Chapter 3

Application of machine learning and metaheuristic algorithms for predicting dust emission ($PM_{2.5}$) induced by drilling operations in open-pit mines

Xuan-Nam Bui[a,b], Hoang Nguyen[a,b], Yosoon Choi[c], Erkan Topal[d], and Tuan-Ngoc Le[e]

[a]*Department of Surface Mining, Mining Faculty, Hanoi University of Mining and Geology, Hanoi, Vietnam,* [b]*Innovations for Sustainable and Responsible Mining (ISRM) Research Group, Hanoi University of Mining and Geology, Hanoi, Vietnam,* [c]*Department of Energy Resources Engineering, Pukyong National University, Busan, South Korea,* [d]*Department of Mining Engineering and Metallurgical Engineering, WA School of Mines, Curtin University, Bentley, Perth, WA, Australia,* [e]*Vinacomin—Minerals Holding Corporation, Hanoi, Vietnam*

1 Introduction

Open-pit mining is known for its safety and efficiency, yet it still has room for improvement in terms of environmental considerations. To address its impact on the climate, a green and climate-sensitive approach must be taken [1–3]. Dust pollution is a major environmental concern in open-pit mining, posing threats to the environment, workers, surrounding ecosystems, and biodiversity [1,4–6]. The dust consists of particulate matter (PM) particles, which are tiny. Total suspended particles (TSP), larger than 100 μm, settle quickly, but particles smaller than 100 μm (e.g., $PM_{1.0}$) can stay suspended for longer. $PM_{2.5}$, which is less than 2.5 μm, contains harmful substances like heavy metals and can cause irreversible harm to human health by penetrating deep into the lungs [7–10]. Particles smaller than 2.5 μm (e.g., $PM_{1.0}$, $PM_{0.5}$, to name a few) are much more dangerous for humans [11,12]. Dust pollution also reduces work efficiency and causes low visibility, equipment malfunctions, higher maintenance costs, and lower labor productivity. Therefore, dust pollution in open-pit mining is extremely serious as it not only carries ordinary small-sized particulate matter but also toxic substances in the dust that have a serious impact on human health.

Applications of Artificial Intelligence in Mining and Geotechnical Engineering
https://doi.org/10.1016/B978-0-443-18764-3.00004-7

In open-pit mines, many operations can generate dust with small-sized particulate matter, such as drilling, blasting, shoving/loading/unloading, transporting, etc. [13–15], and the emission level is different. However, it is not easy to determine and forecast the dust pollution for each operation as they require a lot of time and specialized equipment suitable for each technological stage. In this book chapter, we considered the dust pollution in the drilling operations in open-pit mines and $PM_{2.5}$ was investigated.

A review of related works shows that most of the previous works are only focus on the prediction of $PM_{2.5}$ in the entire space of open-pit mine based on time-series datasets and artificial intelligence (AI) models without the consideration of rock mass properties [13,16–18]. Meanwhile, the main objectives of open-pit mines are rock and ore, and they are influenced by factors related to the mechanical and physical characteristics of rock mass. Regarding to the prediction of dust emissions induced by drilling operations in open-pit mines, Sastry et al. [19] applied the environmental management plan (EMP) model based on meteorological data (e.g., wind speed, wind direction, relative humidity, and temperature) and the geographical data (e.g., distance from the dust source and emission rate of drilling operation) to predict $PM_{2.5}$ and PM_{10} generated from drilling operations in open-pit mines and they concluded that the introduced model is in close agreement with the field measured values than presently USEPA model. We also conducted a study on the prediction of PM_{10} concentration from drilling operations in open-pit mines using the integration of support vector machine for regression problems (SVR) and particle swarm optimization (PSO) model, abbreviated as PSO-SVR model. It then was compared to three benchmark AI-based models, including k-nearest neighbors (KNN), random forest (RF), and classification and regression trees (CART). Different kernel functions were also considered for the PSO-SVR model during predicting PM_{10} concentrations. Finally, we found that AI-based techniques are potential approach for predicting PM_{10} concentration from drilling operations in open-pit mines, and the proposed AI model achieved an accuracy of 95% for this aim.

In this chapter, we introduced an application of a machine learning algorithm (i.e., gradient boosting machine—GBM) and two AI-based metaheuristic algorithms (i.e., particle swarm optimization—PSO and differential evolution—DE) for predicting $PM_{2.5}$ emissions from drilling operations in open-pit mines, resulting in the PSO-GBM and DE-GBM models, respectively. The detail of the methodology and results are presented in the next sections. This work aims to assess the impact of $PM_{2.5}$ dust emissions from drilling operations on air pollution in open-pit mines. Additionally, it seeks to evaluate the feasibility of using new hybrid models that combine GBM models and metaheuristic algorithms to estimate $PM_{2.5}$ levels induced by drilling operations in such mines. The findings of this study will provide valuable insights into innovative approaches for promoting sustainable and responsible mining practices.

2 Methodology

2.1 Gradient boosting machine (GBM)

Gradient boosting machine (GBM) is one of the machine learning algorithms that can be applied for both regression and classification problems, and it was widely applied in a variety of fields including mining and geotechnical engineering [20–25]. It works by combining the results from several weaker models to generate a stronger and more accurate prediction [26,27]. This is achieved through the creation of multiple decision trees in a sequential order, where each new tree attempts to correct the errors made by the previous tree [28].

The fundamental concept of gradient boosting involves fitting an additive model in a step-by-step fashion, where each new tree is developed using information gathered from previously grown trees. During each iteration, a new tree is fit to the negative gradient or residual of the loss function about the current prediction. The final prediction is then obtained by adding up the predictions from all the trees.

The GBM comprises several key components, including a loss function, a weak learner (such as decision trees), and an optimization algorithm (such as gradient descent) [29,30]. The algorithm reduces the loss function by modifying the parameters of the weak learner, while the optimization algorithm ensures that the new tree fits the residuals in a way that minimizes the loss. The details of the GBM model for predicting $PM_{2.5}$ emissions induced by drilling operations in open-pit mines are presented through seven steps in the following section. The framework of GBM is shown in Fig. 1.

1. Loss function: The first step is to define a loss function that measures the difference between the true values and the predicted values of the target variable. Common loss functions used in GBM include mean squared error (MSE) for regression problems and log loss for classification problems.
2. Weak learner: The next step is to define a weak learner, which is typically a decision tree model. A decision tree is a type of model that partitions the input data into regions and makes predictions based on the mean response value in each region.
3. Initial prediction: The initial prediction of the target variable is made using the weak learner, typically using the mean response value for regression problems or the most frequent class for classification problems.
4. Residual computation: The residual is computed as the difference between the true values and the initial prediction. This residual represents the error made by the weak learner and is used to guide the fit of the next tree.
5. Tree fitting: The next step is to fit a decision tree model to the residuals. The tree is fit in a greedy manner, such that it tries to reduce the residual at each split. The tree is grown until a stopping criterion is reached, such as a maximum tree depth or a minimum number of samples in a leaf node.

