objective was to assess its performance in optimizing a ANN and compare it to the proposed RWGWO–ANN model.

Single-Based Genetic Algorithm

The SGA, also preferred as mutation-based genetic algorithm, was proposed by Falco et al.

The pseudo-code of the RWGWO algorithm:

1 Initialize the grey wolf population size xi, iterations

2 Define the encircling prey parameters using the following equations:

$$b = 2 - 2\left(\frac{t}{t_{\max}}\right) \tag{4}$$

$$\mu = 2 \cdot b \cdot r_1 - b \tag{5}$$

$$c = 2 \cdot r_2 \tag{6}$$

r_1, r_2 are the random numbers in the range of 0 and 1; $t_{\max}$ is the maximum number of iterations.

3 Start the algorithm with iteration = 1

4 Evaluate the fitness of each wolf

5 Select the best leaders α, β, δ

6 While iteration t < $t_{\max}$

Evaluate the fitness of each wolf

  For each leader wolf

    Find new position y_i of the leaders x_i by random walk

    If $f(y_i) < f(x_i)$

    Update the leaders

  End

  For each ω wolf

    Update the position of the wolf using the Eq. (7) and apply greedy approach between the current and updated positions.

$$X_{t+1} = \frac{\left(X_1^{'} + X_2^{'} + X_3^{'}\right)}{3} \tag{7}$$

    ($X_1^{'}, X_2^{'}, X_3^{'}$ are the updated positions of the α, β, δ wolves)

  End

7 Update the encircling prey parameters

8 Update the fitness of leaders

9 End

Figure 3. Pseudo-code of the RWGWO algorithm.

(2002) to introduce the undervalued role of mutation in the realm of evolutionary computation, especially focusing on mutations beyond the traditional point mutation. In this algorithm, the nature-inspired mutation operators, namely frame-shift and translocation, were introduced and investigated. Also, their effectiveness in solving test functions was assessed.

As a matter of fact, the SGA was widely applied in various domains, such as structural optimization (Azad et al., 2012), solving the processor configuration problems (Lau & Tsang, 1997), optimization of neural networks (Nadi et al., 2009), optimization of

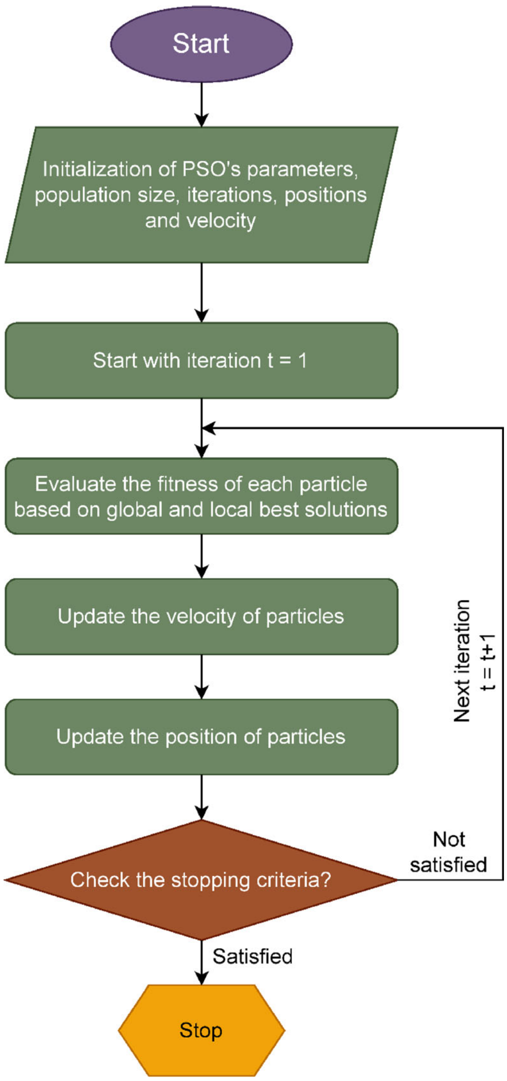


Figure 4. The PSO algorithm flowchart.

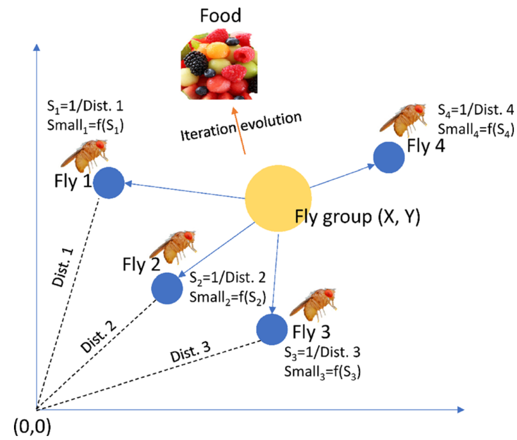


Figure 5. Illustration of the FOA mechanism.

nuclear fusion devices (Gómez-Iglesias et al., 2009), to name a few. However, it has not been applied to optimize a neural network or a machine learning algorithm for predicting AOP in open-pit mines. Therefore, in this study, we introduced this algorithm and used it to optimize the two-layer ANN model for predicting AOP. The pseudo-code of the SGA is presented in Figure 7.

Optimization of AOP Prediction Models

In this study, the ANN model is the key model that was used to predict AOP based on dependent variables. Normally, gradient descend (GD)-based algorithms can be used to train neural networks, including ANN and ANN models, in which back-propagation algorithm is a key component of GD for training neural networks. It computes the gradients of the objective function with respect to the weights and biases of the network by propagating the errors backwards through the layers of the network (Alzubaidi et al., 2021). This allows for updating the weights and biases in a way that reduces the error in the AOP predictions.

However, back-propagation algorithm is susceptible to getting trapped in local optima (Pedram et al., 2022). Because it relies on gradient information to update the model parameters, it can converge to a suboptimal solution if it gets stuck in a local optimal point in the optimization landscape (Bai et al., 2023). Moreover, it is sensitive to the initialization of the weights and biases in the neural network (Sreejith et al., 2020). Starting with poor initial values can result in the network converging to suboptimal solutions or not converging at all. Finding good initial values for the weights and biases can be challenging, especially for deeper networks.

Because prediction of AOP in open-pit mines is a complex, nonlinear and noisy problem, metaheuristic algorithms may be a potential solution to overcome the limitations of the back-propagation algorithm. Indeed, many metaheuristic algorithms were demonstrated to provide promising results, as reviewed above. Metaheuristic algorithms, with their global exploration capabilities, can potentially overcome local optima and search for better solutions across the entire search space. In addition,

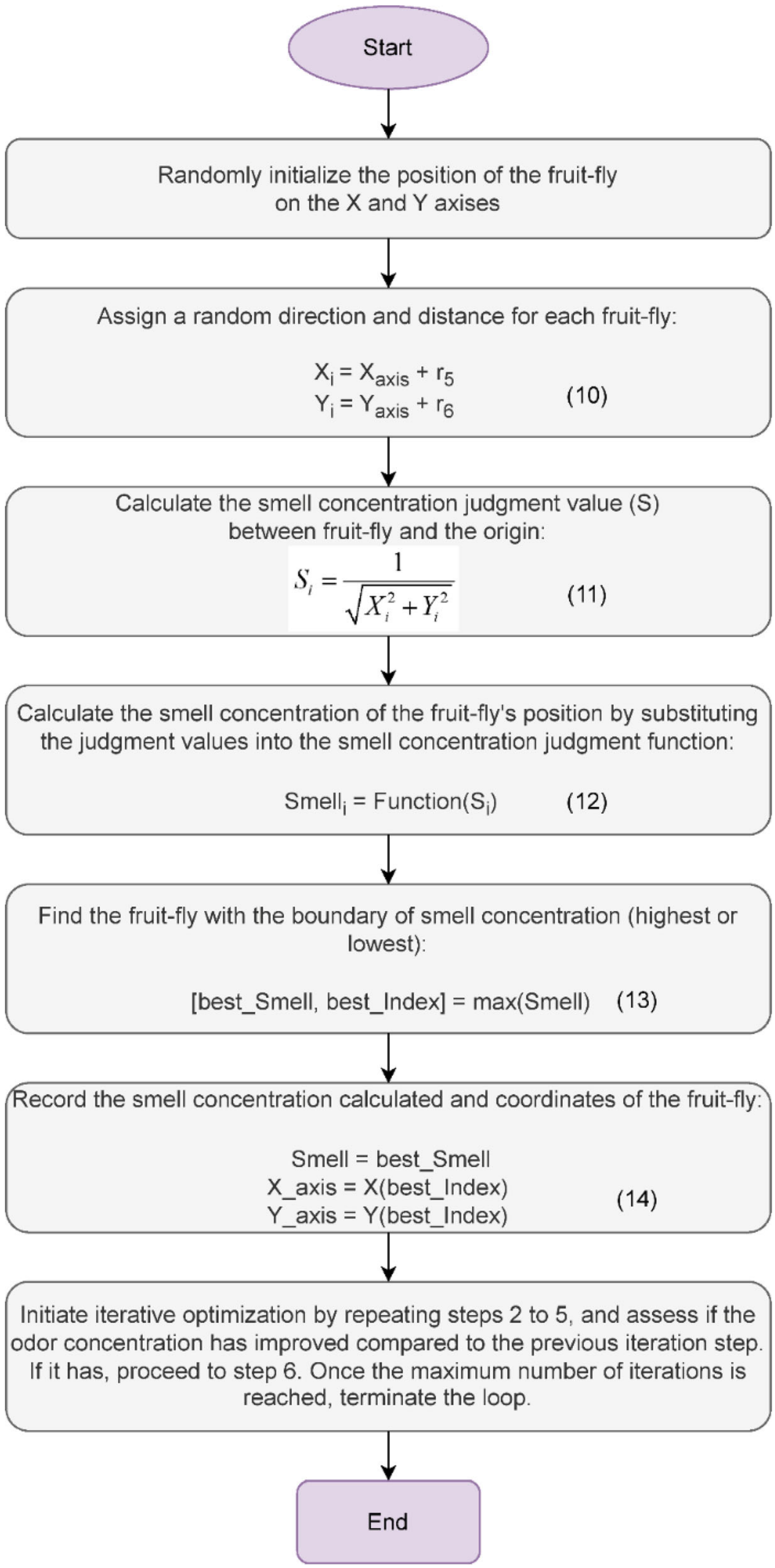


Figure 6. Steps to conduct an optimization problem using the FOA.

The pseudo-code of the SGA algorithm:

```
Begin
        Choose the mutation operator (either frame-shift or translocation)
        Randomly initialize a population P(t) or P elements
        While (a termination criterion is not fulfilled) do
        Evaluate P(t) by using fitness function
        For i = 1 to P do
                Randomly choose one individual in P(t) with tournament
                Apply the mutation operator chosen to it
        End for
        End while
        Update the variables for termination
End
```

Figure 7. Pseudo-code of the SGA algorithm (Falco et al., 2002).

metaheuristic algorithms often have mechanisms to explore the search space more effectively, reducing sensitivity to initialization. Therefore, in this study, we applied four different metaheuristic algorithms for training and optimizing the ANN model in predicting AOP, including RWGWO, PSO, FOA, and SGA. These algorithms have different search strategies, parameters, and strengths. Of those, the RWGWO is an improved algorithm that is combined with the random walk technique.

To enhance the ANN model through the utilization of the RWGWO algorithm, an initial population of potential solutions is established. This process includes the random initialization of the weights and biases within the ANN model. Each potential solution embodies a distinct arrangement of weights and biases, aligned with the architecture and specifications of the ANN. Subsequently, a fitness function (i.e., RMSE or MSE) is selected to evaluate the fitness of each candidate solution in the created population. The fitness value indicates how well a particular candidate solution performs on the optimization objective. To search for better solutions, the optimization process of the RWGWO is performed through iterations, including the selection, variation, evaluation, update, and termination steps. The objective is to continue the iterations until the stopping condition is satisfied. This condition could be either reaching the maximum iteration count or attaining a specific fitness level. When the RWGWO algorithm concludes, the optimal collection of weights and biases for the ANN model is embodied by the best solution identified within the last population. It should be noted that the remaining metaheuristic algorithms, i.e., PSO, FOA, and SGA, also follow the outlined steps of the RWGWO algorithm in optimizing ANN model. The optimization mechanism of the RWGWO–ANN model for predicting blast-induced AOP is presented in Figure 8. The optimization mechanisms of the other hybrid models (i.e., PSO–ANN, SGA–ANN and FOA–ANN) are similar to that of the RWGWO–ANN model. However, the main difference depends on the specific optimization algorithms employed in each model.

EXPERIMENTAL APPLICATION

Study Area

To showcase the effectiveness of the RWGWO–ANN model in forecasting AOP resulting from blasting activities in open-pit mines, we opted for the Coc Sau open-pit coal mine in Northern Vietnam as a representative example (Fig. 9). This mine, a significant player in Vietnam's open-pit coal mining sector, boasts an annual production rate of 2.7 million tons, while currently operating at a depth of − 300 m relative to sea level.

The geological conditions in this region are notably intricate, with rock hardness falling within

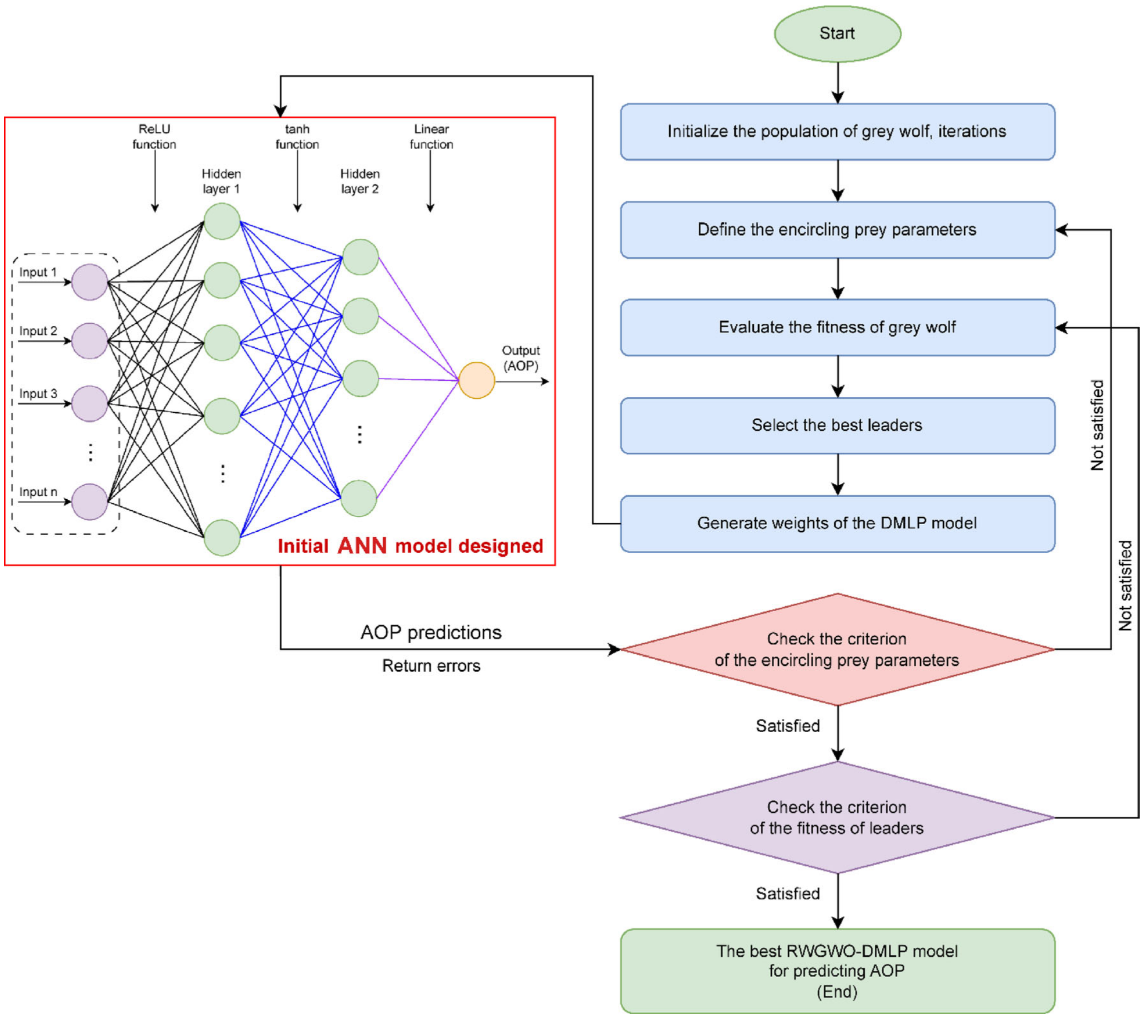


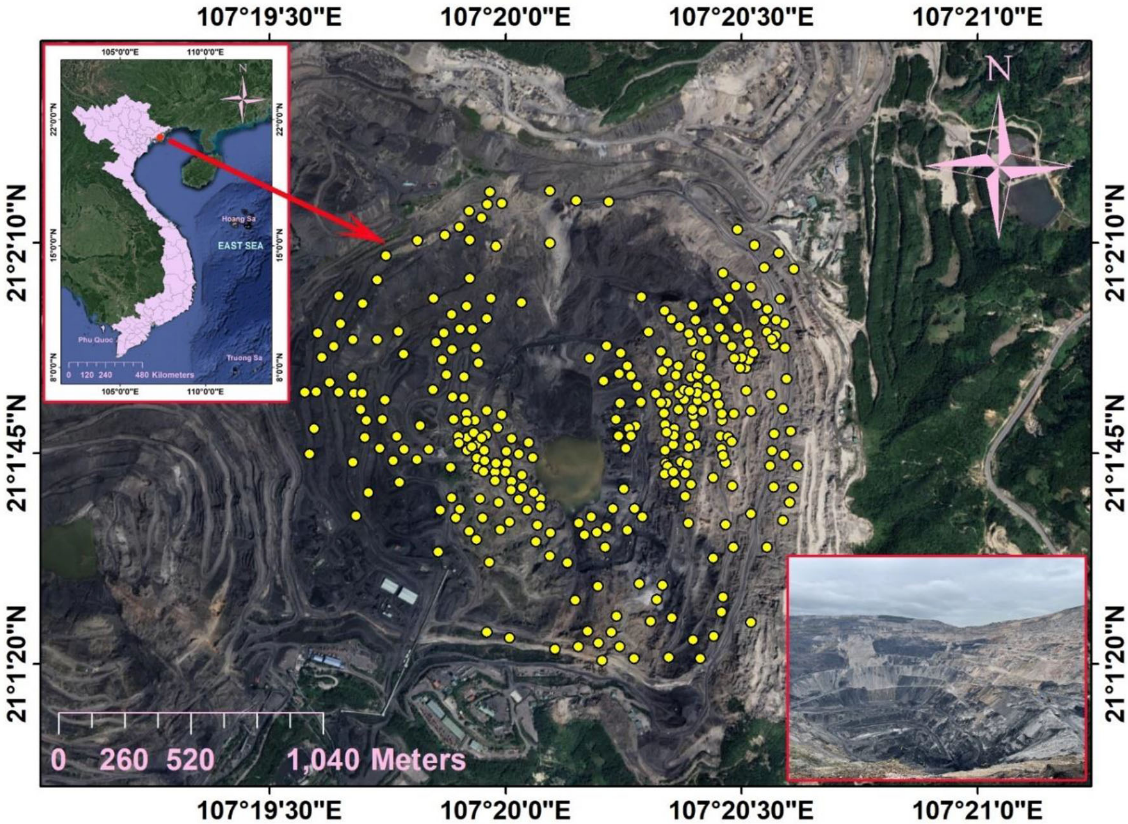
Figure 8. The RWGWO–ANN framework proposed for predicting AOP.

the 11 to 14 range as per the rock hardness grading by Protodiakonov et al. (1964). As a result, employing blasting becomes a pivotal technique for effectively fragmenting rocks prior to their subsequent loading and transportation to waste repositories. Moreover, the Coc Sau open-pit coal mine is surrounded by residential zones at distances ranging from 600 to 700 m, while other open-pit coal mines such as Deo Nai and Cao Son are situated at distances spanning 400 to 500 m (Fig. 9b). Consequently, the adverse repercussions of the blasting operations carried out at the Coc Sau open-pit coal mine hold notable significance.

Data Collection, Preparation and Analysis

Original Dataset

In mine blasting, there are several parameters that can affect the intensity of ground vibration, including blasting parameters and geological/geographical conditions. Among these, blasting parameters are categorized as controllable variables, as they are designed by engineers and can be adjusted in the blast pattern design. Geological and geographical conditions are classified as uncontrollable factors, because their properties cannot be altered due to their inherent uncertainty. Thus, this study



(a)

(b)

Figure 9. (**a**) Location of the Coc Sau open-pit coal mine and blasting events. (**b**) Satellite image of the Coc Sau open-pit coal mine (Google Earth) and surroundings.

Figure 10. Micromate device and AOP measurement by the microphone.

focused on blasting parameters and their relationships with induced AOP.

To gather the dataset for this study, parameters like B (burden), S (spacing), f (rock hardness), PF (powder factor), and Q (maximum explosive charge per delay) were extracted from blast patterns created by mine engineers. The D parameter was calculated based on the blast site's location and the position of the placed Micromate device. AOP values were monitored using the Micromate, produced by Instantel—Canada, as depicted in Figure 10, and the properties of the dataset exhibited weak correlations (Fig. 11). This observation signifies a notable degree of independence among the input variables. Moreover, these weak correlations highlight that each variable can offer distinct insights to the analysis, enhancing the efficacy of prediction models. Hence, all the variables were utilized to formulate the AOP predictive models in this study.

Data Enrichment by ExTree model

As previously introduced, a notable feature of this study is the utilization of the ExTree model for data enrichment prior to the development of AOP predictive models. The ExTree model was crafted using the original dataset, and subsequently, the outcome predictions generated by the ExTree model were incorporated as supplementary input variables. The ExTree model demonstrated a remarkable 93.2% accuracy in explaining the relationships between the input variables and AOP within the original dataset, as listed in Supplementary Table S1. Ultimately, the enriched dataset encompassed seven input variables, and their interrelationships are illustrated in Figure 12, which reveals a robust correlation between the supplementary input variable (ExTree) and S. To tackle the issues of multicollinearity and reduced model interpretability, it was necessary to eliminate one of these variables. Given that the new supplementary input variable, ExTree, possesses the capability to elucidate the relationships among the other variables, the decision was made to remove the S variable from this enriched dataset. Consequently, the enriched dataset comprised solely six input parameters: B, f, PF, Q, D, and ExTree (Fig. 13). This enriched dataset was used to develop predictive models, which were compared to the models developed based on the original dataset.

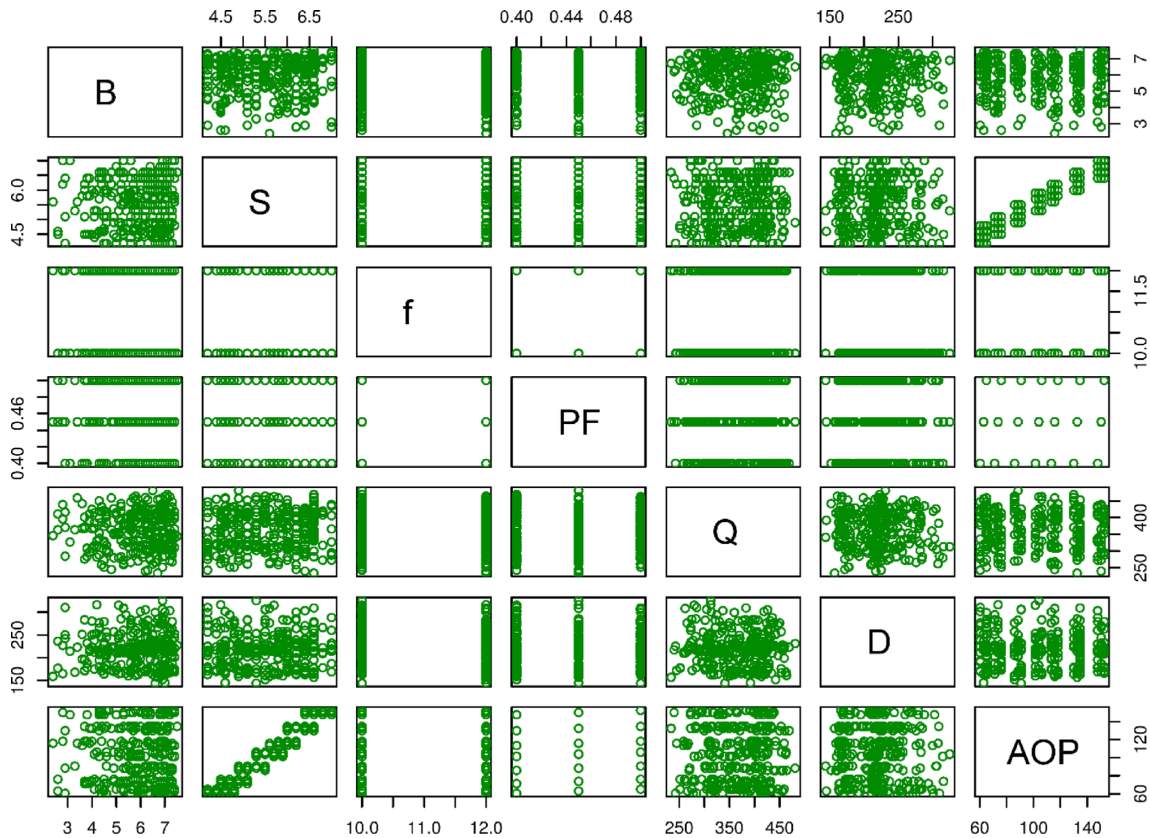


Figure 11. Properties of the dataset collected (original dataset).