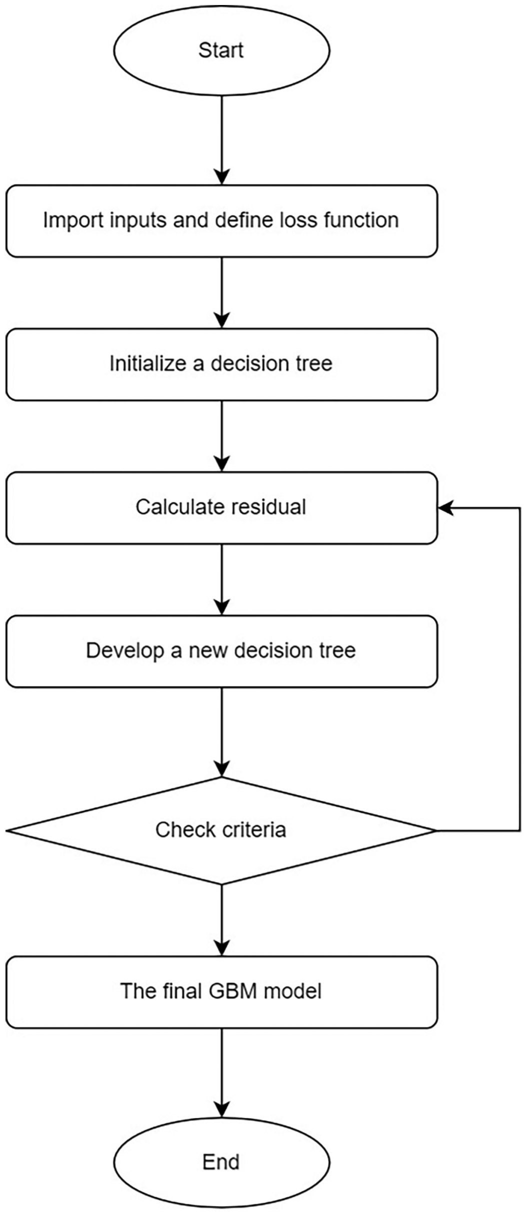

FIG. 1 Framework of GBM algorithm.

6. Update prediction: The prediction made by the tree is added to the initial prediction, and the new prediction is used to compute the residual for the next tree. This process is repeated until a specified number of trees has been fit.
7. Final prediction: The final prediction is obtained by summing up the predictions made by all the trees. The prediction made by each tree is weighted based on its contribution to the reduction of the residual.

GBM is well-regarded for its excellent performance on a variety of problems, and is considered to be one of the most powerful machine learning algorithms available. However, it should be noted that GBM can be computationally intensive, and may also overfit if not properly regulated with too many trees being fit or the trees are too deep. Regularization techniques to overcome this problem include early stopping, tree pruning, and shrinkage, which can be used to reduce the risk of overfitting [31–33]. More details of the GBM algorithm can be found in the literature [27,34,35].

2.2 Differential evolution (DE) algorithm

Differential evolution (DE) is known as one of the metaheuristic algorithms used to find the optimal solution to a multi-dimensional, real-valued problem and it that was designed as a stochastic direct search method [36]. The DE algorithm follows an evolutionary approach, which is inspired by the principles of natural selection and genetics, to find the best solution from a population of candidate solutions.

In DE, each candidate solution is represented as a vector of real numbers. During each iteration of the algorithm, a new solution is generated by combining the differences between three randomly selected candidate solutions [37,38]. This new solution is then compared to the existing solutions, and if it is found to be better, it replaces one of the existing solutions. This process repeats until a stopping criterion is reached, such as a maximum number of iterations or a desired level of accuracy [39,40].

DE can strike a balance between exploration and exploitation, by controlling the size of the difference vector used to generate new solutions. A larger difference vector encourages exploration, while a smaller difference vector emphasizes exploitation. This feature makes DE highly adaptable to various optimization problems [41–43].

DE has been applied successfully to a broad range of optimization problems, including function optimization, parameter tuning, and feature selection, and is known for its ability to handle large and complex optimization problems effectively, such as mining and geotechnical engineering-related problems [44–50]. However, one of the most disadvantages of the DE is computationally intensive, and there is a possibility that it may not converge to the global optimum solution [51]. To mitigate these limitations, variations of DE have been developed, such as SaDE, jDE, ADE, SDE, and JADE, that incorporate additional strategies to

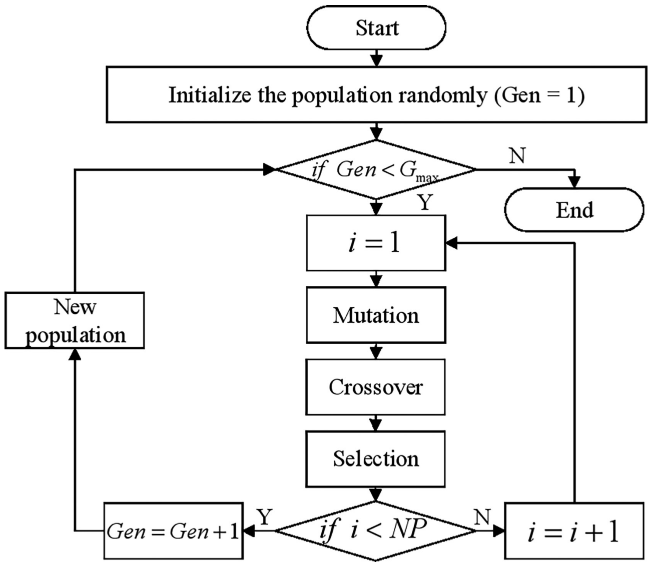

FIG. 2 Framework of the DE algorithm.

enhance the performance of the algorithm [52]. More details of the DE algorithm can be referred to as the literature [53–56]. The framework and pseudo-code of the DE algorithm are shown in Figs. 2 and 3.

2.3 Particle swarm optimization (PSO)

Particle swarm optimization (PSO) is a metaheuristic optimization algorithm that is used similar to the DE algorithm. It was first proposed by Kennedy and Eberhart [57] and has since been widely used in a variety of fields, such as engineering, computer science, mining, geotechnical engineering, mechanical engineering, solar photovoltaic system, to name a few [58–63].

The PSO algorithm works by representing candidate solutions as particles in a search space. Each particle represents a potential solution to the optimization issue. It has its location and velocity within the particle swarm environment. PSO starts with a randomly generated group of particles and continually updates based on personal and global extremes, leading to a new set of populations that continually evolve until the optimal solution is found [64]. In other words, at each iteration, each particle is updated based on its own experience, the experience of its neighbors, and the global best solution found by the swarm. The position and velocity of each particle are updated according to a set of equations that balance exploration and exploitation. Exploration is encouraged by adding randomness to the velocity update, while exploitation is encouraged by steering particles toward the best known solutions. The framework of the PSO algorithm is shown in Fig. 4 and its pseudo-code is presented in Fig. 5.

The pseudo-code of the DE algorithm for optimization problems

```
function DE(func, bounds, pop_size, mut, crossp, max_iter)
  dimensions = length of bounds
  # Initialize the population:
  population = []
  # Evaluate the fitness of the individuals:
  for i = 0 to pop_size-1
    indv = []
    for j = 0 to dimensions-1
      indv[j] = random number between bounds[j][0] and bounds[j][1]
    population[i] = indv

  for i = 0 to max_iter-1
    for j = 0 to pop_size-1
      # Select three individuals randomly:
      r0, r1, r2 = three random individuals from population
      # Generate a mutant vector:
      mutant = []
      # Crossover:
      for k = 0 to dimensions-1
        if random number between 0 and 1 is less than or equal to crossp
          mutant[k] = r1[k] + mut * (r2[k] – r0[k])
        else
          mutant[k] = population[j][k]
      # Selection:
      if f(mutant) is less than f(population[j])
        population[j] = mutant

  best_ individual = 0
  for i = 1 to pop_size-1
    if func(population[i]) is less than func(population[best_ individual])
      best_ individual = i
  return population[best_ individual]
```

FIG. 3 Pseudo-code of the DE algorithm.

One of the main advantages of PSO is its ease of use, as it only requires a few parameters to be specified, such as the number of particles and the weighting factors used in the velocity update. It is also computationally efficient and easy to implement, making it a popular choice for solving optimization problems.

However, PSO has some limitations, including the sensitivity of its performance to the choice of parameters and the risk of getting stuck in local optima. To overcome these limitations, various modifications to the basic PSO algorithm have been proposed, such as inertia weight, constriction coefficient, velocity clamping, adaptive PSO, social-only PSO, hybrid PSO (combining PSO with other optimization algorithms) [65–69]. These variants incorporate additional strategies to improve the performance of the algorithm and to make it more robust to the choice of parameters. Therefore, in this book chapter, a variant of the PSO algorithm with the inertia weight strategy was applied to overcome the limitations of the basic PSO algorithm for the optimization of GBM model in forecasting $PM_{2.5}$ emissions from drilling operations.

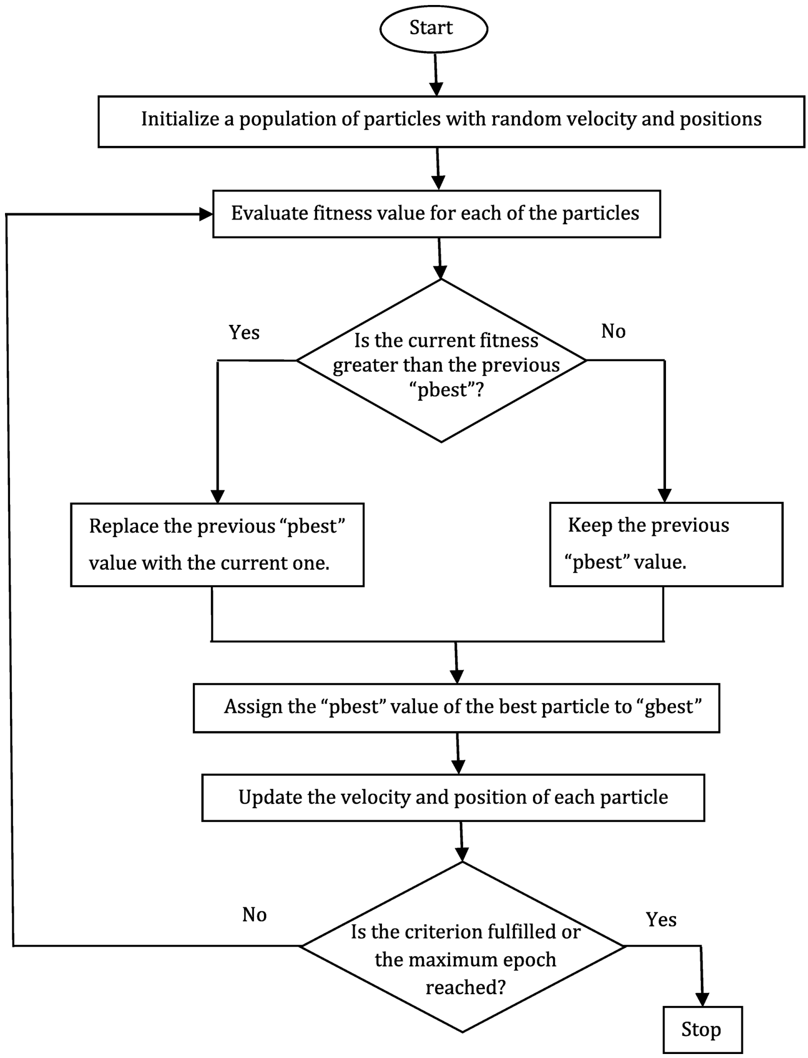

FIG. 4 Framework of the PSO algorithm.

2.4 Integration of DE, PSO and GBM model

The integration of the DE, PSO, and GBM models is motivated by the goal of finding the optimal parameters for the GBM model, with the aim of improving its accuracy in forecasting $PM_{2.5}$ caused by drilling operations in open-pit mines. While the basic GBM model includes parameters such as the learning rate, number of weak learners, and depth of decision trees, selecting the best values for these parameters within a valid range is a challenging task. Normally, the grid search or random search techniques are often applied to select these parameters; however, we do not know which grid is the best and it is time

The pseudo-code of the PSO algorithm for optimization problems

```
Initialize a population of particles with random positions and velocities
Set the best known positions for each particle to their initial positions
Set the global best position to the best known position among all particles
Set the maximum number of iterations

while (iterations < maximum iterations):
  for each particle in the population:
    Evaluate the fitness of the current position
    If the current position is better than the best known position:
      Update the best known position for the particle
      If the best known position for the particle is better than the global best position:
        Update the global best position
    Update the velocity and position of the particle using the following formulae:
      velocity = inertia * velocity +
           cognitive_weight * random() * (best known position - current position) +
           social_weight * random() * (global best position - current position)
      position = position + velocity
  iterations = iterations + 1

Return the global best position
```

FIG. 5 Pseudo-code of the PSO algorithm.

consuming if we search the best parameters by random search. Furthermore, compared to the grid search or random search techniques, metaheuristic algorithms offer several advantages, such as:

1. Exploration of a broader search space: Metaheuristic algorithms are designed to explore a wider range of solutions by utilizing various search strategies. This allows them to effectively navigate complex and high-dimensional search spaces, which may not be efficiently covered by a grid-based approach.
2. Flexibility and adaptability: Metaheuristic algorithms are highly flexible and adaptable to different problem domains. They can dynamically adjust their search behavior based on problem characteristics and constraints. In contrast, the grid search technique relies on a predefined set of grid points, limiting its ability to adapt to changing problem conditions.
3. Efficient handling of continuous and discrete variables: Metaheuristic algorithms are well-suited for optimization problems with both continuous and discrete variables. They can handle mixed variable types effectively, allowing for a more comprehensive exploration of the solution space. However, the grid search technique is primarily suitable for discrete variables and may struggle with continuous variables due to the limited granularity of the grid.
4. Global optimization capabilities: Metaheuristic algorithms excel at finding near-optimal or globally optimal solutions, even in multimodal and non-convex problem landscapes. They can escape local optima and converge toward better solutions by utilizing strategies such as population-based search, randomization, and diversification. The grid search technique, on the other hand, may get stuck in local optima and struggle to identify global optima.

5. Computational efficiency: Metaheuristic algorithms often offer computational efficiency by using smart search heuristics and adaptive techniques. They can significantly reduce the number of function evaluations required to find good solutions compared to exhaustive grid-based methods like grid search, especially in high-dimensional optimization problems.