RESULTS AND DISCUSSIONS

As previously mentioned, the primary model utilized in this study was an ANN model. However, we delved into a more intricate architecture, aiming to enhance the performance of the predictive model. Consequently, it became imperative to formulate the network topology prior to the construction of the hybrid models intended for forecasting AOP within the scope of this investigation.

Design of Topology Network of ANN Model

The topology network of the ANN model for predicting AOP was devised using a trial and error approach. The range of hidden layers explored was from 1 to 3, aiming to assess the performance of the ANN model with a more intricate structure (comprising multiple hidden layers) in predicting AOP. Both the original dataset and the enriched dataset employed the same topology network during the development of predictive models. This facilitated the comprehensive comparison and evaluation of models built upon the original dataset and those constructed using the enriched dataset. To accomplish this objective, an ANN model was utilized, employing the Adam training algorithm along with the mean squared error (MSE) fitness function over the course of 500 epochs. Diverse numbers of hidden layers, spanning from 1 to 3, were experimented with. The outcomes demonstrated that the optimal number of hidden layers was 2 (Figs. 14 and 15).

Once the appropriate number of hidden layers was determined, the task involved designing and calculating the number of neurons (nodes) within each of these hidden layers. The same techniques were employed for this stage, and a range of different quantities of hidden nodes, varying from 6 to 20, were subjected to experimentation. With the decision to employ 2 hidden layers, each of these layers comprised a designated number of hidden nodes. Figure 14 illustrates that the two most effective configurations for hidden node layers were

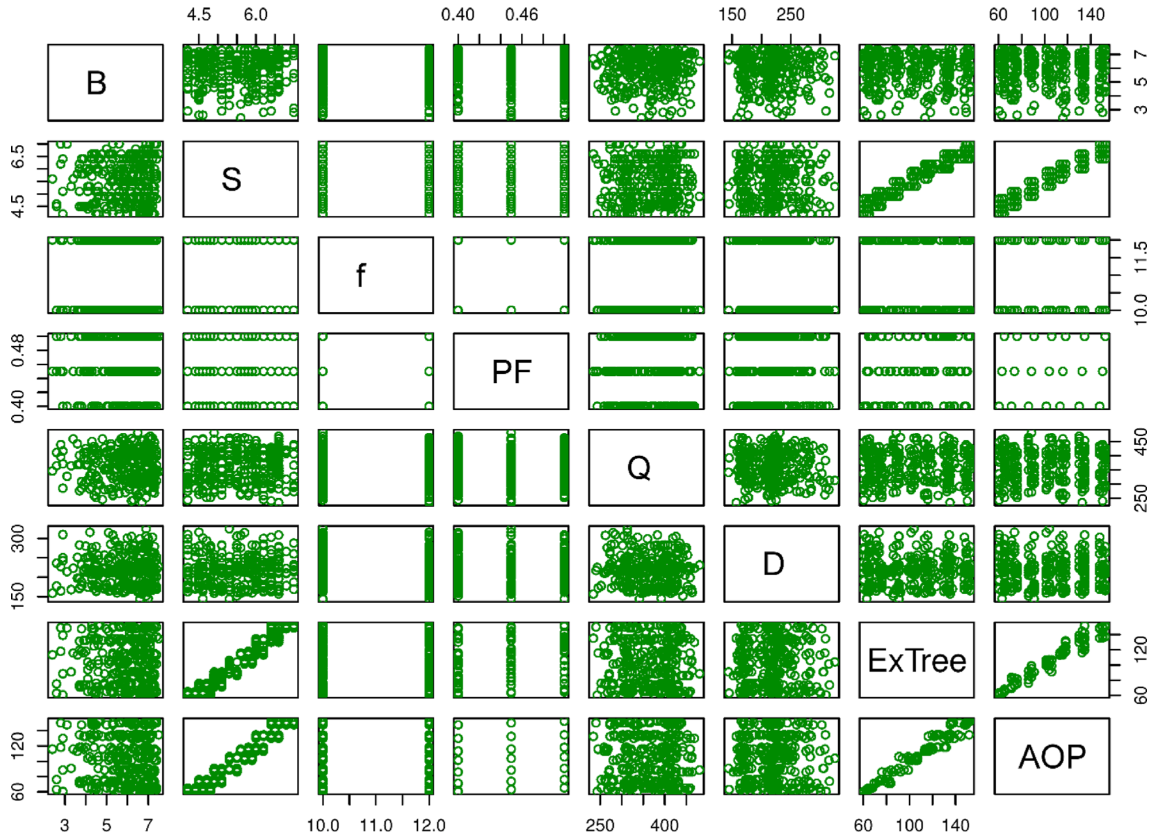


Figure 12. The enriched dataset after applying the Extra Trees algorithm.

15 and 6. Consequently, the initial hidden layer consisted of 15 nodes, while the subsequent hidden layer was composed of 6 nodes. This culminated in the establishment of the optimal structure for the ANN model, meticulously chosen for the purpose of predicting AOP within the context of this study (Fig. 16).

Development of AOP Prediction Models based on the Original Dataset

Once the structure of the ANN was appropriately designed as detailed above, the focus shifted to developing AOP predictive models. The initial development utilized the original dataset, which was subsequently divided into two random portions. The training set comprised 70% of the entire dataset for model training, while the remaining 30% served as a testing set to assess the developed models. To mitigate potential overfitting, a combination of tenfold cross-validation and MinMax scaling techniques were employed. Notably, due to the randomness introduced by predictive models during each iteration of the tenfold cross-validation, this procedure was repeated three times to ensure the objectivity and consistency of the models.

Moreover, as discussed in the preceding section, a technique for enhancing the accuracy of the ANN model involved the use of various activation functions. Illustrated in Figure 2, the ANN model consisted of two hidden layers. Consequently, the "ReLU" activation function was applied to the first layer, while the "tanh" activation function was implemented in the second layer. The evaluation of the ANN model's performance during each iteration of training utilized the RMSE as the chosen loss function.

Subsequent to this, the ANN model underwent training via distinct metaheuristic algorithms (referred to as hybrid models) and the conventional stochastic gradient descent algorithm (conventional model). This training was executed within the framework presented in Figure 8. It is important to

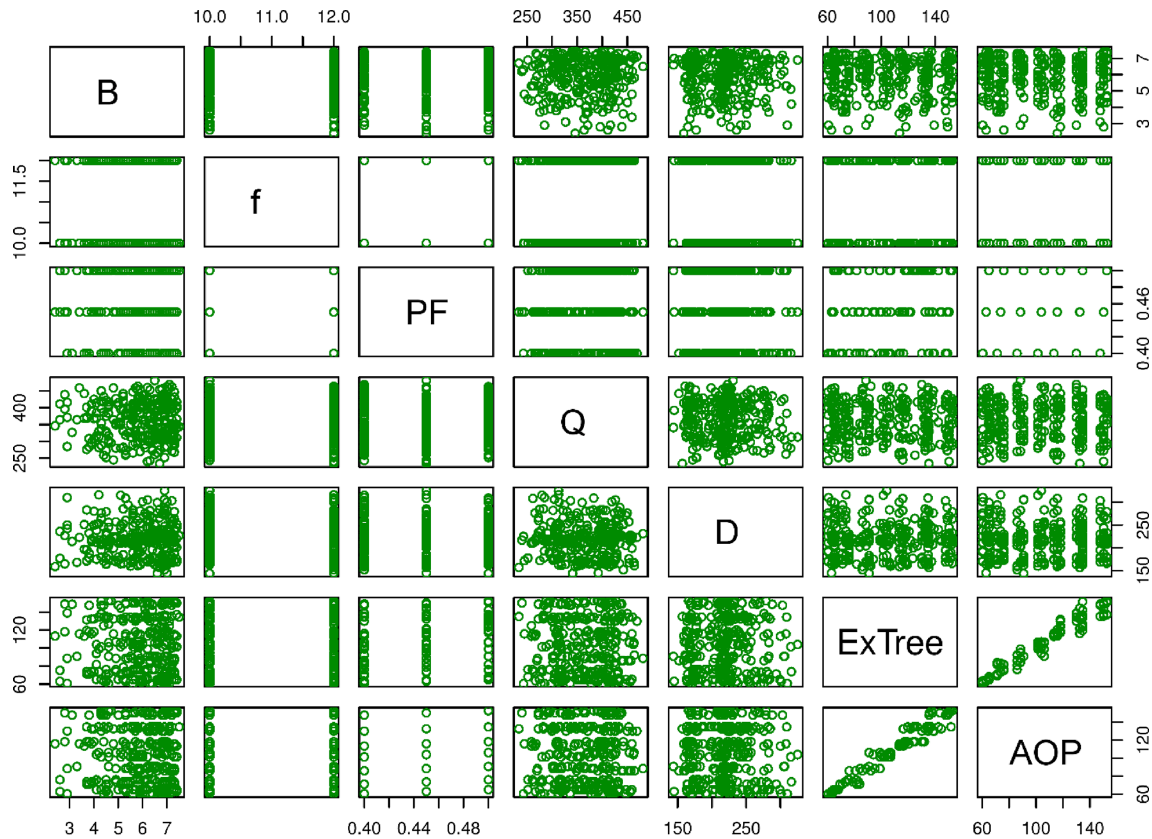


Figure 13. The final enriched dataset after analysis.

note that before commencing the training of the ANN model with the metaheuristic algorithms, proper configuration of the algorithm parameters was a necessary step, which is outlined as follows:

- The SGA algorithm's parameters:

- Crossover probability: 0.75
- Mutation probability: 0.1
- Selection: ''Roulette''
- Crossover: ''Uniform''
- Mutation: ''Swap''
- Population size: 450
- Epoch: 500
- The PSO algorithm's parameters:

- Local coefficient: 1.2
- Global coefficient: 1.2
- Minimum of weight: 0.4
- Maximum of weight: 0.9
- Population size: 450
- Epoch: 500

- The FOA algorithm's parameters:

- Population size: 450
- Epoch: 500
- The RWGWO algorithm's parameters:
- Population size: 450
- Epoch: 500

The established parameters reveal a distinction between the SGA and PSO algorithms, which fall under the category of parametric algorithms, and the FOA and RWGWO algorithms, which are categorized as nonparametric algorithms. Stated differently, the FOA and RWGWO algorithms exhibit simplicity compared to the more intricate PSO and SGA algorithms. Subsequently, employing the framework illustrated in Figure 8, the ANN model's training was conducted using the SGA, PSO, FOA, and RWGWO algorithms. The optimization outcomes for the SGA–ANN, PSO–ANN, FOA–ANN,

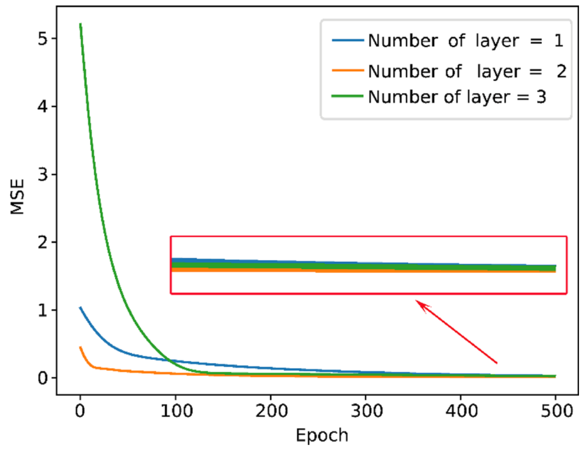


Figure 14. Evaluating the ANN model's performance in predicting AOP using varying numbers of hidden layers.