Therefore, the DE and PSO algorithms were applied to optimize the parameters of the GBM model, and they are then applied to train the GBM model for forecasting $PM_{2.5}$ caused by drilling operations at the Coc Sau open-pit coal mine.

To hybrid the DE-GBM and PSO-GBM models, the DE and PSO algorithms will be used to generate new candidate solutions, and each solution contains a set of parameters, which are used as input to the GBM model for prediction. Subsequently, the GBM model generates $PM_{2.5}$ predictions for each candidate solution, and the DE and PSO algorithms will update the candidate solutions based on the prediction errors. This process is repeated until a stopping criterion is reached, and the best parameters with the lowest error will be defined. Fig. 6 illustrates the flowchart of the DE-GBM and PSO-GBM for estimating PM2.5 in this study. Compared to traditional methods such as grid search or random search, the DE and PSO algorithms are able to generate a greater number of solutions with diverse sets of parameters for the GBM model based on the fitness function. As a result, these algorithms are more effective than traditional methods, and this claim will be demonstrated in the results section.

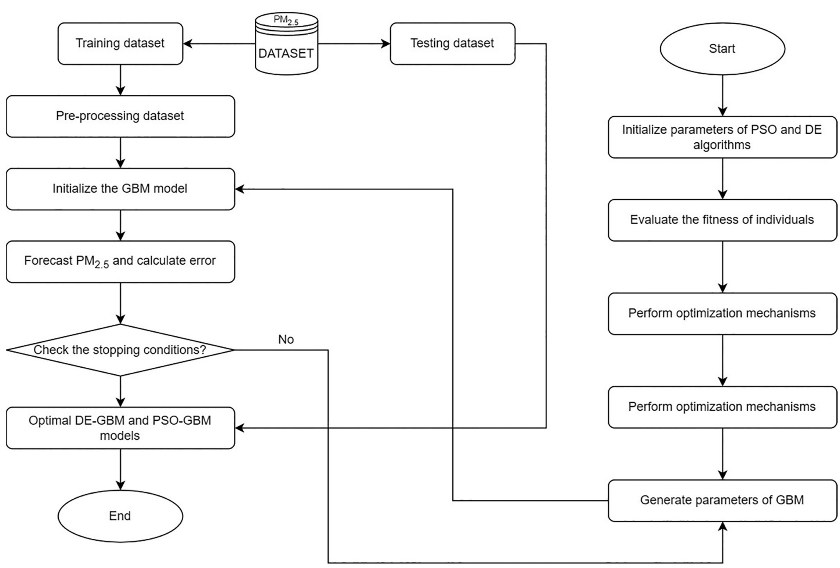

FIG. 6 Framework of the PSO-GBM and DE-GBM models for forecasting $PM_{2.5}$.

2.5 Performance metrics for evaluation

For assessment of the models' efficiency, five metrics, including mean absolute error RMSE, (MAE), determination coefficient (R^2), mean absolute percentage error (MAPE), and variance accounted for (VAF), were used as described in Eqs. (1)–(3).

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

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y}_i)^2} \tag{2}$$

$$\mathrm{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \widehat{y}_i|}{y_i} \tag{3}$$

where n is the total number of datasets used in this book chapter; y_i stands for the ith actual $PM_{2.5}$; $\widehat{y}_i$ denotes the ith predicted $PM_{2.5}$; $\overline{y}_i$ represents for the mean of actual $PM_{2.5}$ values.

3 Data acquisition and preparation

In this chapter, the drilling operations at the Coc Sau open-pit coal mine (Vietnam) and their generated $PM_{2.5}$ were investigated and taken into account. The Coc Sau open-pit coal mine is one of the deepest open-cast coal mine in Vietnam with high production (i.e., 2–3 million t/year, overburden reached 30–40 million m^3/year) at the survey period (2018), and its location is shown in Fig. 7.

The Hon Gai formation (T3n-rhg), which consists of various sedimentary rocks including coal seams, gritstone, conglomerate, sandstone, claystone, siltstone, and shale, covered the entire Coc Sau open-cast coal mine. These sedimentary rocks are generally hard with a Protodyakonov strength index (f) ranging from 8 to 11 [70]. As a result, the method of exploiting coal in the mine was determined to be fragmentation by drilling-blasting, as it is considered effective.

The mine typically used the СБШ-250, D245S, and DML drills for drilling, with borehole diameters ranging from 200 to 250 mm. The average drilling speed was between 10 and 15 m/h, resulting in a significant amount of $PM_{2.5}$ concentration. This increased the amount of dust impacting the surrounding environment and public health, which is especially concerning given the mine's proximity to residential areas (approximately 0.42 miles or 700 m away). The employees working in the mine are particularly at risk of occupational hazards. Fig. 7 illustrates the extent of the impact of dust, particularly the $PM_{2.5}$ concentration.

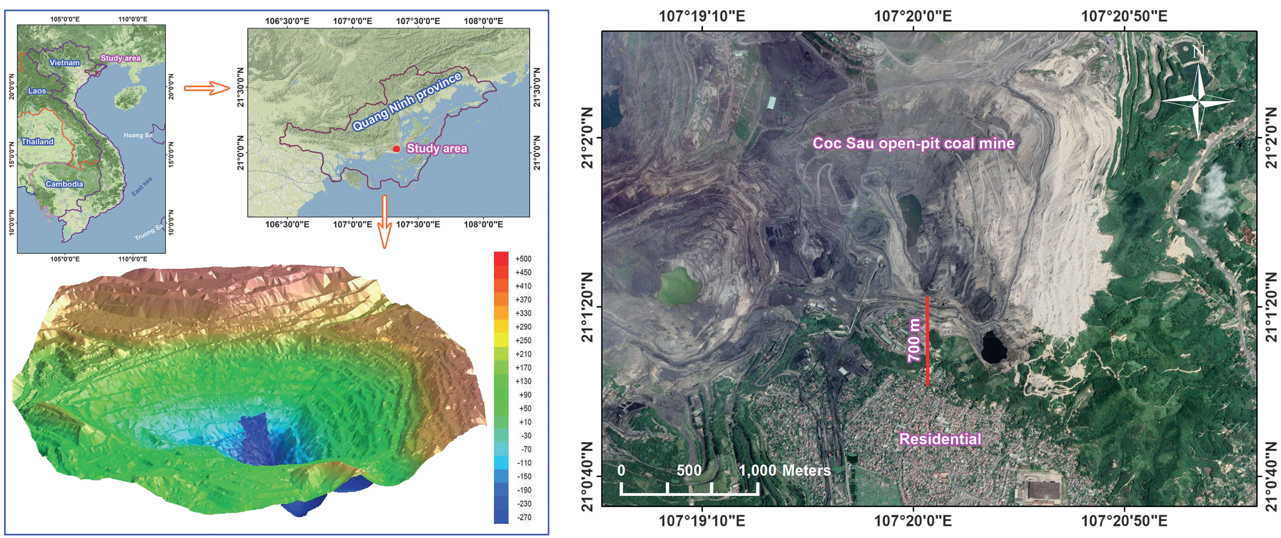

FIG. 7 Location of the COC SAU open-pit coal mine.

At the Coc Sau open-pit coal mine, the СБШ-250, D245S, and DML were used to drill boreholes with diameters ranging from 200 to 250 mm. To forecast $PM_{2.5}$, the following parameters were collected: the diameter of the boreholes (d), the penetration rate of the drill (P), the moisture content (W_{tn}), the silt content (S), the density of the rock mass (ρ), the compressive strength (σ_c), and the rebound hardness number (R), as previously mentioned. The KANOMAX digital dust monitor model no. 3442 was used to measure $PM_{2.5}$ induced by drilling operations in this study. It is important to mention that wind direction and wind speed are also taken into consideration when monitoring and forecasting $PM_{2.5}$ levels caused by drilling operations. However, since this study focuses solely on the $PM_{2.5}$ dust emission from drilling operations, the wind direction was determined before establishing measurement stations at a distance of 30–50 m. Fig. 8 shows the data collection progress and Table 1 presents the characteristics of the dataset used.

FIG. 8 Data collection and devices used in this study.

TABLE 1 Characteristics of the inputs and $PM_{2.5}$ values collected in this study.

d	P	W_{tn}	S
Min.: 200.0	Min.: 0.1600	Min.: 0.29	Min.: 15.20
First Qu.: 200.0	First Qu.: 0.2100	First Qu.: 7.84	First Qu.: 24.70
Median: 230.0	Median: 0.2500	Median: 11.38	Median: 27.60
Mean: 227.3	Mean: 0.2558	Mean: 11.57	Mean: 27.58
Third Qu.: 250.0	Third Qu.: 0.2900	Third Qu.: 15.39	Third Qu.: 30.10
Max.: 250.0	Max.: 0.4100	Max.: 28.12	Max.: 39.20
ρ	σ_c	R	$PM_{2.5}$
Min.: 1.220	Min.: 13.00	Min.: 16.00	Min.: 0.1373
First Qu.: 1.230	First Qu.: 15.00	First Qu.: 20.00	First Qu.: 0.3051
Median: 1.240	Median: 16.00	Median: 22.00	Median: 0.4396
Mean: 1.243	Mean: 15.98	Mean: 21.63	Mean: 0.4602
Third Qu.: 1.260	Third Qu.: 17.00	Third Qu.: 23.00	Third Qu.: 0.5991
Max.: 1.270	Max.: 19.00	Max.: 27.00	Max.: 1.2112

4 Results and discussion

Before developing the models, the dataset should be preprocessed and it was divided into two parts with 80% of the whole dataset was applied for training the model and the remaining 20% of the dataset was used for testing the developed model. The input data was also normalized using the Min-Max scaling method to reduce the error of the models during training. Furthermore, 10-folds cross-validation technique was also applied to evaluate the models during training aiming to avoid the bias evaluations.

Once the dataset was preprocessed, the PSO's parameters including swarm size, maximum iterations, local coefficient, global coefficient, and Min-Max weights of the swarms were setup as 300, 1000, 1.2, 1.2, 0.4, and 0.9, respectively. Subsequently, the parameters of the GBM model were initialized, including learning rate, the number of weak learners, and the depth of the decision trees. In the next step, the population of particles with random positions and velocities in the search space of GBM hyperparameters were initialized. Then, the RMSE was used as the fitness function to evaluate the fitness of each particle by training the GBM model using the training dataset and calculating the RMSE on the validation dataset. Based on the fitness of each particle, the personal best solution and the global best solution were updated. The PSO algorithm then updated the velocity and position to obtain the best solution. In this phase, each population is considered a solution with a set of GBM's parameters was generated in their feasible range that were generated by the PSO algorithm, and the GBM model was then trained with the generated parameters by the PSO and evaluated through the RMSE values calculated. The progress was repeated in 1000 iterations until the maximum number of iterations is reached or the convergence criterion is met.

For the DE-GBM model, similar steps were performed to develop the DE-GBM model for forecasting $PM_{2.5}$ induced by drilling operations. However, the parameters of the DE algorithm is different from the PSO algorithm except the population size (i.e., 300) and maximum iterations (i.e., 1000). The training processes of the PSO-GBM and DE-GBM models are shown in Fig. 9, and the prediction results are shown in Fig. 10. In addition, the summary of the performances is shown in Table 2.

Fig. 9 shows that both models converge well, indicating that they have overcome the overfitting problem during training and testing. This result may be attributed to the preprocessing of the dataset and the use of the 10-fold cross-validation technique for model evaluation during training.

Fig. 10 shows that both the DE-GBM and PSO-GBM models accurately forecasted $PM_{2.5}$ levels induced by drilling operations. It is difficult to determine which model performed better, as both achieved high accuracy. This indicated that the parameters optimized by the DE and PSO algorithms are in line with the forecast of $PM_{2.5}$ induced by drilling operations at the Coc Sau open-pit coal mine. For further assessment of the obtained results by the DE-GBM and PSO-GBM models, performance metrics such as RMSE, R^2, and MAPE were calculated, as listed in Table 2.

Based on the metrics computed in Table 2, it is clear that both models perform well, with MAPE values in the range of 10%–11%. However, upon closer examination of Table 2, we can observe that the DE-GBM model outperforms the PSO-GBM model slightly, but the difference is not significant. These results showed that machine learning algorithms (i.e., GBM model) and metaheuristic algorithms are potential approaches to forecast air pollution (i.e., dust emissions) generated by drilling operations in open-pit mines. Fig. 11 shows the regression between the measured and forecasted $PM_{2.5}$ by the developed models.

Observing the datasets in Fig. 11, it is clear that most of the forecasted datasets are close to the measured datasets, and both models exhibit a high degree of similarity when it comes to data points lying outside the regression line, even though their exact positions may vary slightly. Overall, the results demonstrated