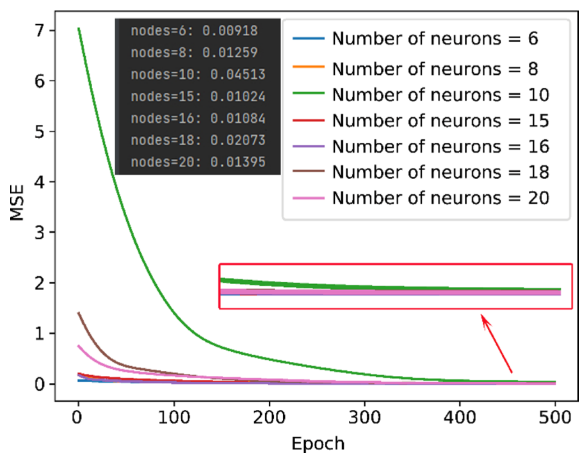


Figure 15. The ANN model's performance in predicting AOP using varying numbers of hidden nodes.

and RWGWO–ANN models on the original dataset are illustrated in Figures 17 and 18.

Development of AOP Prediction Models Based on Enriched Dataset

Regarding the development of predictive models based on the enriched dataset for predicting AOP, the same techniques and parameters as those used for the models developed based on the original dataset were applied. The main difference was the use of the enriched dataset instead of the original dataset. The optimization outcomes for the SGA–

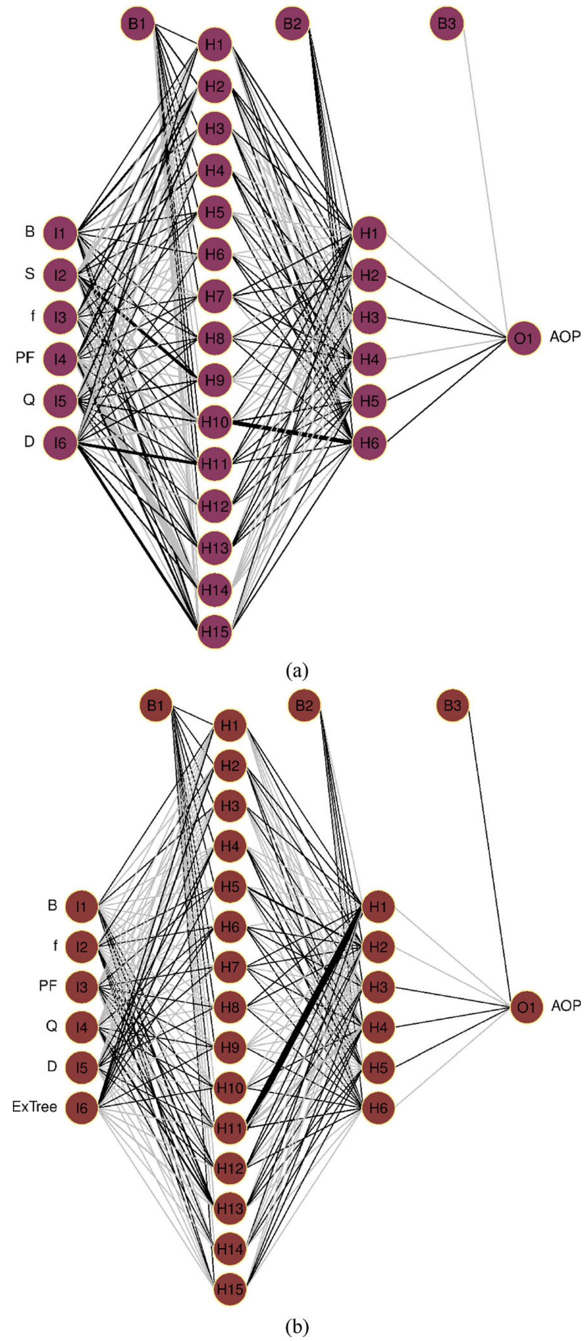


Figure 16. The network topology of the ANN model crafted for AOP prediction in this research. (**a**) The topology network for the original dataset; (**b**) The topology network for the enriched dataset.

ANN, PSO–ANN, FOA–ANN, and RWGWO–ANN models on the enriched dataset are illustrated in Figures 19 and 20.

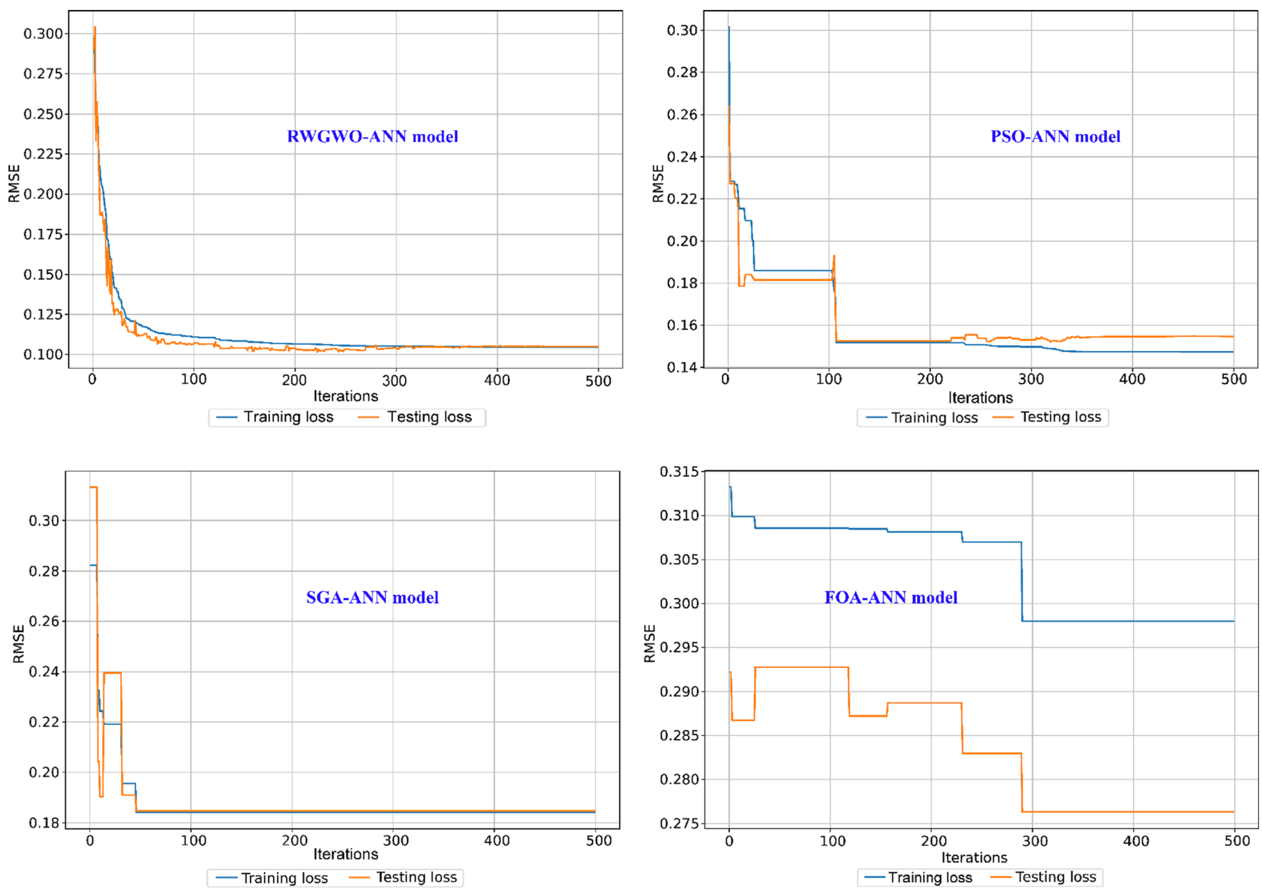


Figure 17. Training and testing loss curves of the optimization-based ANN models for predicting AOP on the original dataset.

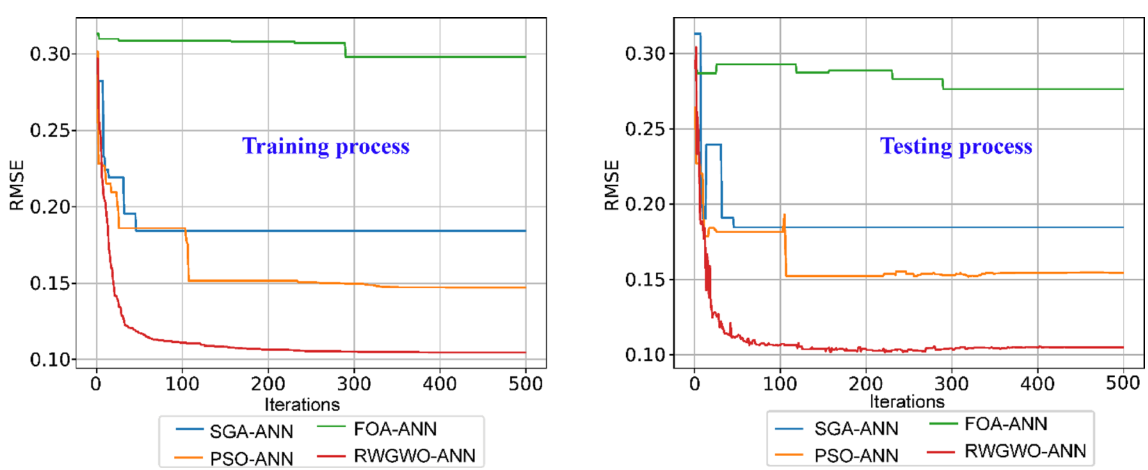


Figure 18. Comparison of the hybrid models for predicting AOP through the fitness function on the original dataset.

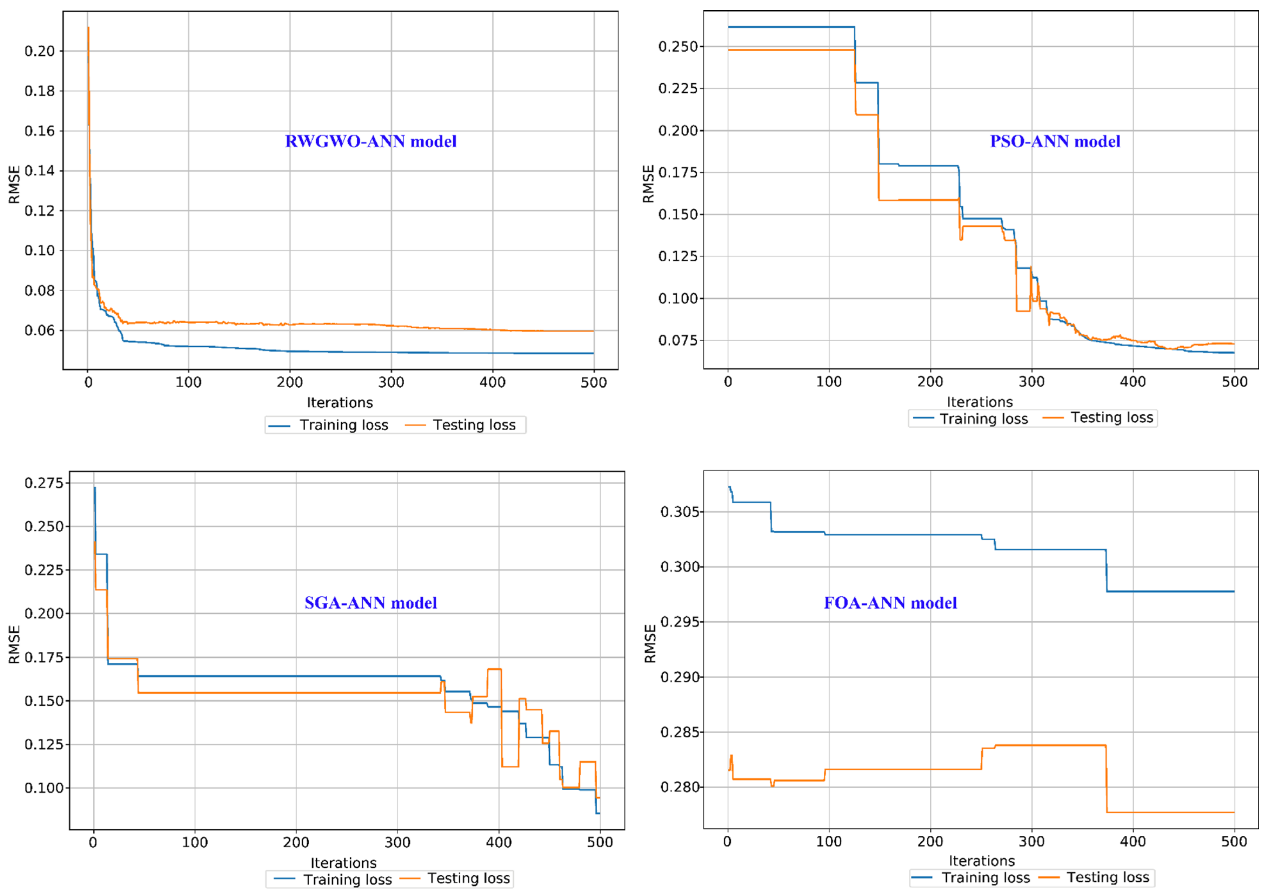


Figure 19. Training and testing loss curves of the optimization-based ANN models for predicting AOP on the enriched dataset.