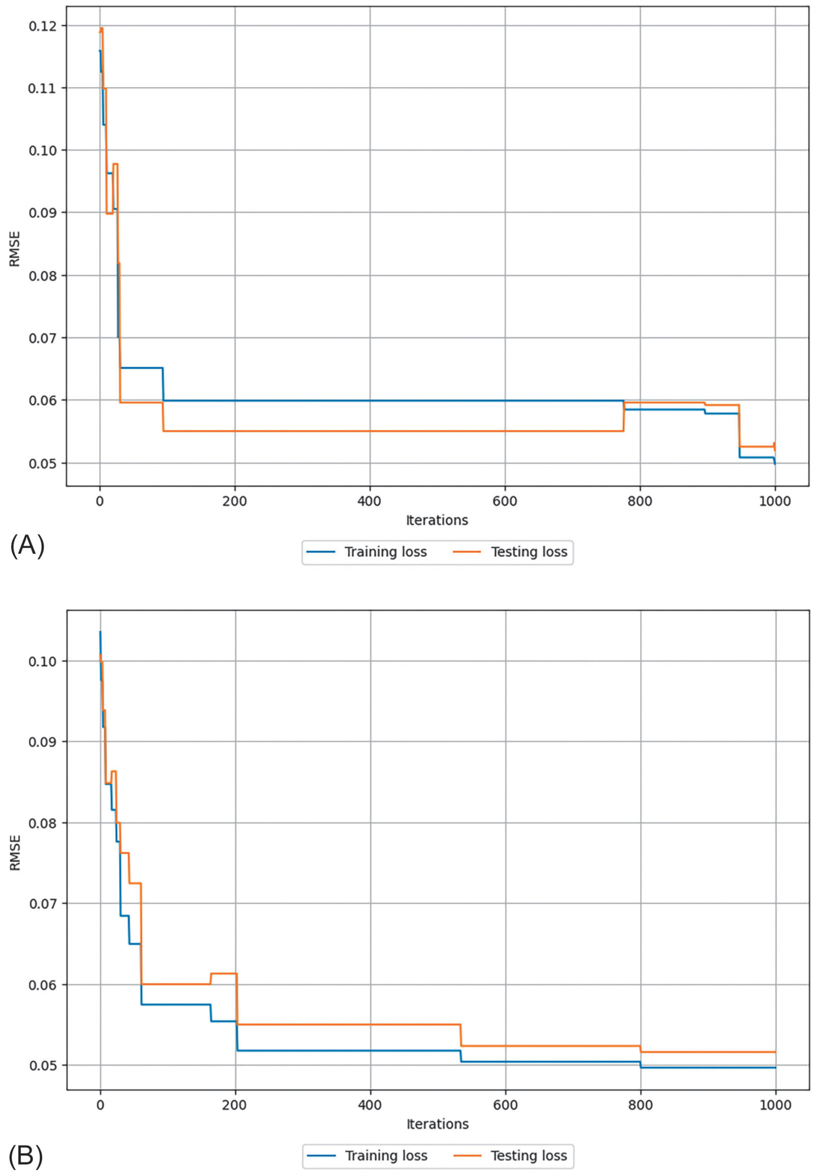

FIG. 9 Optimization performance of the PSO-GBM and DE-GBM models. (A) PSO-GBM model and (B) DE-GBM model.

the potential application of AI for forecasting $PM_{2.5}$ levels resulting from drilling operations in open-pit mines, as well as for predicting air pollution in future studies that utilize AI techniques. This represents one of the pathways to achieving green mining as quickly as possible in the future.

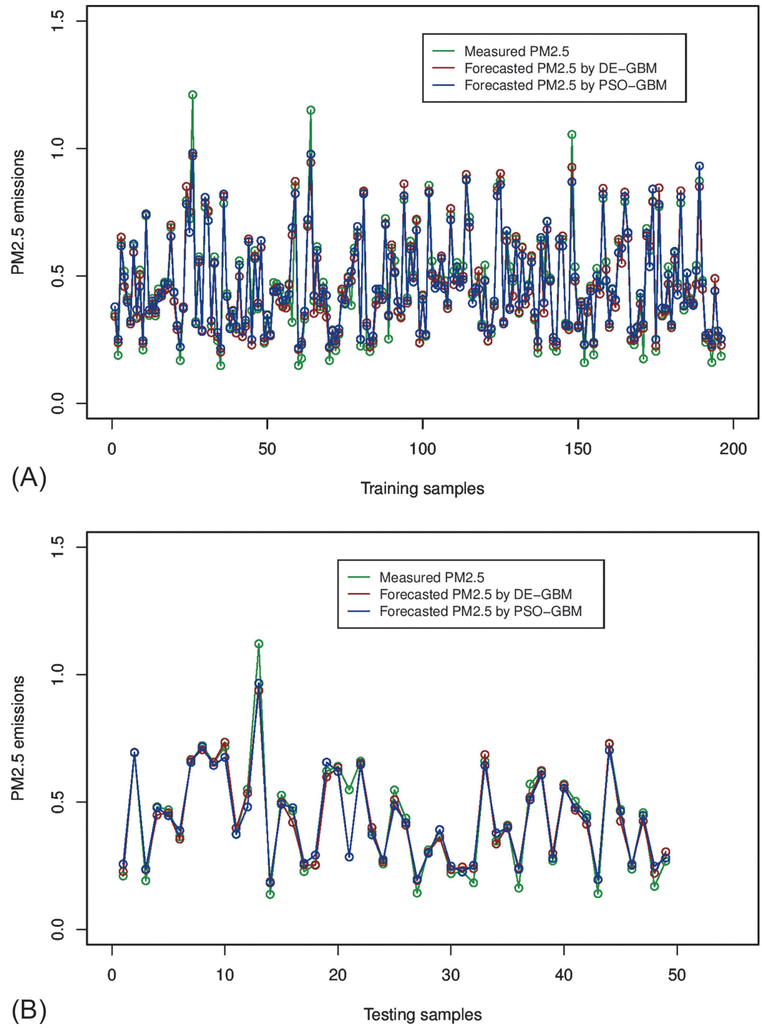

FIG. 10 Comparison of the measured and predicted $PM_{2.5}$. (A) Training datasets and (B) testing dataset.

TABLE 2 Performance of the PSO-GBM and DE-GBM models for forecasting $PM_{2.5}$ induced by drilling operations.

	Training dataset			Testing dataset		
Model	*RMSE*	R^2	*MAPE*	*RMSE*	R^2	*MAPE*
PSO-GBM	0.053	0.931	0.092	0.055	0.929	0.110
DE-GBM	0.053	0.931	0.084	0.055	0.930	0.102

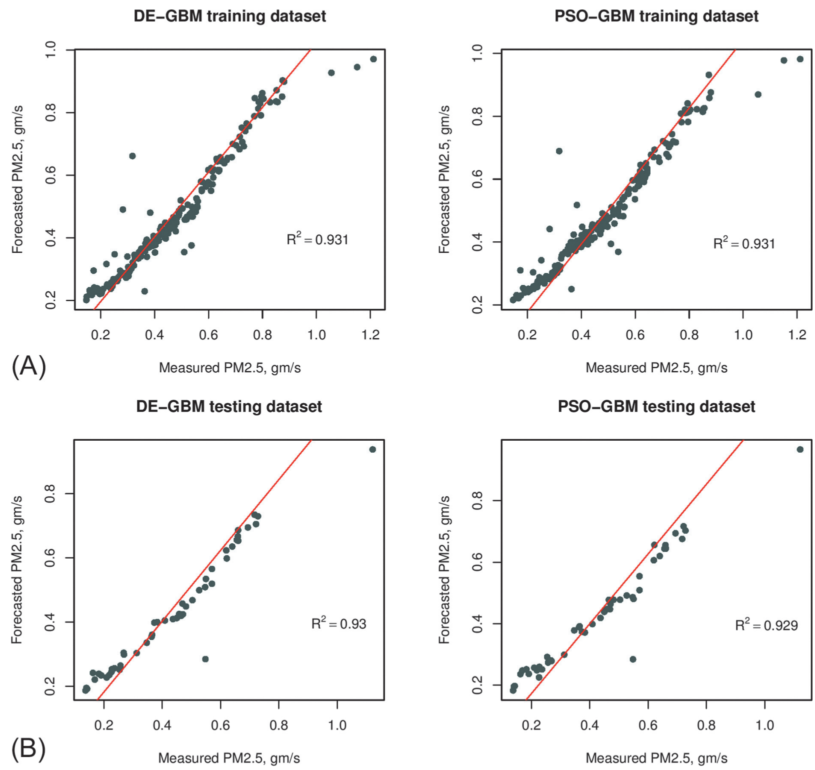

FIG. 11 Regression analysis of the developed models for forecasting $PM_{2.5}$ induced by drilling operations. (A) Training dataset and (B) testing dataset.

5 Conclusion

Dust pollution in open-pit mining is one of the barriers on the road to achieving green mining. Therefore, predicting dust pollution, especially $PM_{2.5}$ dust, with the support of AI in this study is a breakthrough in reaching this goal. This study demonstrated the effective application of AI in forecasting $PM_{2.5}$ caused by drilling operations in open-pit mines, while also uncovering and clarifying the relationships between drilling parameters, soil moisture, and the amount of dust generated during the drilling process. Finally, two optimized models, that is, DE-GBM and PSO-GBM, were proposed to address this issue and they have achieved promising results.

While the results obtained suggest that the GBM model, optimized using DE and PSO algorithms, shows promise in predicting and mitigating $PM_{2.5}$ dust emissions from drilling operations in open-pit mines, several limitations warrant further research in future studies. These limitations include:

1. Inclusion of meteorological conditions: Future research should delve into considering the impact of meteorological conditions on $PM_{2.5}$ emissions

from drilling operations. This will enhance the comprehensiveness of the models and provide a more accurate understanding of the contributing factors.

2. Real-time measurement device and time-series analysis: It is crucial to employ a real-time measurement device capable of capturing $PM_{2.5}$ levels in real time. Additionally, the issue of time-series analysis should be taken into account to effectively address the temporal aspects of $PM_{2.5}$ emissions.
3. Expansion to different areas: To ensure consistent evaluations of the proposed models in practical engineering scenarios, it is necessary to survey and consider different areas. This will provide a broader perspective and account for potential variations in the performance of the models.

By addressing these limitations in future research, we can enhance the applicability and robustness of the proposed models in predicting and mitigating $PM_{2.5}$ dust emissions from drilling operations in open-pit mines.

Acknowledgments

The authors would like to thanks to Prof. Changwoo Lee (Dong A University, South Korea), Dr. Nguyen Van Duc (KIGAM Institute, South Korea), and the Coc Sau open-pit coal mine company (Vietnam) for their support in collecting the dataset for this study. The authors also extend heartfelt gratitude to Hanoi University of Mining and Geology (Vietnam), Pukyong National University (South Korea), and Curtin University (Australia) for their invaluable cooperation.

References

[1] Z. Wang, et al., Annual dust pollution characteristics and its prevention and control for environmental protection in surface mines, Sci. Total Environ. 825 (2022) 153949.
[2] I.M. Jiskani, et al., Green and climate-smart mining: a framework to analyze open-pit mines for cleaner mineral production, Resour. Policy 71 (2021) 102007.
[3] A.A.P. De Torres, B.I.M. Andres, Mining and Climate Change: Revisiting the Alternative Minerals Management Bill in the Age of Climate Crisis, 93, Phil. LJ, 2020, p. 1037.
[4] S. Abdul-Wahab, F. Marikar, The environmental impact of gold mines: pollution by heavy metals, Open Eng. 2 (2) (2012) 304–313.
[5] L. Dong, et al., Some developments and new insights of environmental problems and deep mining strategy for cleaner production in mines, J. Clean. Prod. 210 (2019) 1562–1578.
[6] J. Phillips, Climate change and surface mining: a review of environment-human interactions & their spatial dynamics, Appl. Geogr. 74 (2016) 95–108.
[7] K. Mohapatra, S. Biswal, Effect of particulate matter (PM) on plants, climate, ecosystem and human health, Int. J. Adv. Technol. Eng. Sci. 2 (4) (2014) 2348–7550.
[8] F. Schulze, et al., Air quality effects on human health and approaches for its assessment through microfluidic chips, Gene 8 (10) (2017) 244.
[9] R. Chen, et al., Beyond PM2.5: the role of ultrafine particles on adverse health effects of air pollution, Biochim. Biophys. Acta (BBA) Gen. Subjects 1860 (12) (2016) 2844–2855.

[10] Q. Huang, et al., Effects of water soluble PM2.5 extracts exposure on human lung epithelial cells (A549): a proteomic study, J. Appl. Toxicol. 34 (6) (2014) 675–687.
[11] L.R. Crilley, Investigations of the Physical and Chemical Properties of Ambient Fine Particles in Urban Schools, Queensland University of Technology, 2013.
[12] Q.A. Hathaway, Environmental Regulation of the Heart: The Role of Non-Coding RNA and Epigenetics in Influencing Mitochondrial and Cellular Health, 2019.
[13] L. Li, et al., Monitoring and prediction of dust concentration in an open-pit mine using a deep-learning algorithm, J. Environ. Health Sci. Eng. 19 (2021) 401–414.
[14] Z. Wang, et al., Dust pollution in cold region surface mines and its prevention and control, Environ. Pollut. 292 (2022) 118293.
[15] Z.-M. Wang, et al., Dust reduction method based on water infusion blasting in open-pit mines: a step toward green mining, in: Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 2021, pp. 1–15.
[16] X. Lu, et al., Prediction into the future: a novel intelligent approach for PM2.5 forecasting in the ambient air of open-pit mining, Atmos. Pollut. Res. 12 (6) (2021) 101084.
[17] M. Wang, et al., Design of PM2.5 monitoring and forecasting system for opencast coal mine road based on internet of things and ARIMA mode, PloS One 17 (5) (2022) e0267440.
[18] C. Richardson, S. Rutherford, I. Agranovski, Characterization of particulate emissions from Australian open-cut coal mines: toward improved emission estimates, J. Air Waste Manage. Assoc. 68 (6) (2018) 598–607.
[19] V. Sastry, et al., Prediction and analysis of dust dispersion from drilling operation in opencast coal mines, Procedia Earth Planet. Sci. 11 (2015) 303–311.
[20] J. Zhou, et al., Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories, Saf. Sci. 118 (2019) 505–518.
[21] H. Ding, et al., Prediction of pillar stability for underground mines using the stochastic gradient boosting technique, IEEE Access 6 (2018) 69253–69264.
[22] P.K. Ganti, H. Naik, M.K. Barada, Environmental impact analysis and enhancement of factors affecting the photovoltaic (PV) energy utilization in mining industry by sparrow search optimization based gradient boosting decision tree approach, Energy 244 (2022) 122561.
[23] W. Zhang, et al., Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization, Geosci. Front. 12 (1) (2021) 469–477.
[24] R. Zhang, et al., Analysis of ground surface settlement in anisotropic clays using extreme gradient boosting and random forest regression models, J. Rock Mech. Geotech. Eng. 13 (6) (2021) 1478–1484.
[25] H. Chen, et al., Prediction of undrained failure envelopes of skirted circular foundations using gradient boosting machine algorithm, Ocean Eng. 258 (2022) 111767.
[26] A. Natekin, A. Knoll, Gradient boosting machines, a tutorial, Front. Neurorobot. 7 (2013) 21.
[27] V.K. Ayyadevara, V.K. Ayyadevara, Gradient boosting machine, in: Pro Machine Learning Algorithms: A Hands-on Approach to Implementing Algorithms in Python and R, 2018, pp. 117–134.
[28] O. Sagi, L. Rokach, Ensemble learning: a survey, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 8 (4) (2018) e1249.
[29] U. Singh, et al., A machine learning-based gradient boosting regression approach for wind power production forecasting: a step towards smart grid environments, Energies 14 (16) (2021) 5196.
[30] V.A. Dev, M.R. Eden, Formation lithology classification using scalable gradient boosted decision trees, Comput. Chem. Eng. 128 (2019) 392–404.