Development of the empirical model (USBM)

In an effort to establish a comparison with the ANN and hybrid optimization-based ANN models, the USBM empirical equation (Eq. 15) was employed to estimate AOP resulting from blasting activities in open-pit mines, thus:

$$\mathrm{AOP} = k(\mathrm{SD})^{-\beta} \tag{15}$$

where *SD* is the distance scale that is calculated based on the relationship between D and Q parameters, thus:

$$\mathrm{SD} = \mathrm{DQ}^{-0.33} \tag{16}$$

where k and β are site coefficients and computed by multiple regression analysis. The site coefficients vary based on the different conditions of each mine.

In constructing the USBM empirical model to predict AOP, both the training and testing datasets employed were identical to those utilized in the development of the AI-based models. However, given that the USBM empirical model relies exclusively on two variables (i.e., D and Q), there was no distinction between the USBM empirical model constructed using either the original dataset or the enriched dataset. After analyzing the training dataset by multiple regression analysis, the site coefficients k and β were determined, as shown in the empirical Eq. (17). The convergence of the USBM empirical model for predicting AOP on the training dataset based on the distance scale is shown in Figure 21.

$$\mathrm{AOP} = 117.879(\mathrm{SD})^{-0.055} \tag{17}$$

Observing the outcomes presented in Figure 21, it became evident that the predicted AOP and scaled distance exhibited strong correlation with power function, yielding a $R^2$of 1. This suggests that the constructed USBM empirical model aligns well

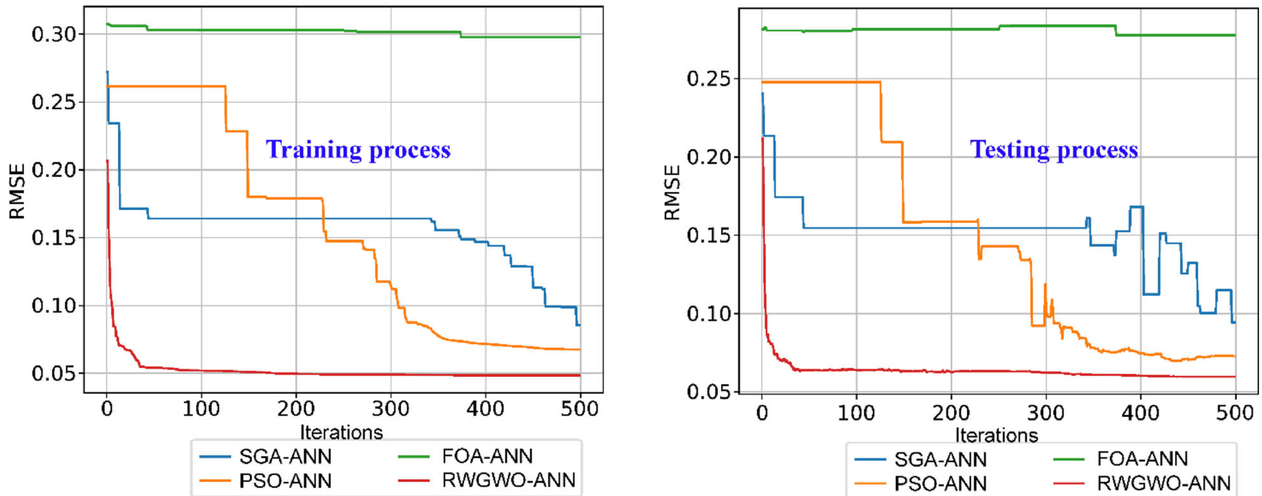


Figure 20. Comparison of the hybrid models for predicting AOP through the fitness function on the enriched dataset.

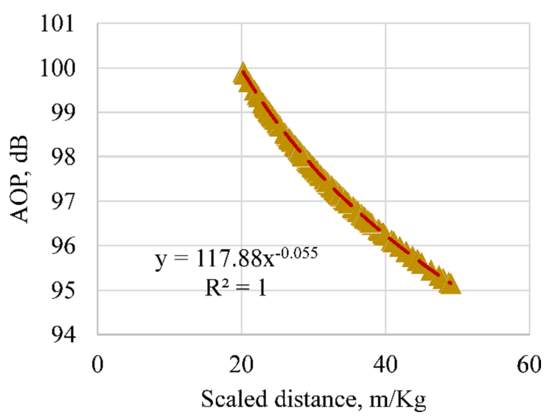


Figure 21. The USBM model's convergence for forecasting AOP.

with the scaled distance parameter. Nonetheless, a more comprehensive assessment is required to gauge the model's precision in forecasting AOP.

Evaluation and Discussion

Once all predictive models, encompassing AI-based and empirical approaches, were developed, their performances were evaluated using both training and testing datasets. This assessment aimed to determine the adequacy of their performance and to identify potential occurrences of overfitting. Illustrated in Figure 17, the training and testing curves of the RWGWO–ANN, PSO–ANN, and SGA–ANN models displayed remarkable performance on the original dataset. Their errors exhibited similarity, indicating appreciable convergence and absence of overfitting issues. Among these models, the RWGWO–ANN variant demonstrated the most favorable performance curves. Conversely, the FOA–ANN model exhibited less desirable performance curves, with noticeable dissimilarity between its training and testing curves for the original dataset. Despite this, the error disparity between its training and testing phases remained relatively modest, suggesting absence of overfitting. The comparison extended to Figure 18 gives evidence that the RWGWO–ANN model exhibited the lowest error among the four hybrid models developed in this study for the original dataset.

Similarly, examining the outcomes on the enriched dataset, as depicted in Figures 19 and 20, highlighted the superior performance curves of the RWGWO–ANN, PSO–ANN, and SGA–ANN models compared to the FOA–ANN model. These models demonstrated convergent performance curves. The FOA–ANN model's curves, conversely, exhibited greater disparity, akin to the trends observed in the original dataset analysis. Moreover, Figure 20 underscored the continued prominence of the RWGWO–ANN model in terms of accuracy when predicting AOP on the enriched dataset. Figure 22 presents a comprehensive comparison of the variations between the predicted AOP values from various AI-based models and the USBM empirical

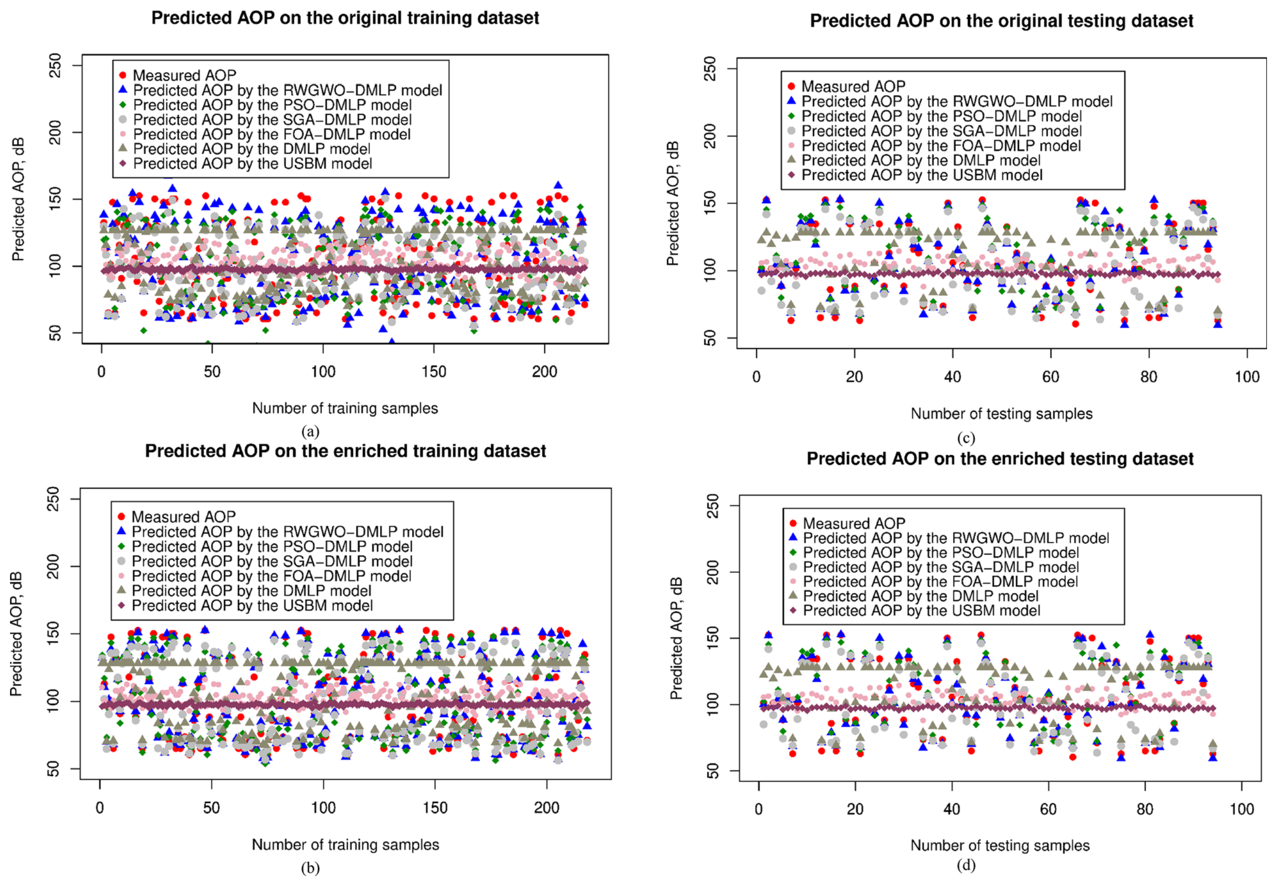


Figure 22. Comparison of the measured AOP values and predicted AOP values by different AI-based models and USBM empirical model with respect to the: (**a**) original training dataset; (**b**) enriched training dataset; (**c**) original testing dataset; and (**d**) enriched testing dataset.

model, both on the original dataset and the enriched dataset.

Illustrated in Figure 22, the RWGWO–ANN model's predicted AOP values closely align with the measured AOP values, outperforming the predictions by other models with respect to the original dataset, encompassing both training and testing data. Subsequently, the AOP values projected by the ANN model are presented. Notably, these values exhibited higher accuracy than those generated by the PSO–ANN, SGA–ANN, and FOA–ANN models for the original dataset. This discrepancy suggests that the PSO, SGA, and FOA algorithms are comparatively weak compared to the conventional SGD algorithm in training the ANN model using the initial AOP dataset. In contrast, the USBM empirical model showcased the worst predictive accuracy in this study although the relationship between the scaled distance and AOP of this model was 1 (Fig. 21).

Meanwhile, turning attention to the AI-based models constructed using the enriched dataset, remarkable results come to the fore in Figure 22. Most of the hybrid AI-based models demonstrated heightened accuracy, barring the FOA–ANN model. Compared to the individual ANN model, the RWGWO–ANN, PSO–ANN, and SGA–ANN models yielded more precise AOP predictions with the enriched dataset. Moreover, when measured against their AOP predictions on the original dataset, a substantial enhancement in accuracy was evident for the RWGWO–ANN, PSO–ANN, and SGA–ANN models. This underscores the pivotal role played by the ExTree model in data enrichment, profoundly ameliorating the accuracy of AI-based models (namely, RWGWO–ANN, PSO–ANN, and SGA–ANN) alongside fortifying the potency of metaheuristic algorithms.