[31] J.P. Bharti, et al., Slope stability analysis using Rf, gbm, cart, bt and xgboost, Geotech. Geol. Eng. 39 (2021) 3741–3752.

[32] D.-S. Cao, et al., The boosting: a new idea of building models, Chemom. Intel. Lab. Syst. 100 (1) (2010) 1–11.

[33] C. Kim, T. Park, Predicting determinants of lifelong learning intention using gradient boosting machine (GBM) with grid search, Sustainability 14 (9) (2022) 5256.

[34] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. (2001) 1189–1232.

[35] S. Touzani, J. Granderson, S. Fernandes, Gradient boosting machine for modeling the energy consumption of commercial buildings, Energ. Buildings 158 (2018) 1533–1543.

[36] R. Storn, K. Price, Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces, J. Glob. Optim. 11 (4) (1997) 341.

[37] V. Kachitvichyanukul, Comparison of three evolutionary algorithms: GA, PSO, and DE, Indus. Eng. Manag. Syst. 11 (3) (2012) 215–223.

[38] A. Gogna, A. Tayal, Metaheuristics: review and application, J. Exp. Theor. Artif. Intell. 25 (4) (2013) 503–526.

[39] X. Yuan, et al., A hybrid differential evolution method for dynamic economic dispatch with valve-point effects, Expert Syst. Appl. 36 (2) (2009) 4042–4048.

[40] S. Ji, J. Karlovšek, Optimized differential evolution algorithm for solving DEM material calibration problem, Eng. Comput. (2022) 1–16.

[41] G. Sun, G. Xu, N. Jiang, A simple differential evolution with time-varying strategy for continuous optimization, Soft. Comput. 24 (2020) 2727–2747.

[42] A.K. Qin, V.L. Huang, P.N. Suganthan, Differential evolution algorithm with strategy adaptation for global numerical optimization, IEEE Trans. Evol. Comput. 13 (2) (2008) 398–417.

[43] N. Ye, et al., Radial basis function-assisted adaptive differential evolution using cooperative dual-phase sampling for high-dimensional expensive optimization problems, Struct. Multidiscip. Optim. 65 (9) (2022) 241.

[44] A. Khan, C. Niemann-Delius, A differential evolution based approach for the production scheduling of open pit mines with or without the condition of grade uncertainty, Appl. Soft Comput. 66 (2018) 428–437.

[45] K. Chen, et al., Optimization of air quantity regulation in mine ventilation networks using the improved differential evolution algorithm and critical path method, Int. J. Min. Sci. Technol. 25 (1) (2015) 79–84.

[46] W.B. de Melo, Optimization of truck allocation in open pit mines using differential evolution algorithm, Int. J. Innov. Res. 9 (8) (2021) 338–350.

[47] S.K. Das, et al., Classification of slopes and prediction of factor of safety using differential evolution neural networks, Environ. Earth Sci. 64 (2011) 201–210.

[48] P.G.A. Njock, et al., Artificial neural network optimized by differential evolution for predicting diameters of jet grouted columns, J. Rock Mech. Geotech. Eng. 13 (6) (2021) 1500–1512.

[49] S. Vardakos, M. Gutierrez, C. Xia, Parameter identification in numerical modeling of tunneling using the differential evolution genetic algorithm (DEGA), Tunn. Undergr. Space Technol. 28 (2012) 109–123.

[50] J.-S. An, et al., Analysis for applicability of differential evolution algorithm to geotechnical engineering field, J. Korean Geotech. Soc. 35 (4) (2019) 27–35.

[51] S. Das, P.N. Suganthan, Differential evolution: a survey of the state-of-the-art, IEEE Trans. Evol. Comput. 15 (1) (2010) 4–31.

[52] L. Tang, Y. Dong, J. Liu, Differential evolution with an individual-dependent mechanism, IEEE Trans. Evol. Comput. 19 (4) (2014) 560–574.

[53] K.P. Wong, Z.Y. Dong, Differential evolution, an alternative approach to evolutionary algorithm, in: *Proceedings of the 13th International Conference on, Intelligent Systems Application to Power Systems*, IEEE, 2005.
[54] A.K. Qin, P.N. Suganthan, Self-adaptive differential evolution algorithm for numerical optimization, in: 2005 IEEE Congress on Evolutionary Computation, IEEE, 2005.
[55] U.K. Chakraborty, Advances in Differential Evolution, vol. 143, Springer, 2008.
[56] P.N. Suganthan, Differential evolution algorithm: recent advances, in: *Theory and Practice of Natural Computing: First International Conference, TPNC 2012, Tarragona, Spain, October 2–4, 2012. Proceedings 1*, Springer, 2012.
[57] J. Kennedy, R. Eberhart, Particle swarm optimization, in: *Proceedings of ICNN'95-international conference on neural networks*, IEEE, 1995.
[58] K.E. Parsopoulos, M.N. Vrahatis, Particle Swarm Optimization and Intelligence: Advances and Applications, IGI Global, 2010.
[59] M.N.K. Kulkarni, et al., Particle swarm optimization applications to mechanical engineering—a review, Mater. Today Proc. 2 (4–5) (2015) 2631–2639.
[60] A. Khare, S. Rangnekar, A review of particle swarm optimization and its applications in solar photovoltaic system, Appl. Soft Comput. 13 (5) (2013) 2997–3006.
[61] A.G. Gad, Particle swarm optimization algorithm and its applications: a systematic review, Arch. Comput. Methods Eng. 29 (5) (2022) 2531–2561.
[62] M. Hajihassani, D. Jahed Armaghani, R. Kalatehjari, Applications of particle swarm optimization in geotechnical engineering: a comprehensive review, Geotech. Geol. Eng. 36 (2018) 705–722.
[63] X. Zhang, et al., Novel soft computing model for predicting blast-induced ground vibration in open-pit mines based on particle swarm optimization and XGBoost, Nat. Resour. Res. 29 (2020) 711–721.
[64] F. Yang, T. Yu, T.Q. Bui, Morphogenesis of free-form surfaces by an effective approach based on isogeometric analysis and particle swarm optimization, Structure 47 (2023) 2347–2353.
[65] A. Banks, J. Vincent, C. Anyakoha, A review of particle swarm optimization. Part I: background and development, Nat. Comput. 6 (2007) 467–484.
[66] Z.-H. Zhan, et al., Adaptive particle swarm optimization, IEEE Trans. Syst. Man Cybern. B Cybern. 39 (6) (2009) 1362–1381.
[67] E.H. Houssein, et al., Major advances in particle swarm optimization: theory, analysis, and application, Swarm Evol. Comput. 63 (2021) 100868.
[68] J.C. Vazquez, F. Valdez, P. Melin, Comparative study of particle swarm optimization variants in complex mathematics functions, Recent Adv. Hybrid Intell. Syst. (2013) 223–235.
[69] M. Isiet, M. Gadala, Self-adapting control parameters in particle swarm optimization, Appl. Soft Comput. 83 (2019) 105653.
[70] M. Protodiakonov, et al., Rock Strength Passports and Methods for Their Determination, Nauka, Moscow, 1964.