To further compare these hybrid models across both the original and enriched datasets, and to assess

Table 1. Performances of AI-based and empirical models for AOP prediction based on the original and enriched datasets

Model	Training dataset				Testing dataset			
	MAE	RMSE	R^2	VAF	MAE	RMSE	R^2	VAF
Original dataset								
SGA–ANN	13.391	16.957	0.688	70.619	13.094	17.001	0.632	68.531
PSO–ANN	10.660	13.565	0.800	80.048	11.465	14.241	0.741	74.416
FOA–ANN	24.107	27.445	0.183	18.602	21.807	25.447	0.175	17.981
RWGWO–ANN	8.292	9.634	0.899	89.939	8.173	9.666	0.881	89.079
ANN	11.585	13.692	0.797	80.241	10.229	12.402	0.804	80.528
USBM	27.184	30.700	−0.022	77.933	25.148	29.838	−0.092	67.130
Enriched dataset								
SGA–ANN	6.153	7.877	0.933	94.068	7.193	8.693	0.904	92.038
PSO–ANN	5.108	6.229	0.958	95.793	5.478	6.713	0.943	94.284
FOA–ANN	24.239	27.423	0.185	18.517	21.904	25.578	0.166	17.948
RWGWO–ANN	3.075	4.472	0.978	97.833	3.860	5.499	0.961	96.186
ANN	11.928	13.453	0.804	82.512	12.739	14.438	0.734	76.035
USBM	27.184	30.700	−0.022	77.933	25.148	29.838	−0.092	67.130

their advancements relative to the ANN and empirical models, four performance metrics—MAE, RMSE, R^2, and VAF (variance accounted for)—were computed using Eqs. (18–21), with the outcomes detailed in Table 1.

$$\mathrm{MAE} = \frac{1}{n_{\mathrm{blast}}} \sum_{\mathrm{blast}=1}^{n_{\mathrm{blast}}} \left| \mathrm{AOP}_i - \widehat{\mathrm{AOP}}_i \right| \tag{18}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n_{\mathrm{blast}}} \sum_{\mathrm{blast}=1}^{n_{\mathrm{blast}}} \left(\mathrm{AOP}_i - \widehat{\mathrm{AOP}}_i \right)^2} \tag{19}$$

$$R^2 = 1 - \frac{\sum_{\mathrm{blast}=1}^{n_{\mathrm{blast}}} \left(\mathrm{AOP}_i - \widehat{\mathrm{AOP}}_i \right)^2}{\sum_{\mathrm{blast}=1}^{n_{\mathrm{blast}}} \left(\mathrm{AOP}_i - \overline{\mathrm{AOP}}_i \right)^2} \tag{20}$$

$$VAF = \left(1 - \frac{\mathrm{var}\left(\mathrm{AOP_i} - \widehat{\mathrm{AOP}}_\mathrm{i} \right)}{\mathrm{var}\left(\widehat{AOP}_i \right)} \right) \times 100 \tag{21}$$

where n_{blast} represents the overall count of datasets employed; AOP_i, $\widehat{\mathrm{AOP}}_i$, and $\overline{\mathrm{AOP}}_i$ symbolize the measured AOP, predicted AOP, and the mean of measured AOP values, respectively.

Drawing from the performance metrics in Table 1, it is evident that the RWGWO–ANN model exceled not only on the original dataset but also on the enriched dataset. Remarkably, its performance improved by approximately 7% with the enriched dataset due to the ExTree model. The other models, such as PSO–ANN and SGA–ANN, also showed significant enhancements (ranging from 20 to 27%) when compared to their performance on the original dataset. In essence, the ExTree algorithm added valuable information to the original dataset, aiding the RWGWO, PSO, and SGA algorithms in exploring the nuances and value of the enriched dataset, resulting in improved accuracies during the training of the ANN model.

Clearly, the amalgamation of data enrichment (utilizing the ExTree algorithm) and metaheuristic algorithms, combined with an ANN model (two hidden layers), substantially enhanced the accuracy of the AOP predictive models in this study. Additionally, it appeared that the FOA algorithm did not align well with the AOP datasets in this study, even when enriched by the ExTree algorithm. In Figure 23, the correlations between measured AOP values and predicted AOP values were depicted for both the original and enriched datasets.

The analysis depicted in Figure 23 reveals that the RWGWO–ANN model exhibited the most favorable convergent outcomes for both the original and enriched datasets, outperforming the other models. Notably, the convergences achieved on the enriched dataset, for both training and testing phases, surpassed those obtained from the original dataset. This is evident from the tighter clustering of the grey circles and purple triangles in Figure 23, compared to the dispersion between the blue circles and red triangles. The subsequent models in consideration are the PSO–ANN and SGA–ANN

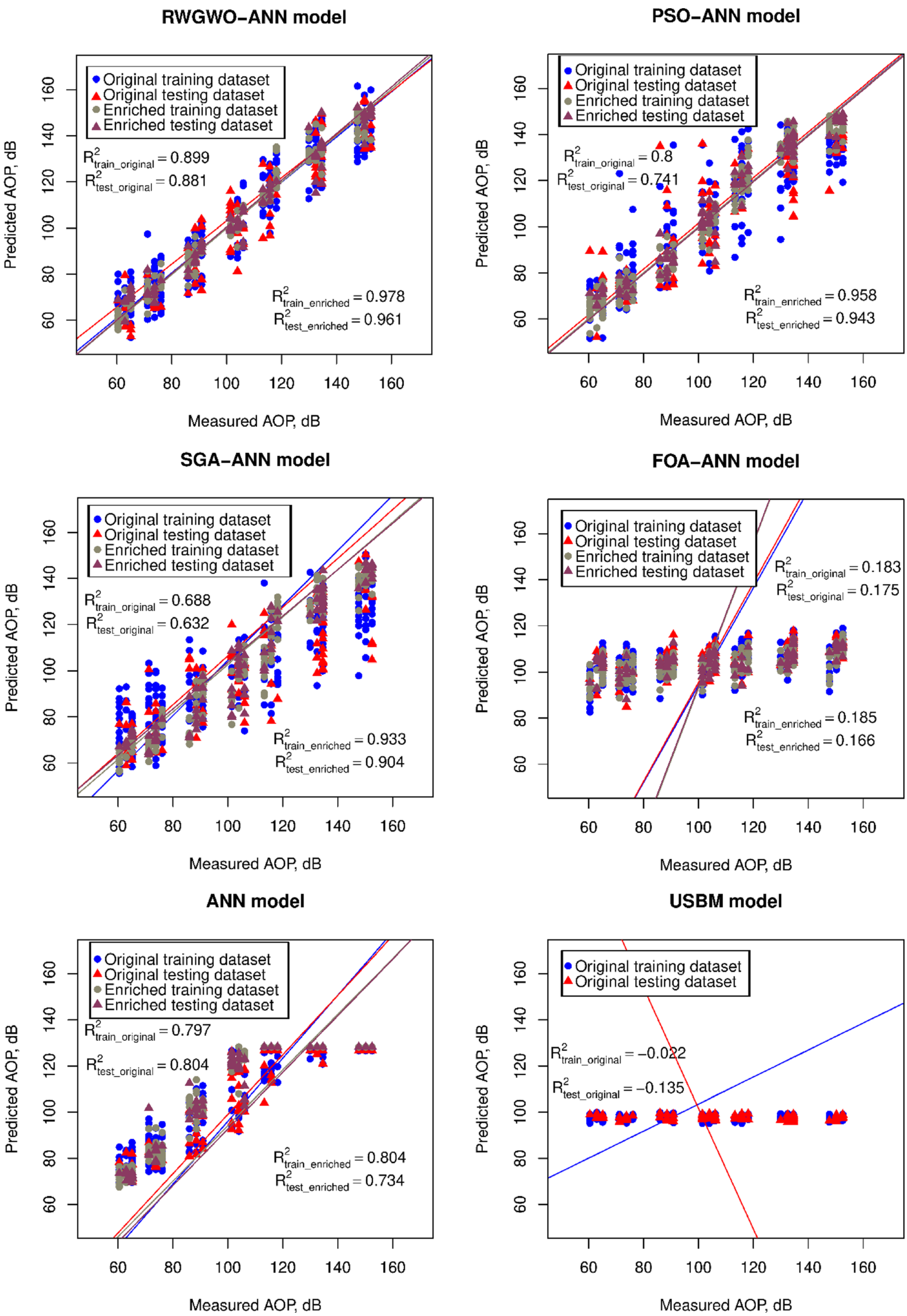


Figure 23. Relationship between measured AOP values and predicted AOP values through regression line fitting.

models. It is worth highlighting that the FOA–ANN model exhibited notably poor convergence. This model seemed to generate AOP predictions spanning a range of 80 to 120 dB, with data points showing minimal vertical displacement. Surprisingly, this performance even lagged behind that of the single ANN model.

In stark contrast, the USBM empirical model demonstrated the least favorable convergence, with majority of predicted AOP values falling within the range of 90 to 100 dB. This implies that the USBM empirical model, as indicated by the regression line fitting in Figure 23, lacked accuracy in predicting AOP in the context of this study. Consequently, its application for AOP prediction in this area is not recommended.

Further evaluations of the developed AI-based and USBM empirical models in this study for predicting AOP were made through the histogram of errors (Fig. 24). Based on the histogram of errors, we see the error range of each model and how each model fits with the prediction of AOP in practical engineering.

As depicted in Figure 24a, the distinction is clear: the RWGWO–ANN model applied to the enriched dataset displayed reduced errors (± 4 dB) in contrast with its counterpart developed using the original dataset (± 15 dB). Similarly, the PSO–ANN model applied to the enriched dataset exhibited errors within the range of ± 10 dB, the errors expanded to ± 15 dB for predictions of AOP on the original dataset. On the other hand, the SGA–ANN model showed marginally greater errors, spanning from − 20 to 35 dB on the original dataset and narrowing within − 10 and 15 dB on the enriched dataset.

It is worth highlighting that the error ranges for the FOA–ANN model were similar between the original and enriched datasets, both residing within the range of ± 40 dB. This deviation was notably higher than that observed in the previous models. Despite the conventional ANN model having smaller error ranges than the FOA–ANN model, its performance did not seem to improve on the enriched dataset compared to the original dataset, despite a distinct difference in error distribution. Lastly, the USBM model exhibited the most substantial errors, spanning from − 40 to 60 dB, and demonstrating a non-normal distribution.

In summary, a marked enhancement was observed in most of the developed AOP prediction models when applied to the enriched dataset, particularly notable for the RWGWO–ANN, PSO–ANN, and SGA–ANN models. Among these models, the RWGWO–ANN model stood out as the dominant choice for predicting AOP at the Coc Sau open-pit coal mine.

SENSITIVITY ANALYSIS

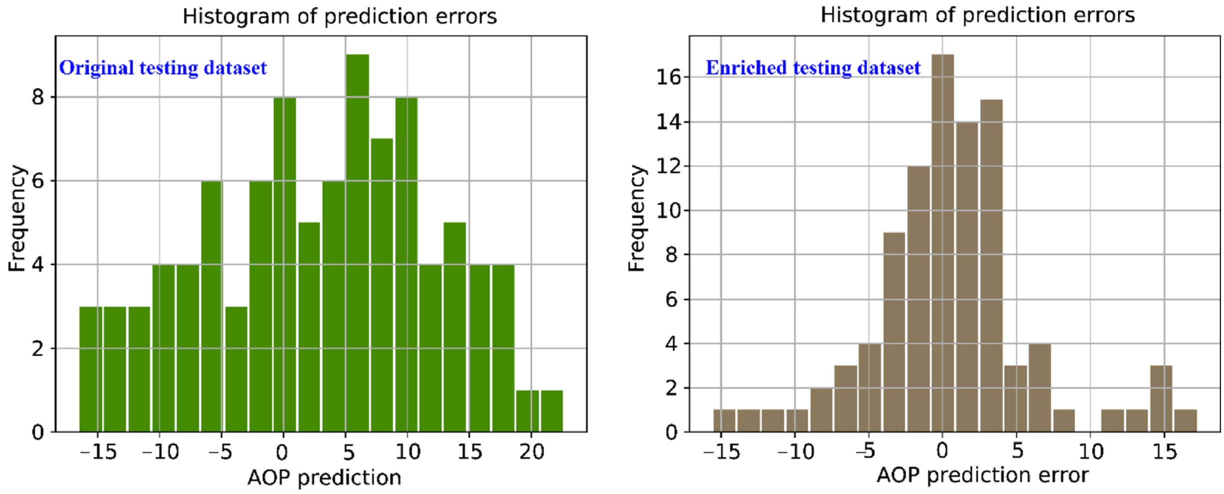
The RWGWO–ANN model was chosen as the optimal predictor for AOP in this study due to its strong performance on both the original and enriched datasets. To assess the significance of the input variables, a sensitivity analysis using the Morris method (Morris, 1991) was conducted (Fig. 25).

Figure 25 presents the sensitivity analysis of input variables using the chosen RWGWO–ANN model. This analysis was performed on datasets both with and without enrichment. Examination of the insights provided by Figure 25 unveiled the significance and individual contributions of each input variable throughout the RWGWO–ANN model's development. Notably, this model was identified as the most proficient for AOP prediction in this study. In the case of the original dataset, the spacing parameter (S) emerged as the most influential variable, closely followed by the burden (B) with slightly lower significance. Other inputs, namely f, PF, Q, and D, exhibited comparable and relatively lower importance levels compared to B and S in the construction of the RWGWO–ANN model for AOP prediction.

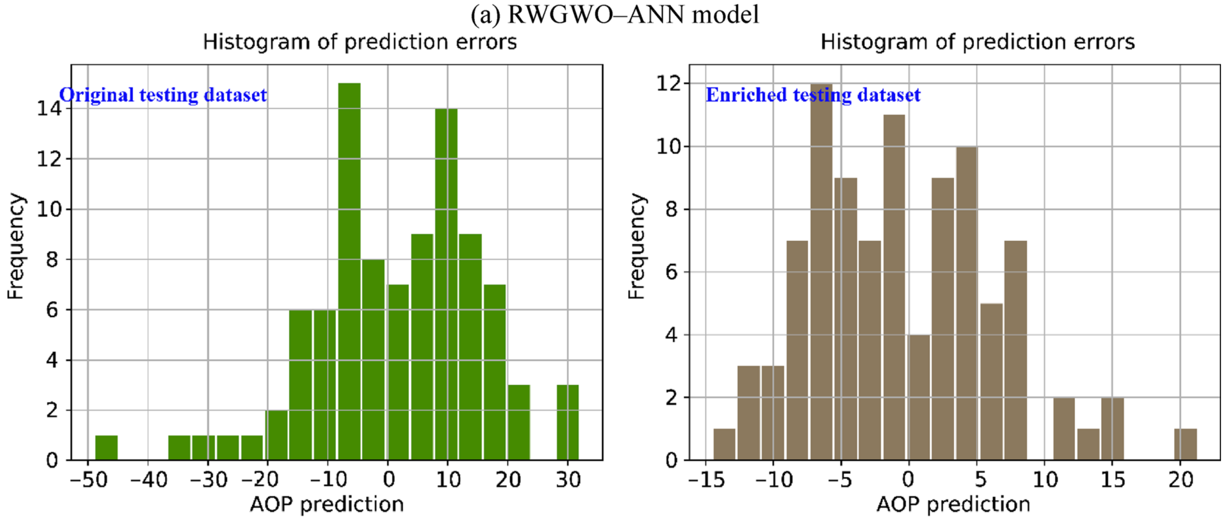
In contrast, examining the sensitivity analysis outcomes for the RWGWO–ANN model using the enriched dataset highlights that the additional variable, ExTree, took on the highest importance in predicting AOP. The remaining parameters contributed similarly in importance to the RWGWO–ANN model's AOP prediction. Despite the similar importance levels of the ExTree variable and the S variable, the enriched dataset demonstrated that the ExTree variable significantly enhanced the model's understanding of the interrelationships among variables, surpassing the lone S variable. The ExTree variable imparts a more comprehensive dataset representation to the RWGWO–ANN model, resulting in heightened accuracy compared to the original dataset.

CONCLUSIONS AND RECOMMENDATIONS

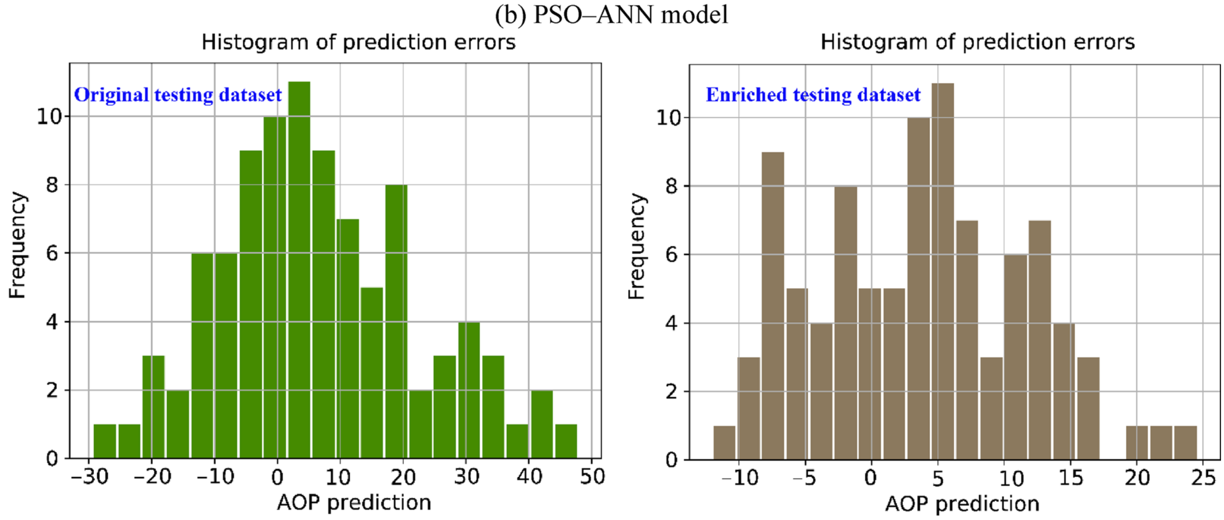
Ensuring the safety of surrounding areas relies heavily on precise prediction and control of AOP,



(a) RWGWO–ANN model

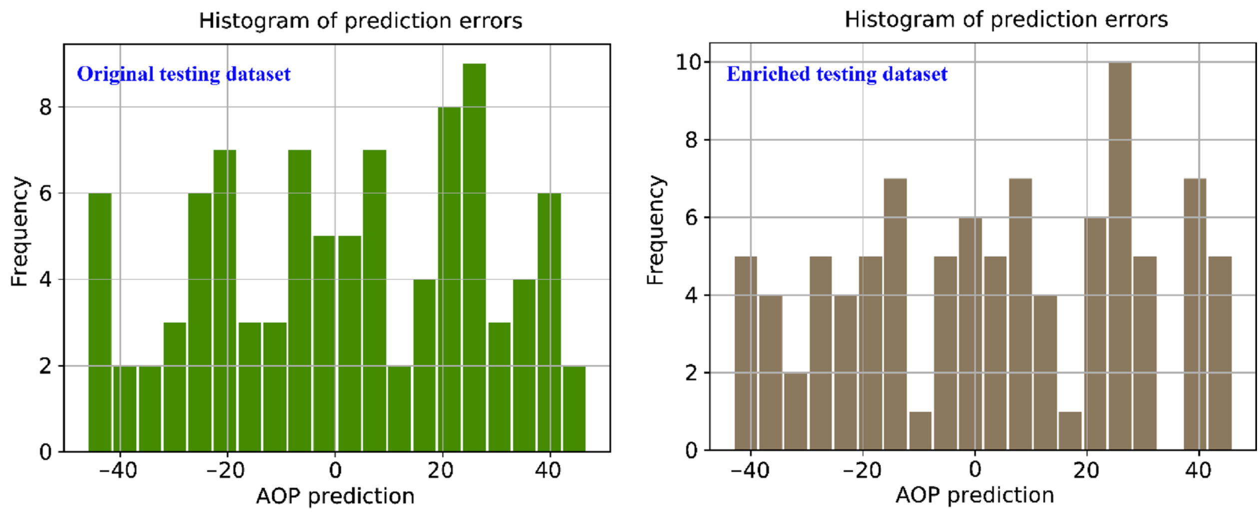


(b) PSO–ANN model

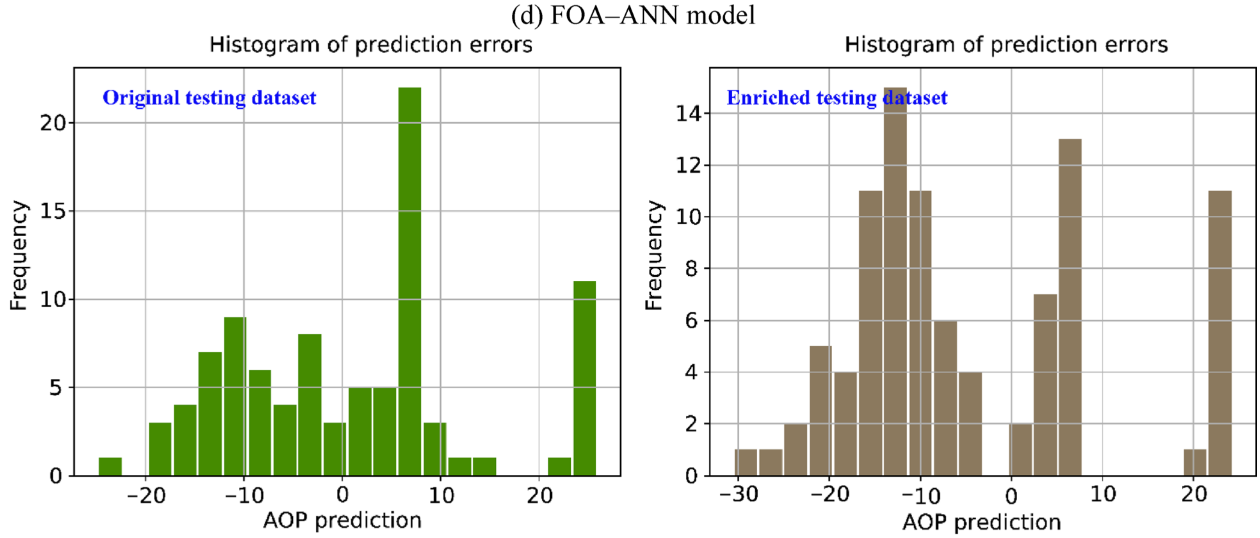


(c) SGA–ANN model

Figure 24. Error distributions of the developed models for predicting AOP.



(d) FOA–ANN model



(e) ANN model

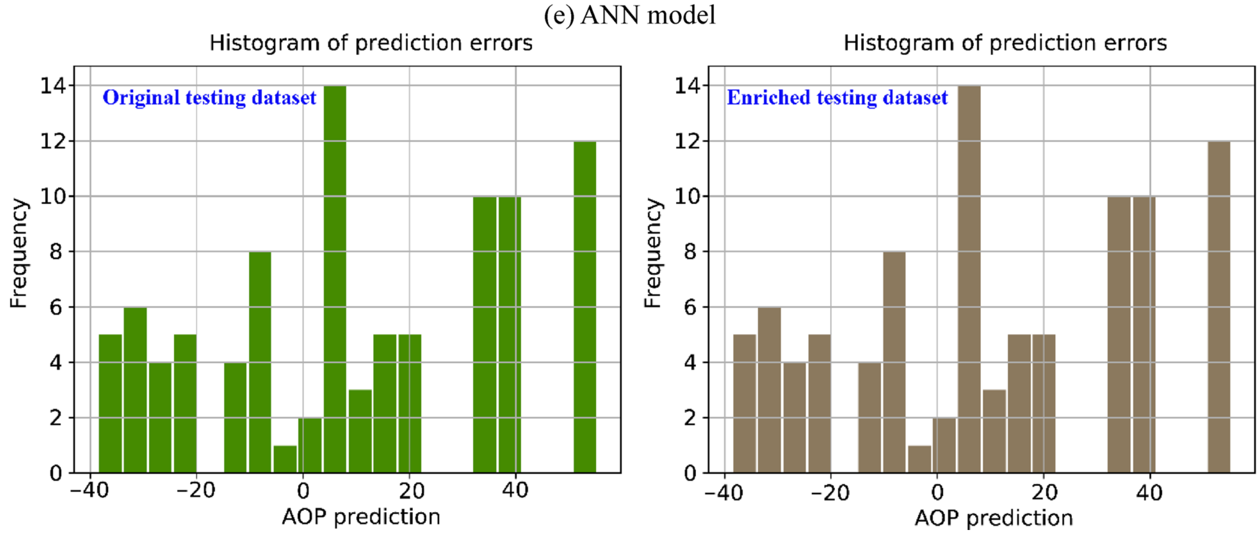


Figure 24. continued.

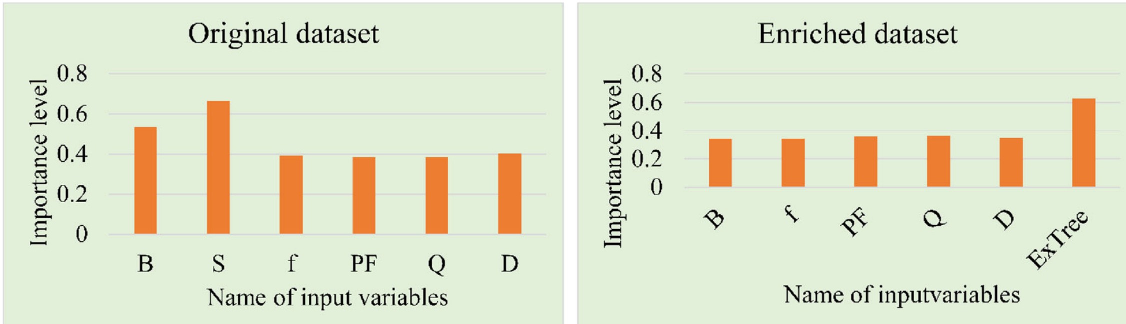


Figure 25. Exploring input variable sensitivity using the chosen RWGWO–ANN model with enriched and non-enriched datasets.

which arises from blasting activities in open-pit mines. This study introduced two innovative strategies toward this goal:

1. Utilizing the ExTree algorithm, we enriched the dataset to provide comprehensive information for AOP predictive models, enhancing their understanding of the dataset's nuances.
2. We introduced the novel RWGWO–ANN model, which combines an ANN model with two hidden layers and the new RWGWO optimization algorithm (developed based on the GWO algorithm and random walk strategy). This advancement aims to enhance the accuracy of conventional ANN models in AOP prediction.

The outcomes obtained underscore the significant role of the ExTree algorithm in data enrichment. It substantially improves the models' grasp of the dataset, thereby elevating the accuracy of prediction outcomes. Moreover, the proposed RWGWO–ANN model emerged as an exceptional contender for AOP prediction. Notably, its accuracy surged from 89 to 96.2% following dataset enrichment through the ExTree model.

We strongly recommend the adoption of these techniques in practical applications and future research endeavors. By doing so, the accuracy of predictions, especially concerning the adverse effects of blasting in open-pit mines, can be enhanced substantially.

DATA AVAILABILITY STATEMENT

The authors do not have permission to share the dataset.

ACKNOWLEDGMENTS

Dr. Hoang Nguyen was funded by the Postdoctoral Scholarship Programme of Vingroup Innovation Foundation (VINIF), code VINIF.2022.STS.23.

DECLARATIONS

Conflict of Interest The authors declare that they have no conflicts of interest.

SUPPLEMENTARY INFORMATION

The online version contains supplementary material available at https://doi.org/10.1007/s11053-023-10299-w.

REFERENCES

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data, 8*, 1–74.

Amiri, M., Bakhshandeh Amnieh, H., Hasanipanah, M., & Mohammad Khanli, L. (2016). A new combination of artificial neural network and K-nearest neighbors models to pre-

dict blast-induced ground vibration and air-overpressure. *Engineering with Computers, 32*, 631–644.
Armaghani, D. J., Hajihassani, M., Marto, A., Faradonbeh, R. S., & Mohamad, E. T. (2015a). Prediction of blast-induced air overpressure: A hybrid AI-based predictive model. *Environmental Monitoring and Assessment, 187*, 1–13.
Armaghani, D. J., Hajihassani, M., Sohaei, H., Mohamad, E. T., Marto, A., Motaghedi, H., & Moghaddam, M. R. (2015b). Neuro-fuzzy technique to predict air-overpressure induced by blasting. *Arabian Journal of Geosciences, 8*(12), 10937–10950.
Armaghani, D. J., Hasanipanah, M., & Mohamad, E. T. (2016). A combination of the ICA-ANN model to predict air-overpressure resulting from blasting. *Engineering with Computers, 32*(1), 155–171.
Azad, S. K., Azad, S. K., & Kulkarni, A. J. (2012). Structural optimization using a mutation-based genetic algorithm. *International Journal of Optimization in Civil Engineering, 2*(1), 80–100.
Baghbani, A., Choudhury, T., Costa, S., & Reiner, J. (2022). Application of artificial intelligence in geotechnical engineering: A state-of-the-art review. *Earth-Science Reviews, 228*, 103991.
Bai, Y., Gautam, T., & Sojoudi, S. (2023). Efficient global optimization of two-layer relu networks: Quadratic-time algorithms and adversarial training. *SIAM Journal on Mathematics of Data Science, 5*(2), 446–474.
Bansal, J. C. (2019). *Particle Swarm Optimization* (pp. 11–23). Berlin: Springer.
Bensingh, R. J., Machavaram, R., Boopathy, S. R., & Jebaraj, C. (2019). Injection molding process optimization of a bi-aspheric lens using hybrid artificial neural networks (ANNs) and particle swarm optimization (PSO). *Measurement, 134*, 359–374.
Bui, X.-N., Bui, H.-B., & Nguyen H. (2021). A review of artificial intelligence applications in mining and geological engineering. In *Proceedings of the international conference on innovations for sustainable and responsible mining: ISRM 2020*-Vol. 1, Springer. Berlin.
Chao, Z., Ma, G., Zhang, Y., Zhu, Y., & Hu, H. (2018). The application of artificial neural network in geotechnical engineering. In *IOP conference series: Earth and environmental science.* IOP Publishing.
de Paz-Centeno, I., García-Ordás, M. T., García-Olalla, Ó., & Alaiz-Moretón, H. (2023). Imputation of missing measurements in PV production data within constrained environments. *Expert Systems with Applications, 217*, 119510.
Dramsch, J. S. (2020). 70 years of machine learning in geoscience in review. *Advances in Geophysics, 61*, 1–55.
Falco, I. D., Cioppa, A. D., & Tarantino, E. (2002). Mutation-based genetic algorithm: Performance evaluation. *Applied Soft Computing, 1*(4), 285–299.
García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining*. Berlin: Springer.
Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning, 63*(1), 3–42.
Gómez-Iglesias, A., Vega-Rodríguez, M. A., Castejón-Magaña, F., Cárdenas-Montes, M., & Morales-Ramos, E. (2009). Grid-enabled mutation-based genetic algorithm to optimise nuclear fusion devices. In *Computer aided systems theory-EUROCAST 2009: 12th international conference, Las Palmas de Gran Canaria, Spain, February 15–20, 2009, Revised selected papers 12.* Springer, Berlin.
Guo, H., Li, H., Xiong, J., & Yu, M. (2019). Indoor positioning system based on particle swarm optimization algorithm. *Measurement, 134*, 908–913.
Gupta, S., & Deep, K. (2019). A novel random walk grey wolf optimizer. *Swarm and Evolutionary Computation, 44*, 101–112.
Han, K., & Wang, Y. (2021). A review of artificial neural network techniques for environmental issues prediction. *Journal of Thermal Analysis and Calorimetry, 145*(4), 2191–2207.
Harandizadeh, H., & Armaghani, D. J. (2021). Prediction of air-overpressure induced by blasting using an ANFIS-PNN model optimized by GA. *Applied Soft Computing, 99*, 106904.
Hasanipanah, M., Shahnazar, A., Amnieh, H. B., & Armaghani, D. J. (2017). Prediction of air-overpressure caused by mine blasting using a new hybrid PSO–SVR model. *Engineering with Computers, 33*(1), 23–31.
He, Z., Armaghani, D. J., Masoumnezhad, M., Khandelwal, M., Zhou, J., & Murlidhar, B. R. (2021). A combination of expert-based system and advanced decision-tree algorithms to predict air–overpressure resulting from quarry blasting. *Natural Resources Research, 30*, 1889–1903.
Hosseini, S., Monjezi, M., Bakhtavar, E., & Mousavi, A. (2021). Prediction of dust emission due to open pit mine blasting using a hybrid artificial neural network. *Natural Resources Research, 30*(6), 4773–4788.
John, V., Liu, Z., Guo, C., Mita, S., & Kidono, K. (2015). *Real-time lane estimation using deep features and extra trees regression*. Berlin: Springer.
Kan, J. (2017). Evaluation of mining engineering technology innovation ability and application based on BP neural network. In *2017 6th international conference on industrial technology and management (ICITM), IEEE.*
Eberhart & J. Kennedy (1995). A new optimizer using particle swarm theory. MHS'95. In *Proceedings of the sixth international symposium on micro machine and human science, IEEE.*
Kennedy, J. (2011). *Particle swarm optimization* (pp. 760–766). Berlin: Springer.
Lau, T., & Tsang, E. P. (1997). Solving the processor configuration problems with a mutation-based genetic algorithm. *International Journal on Artificial Intelligence Tools, 6*(04), 567–585.
Li, S., & Zheng, D. (2003). Applications of artificial neural networks to geosciences: Review and prospect. *Advances in Earth Science, 18*(1), 68.
Liu, Z., Peng, C., Xiang, W., Tian, D., Deng, X., & Zhao, M. (2010). Application of artificial neural networks in global climate change and ecological research: An overview. *Chinese Science Bulletin, 55*, 3853–3863.
Maier, O., Wilms, M., von der Gablentz, J., Krämer, U. M., Münte, T. F., & Handels, H. (2015). Extra tree forests for sub-acute ischemic stroke lesion segmentation in MR sequences. *Journal of Neuroscience Methods, 240*, 89–100.
Mohamad, E. T., Armaghani, D. J., Hasanipanah, M., Murlidhar, B. R., & Alel, M. N. A. (2016). Estimation of air-overpressure produced by blasting operation through a neuro-genetic technique. *Environmental Earth Sciences, 75*, 1–15.
Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics, 33*(2), 161–174.
Murlidhar, B. R., Nguyen, H., Rostami, J., Bui, X., Armaghani, D. J., Ragam, P., & Mohamad, E. T. (2021). Prediction of flyrock distance induced by mine blasting using a novel Harris Hawks optimization-based multi-layer perceptron neural network. *Journal of Rock Mechanics and Geotechnical Engineering, 13*(6), 1413–1427.
Nadi, A., Tayarani-Bathaie, S., & Safabakhsh, R. (2009). Evolution of neural network architecture and weights using mutation based genetic algorithm. In *2009 14th international CSI computer conference, IEEE.*
Nguyen, H., & Bui, X.-N. (2018). Predicting blast-induced air overpressure: a robust artificial intelligence system based on artificial neural networks and random forest. *Natural Resources Research, 28*(3), 893–907.
Nguyen, H., & Bui, X.-N. (2019). Predicting blast-induced air overpressure: A robust artificial intelligence system based on

artificial neural networks and random forest. *Natural Resources Research, 28*(3), 893–907.

Nguyen, H., Bui, X.-N., Tran, Q.-H., Van Hoa, P., Nguyen, D.-A., Hoa, L. T. T., Le, Q.-T., Do, N.-H., Bao, T. D., & Bui, H.-B. (2020). A comparative study of empirical and ensemble machine learning algorithms in predicting air over-pressure in open-pit coal mine. *Acta Geophysica, 68*, 325–336.

Pedram, M., Mousavirad, S. J., & Schaefer, G. (2022). Training neural networks with Lévy flight distribution algorithm. In *Proceedings of 7th international conference on harmony search, soft computing and applications: ICHSA 2022*, Springer, Berlin.

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

Ramesh Murlidhar, B., Yazdani Bejarbaneh, B., Jahed Armaghani, D., Mohammed, A. S., & Tonnizam Mohamad, E. (2021). Application of tree-based predictive models to forecast air overpressure induced by mine blasting. *Natural Resources Research, 30*(2), 1865–1887.

Refaat, M. (2010). *Data preparation for data mining using SAS*. Amsterdam: Elsevier.

Saadat, M., Khandelwal, M., & Monjezi, M. (2014). An ANN-based approach to predict blast-induced ground vibration of Gol-E-Gohar iron ore mine, Iran. *Journal of Rock Mechanics and Geotechnical Engineering, 6*(1), 67–76.

Saeed, U., Jan, S. U., Lee, Y.-D., & Koo, I. (2021). Fault diagnosis based on extremely randomized trees in wireless sensor networks. *Reliability Engineering & System Safety, 205*, 107284.

Sreejith, S., Nehemiah, H. K., & Kannan, A. (2020). A classification framework using a diverse intensified strawberry optimized neural network (DISON) for clinical decision-making. *Cognitive Systems Research, 64*, 98–116.

Tadeusiewicz, R. (2015). Neural networks in mining sciences–general overview and some representative examples. *Archives of Mining Sciences, 60*(4), 971–984.

Tran, Q.-H., Nguyen, H., & Bui, X.-N. (2023). Novel soft computing model for predicting blast-induced ground vibration in open-pit mines based on the bagging and sibling of extra trees models. *Computer Modeling in Engineering & Sciences, 134*(3), 2227–2246.

Xing, B., & Gao, W.-J. (2014). Fruit fly optimization algorithm. In *Innovative computational intelligence: A rough guide to 134 clever algorithms* (pp. 167–170).

Ye, J., Dalle, J., Nezami, R., Hasanipanah, M., & Armaghani, D. J. (2022). Stochastic fractal search-tuned ANFIS model to predict blast-induced air overpressure. *Engineering with Computers, 38*(1), 497–511.

Zhang, R., Li, Y., Gui, Y., & Zhou, J. (2022). Prediction of blasting induced air-overpressure using a radial basis function network with an additional hidden layer. *Applied Soft Computing, 127*, 109343.

Zhang, S., Carranza, E. J. M., Wei, H., Xiao, K., Yang, F., Xiang, J., Zhang, S., & Xu, Y. (2021). Data-driven mineral prospectivity mapping by joint application of unsupervised convolutional auto-encoder network and supervised convolutional neural network. *Natural Resources Research, 30*, 1011–1031.