Defence Technology 52 (2025) 32–46

Contents lists available at ScienceDirect

Defence Technology

journal homepage: www.keaipublishing.com/en/journals/defence-technology

Accurate prediction of blast-induced ground vibration intensity using optimized machine learning models

Lihua Chen [a], Yewuhalashet Fissha [b,c,*], Mahdi Hasanipanah [d,e,**], Refka Ghodhbani [f,***], Hesam Dehghani [g], Jitendra Khatti [h]

[a] Chongqing Vocational Institute of Engineering, Chongqing, 402260, China
[b] Department of Electrical and Computer Engineering, National Institute of Technology, Asahikawa College, Asahikawa city, 071-8142, Japan
[c] Department of Mining Engineering, Aksum University, Aksum, 7080, Ethiopia
[d] Institute of Research and Development, Duy Tan University, Da Nang, Viet Nam
[e] School of Engineering & Technology, Duy Tan University, Da Nang, Viet Nam
[f] Center for Scientific Research and Entrepreneurship, Northern Border University, Arar, 73213, Saudi Arabia
[g] Department of Mining, Faculty of Engineering, Hamedan University of Technology, Hamedan, 65168, Iran
[h] Department of Civil Engineering, Rajasthan Technical University, Kota, 324010, India

ARTICLE INFO

Article history:
Received 5 April 2025
Received in revised form
17 May 2025
Accepted 22 June 2025
Available online 24 June 2025

Keywords:
Ground vibrations
Peak particle velocity
Machine learning
CatBoost
Nature-inspired optimization
Blasting safety

ABSTRACT

Blast-induced ground vibration, quantified by peak particle velocity (PPV), is a crucial factor in mitigating environmental and structural risks in mining and geotechnical engineering. Accurate PPV prediction facilitates safer and more sustainable blasting operations by minimizing adverse impacts and ensuring regulatory compliance. This study presents an advanced predictive framework integrating CatBoost (CB) with nature-inspired optimization algorithms, including the Bat Algorithm (BAT), Sparrow Search Algorithm (SSA), Butterfly Optimization Algorithm (BOA), and Grasshopper Optimization Algorithm (GOA). A comprehensive dataset from the Sarcheshmeh Copper Mine in Iran was utilized to develop and evaluate these models using key performance metrics such as the Index of Agreement (IoA), Nash-Sutcliffe Efficiency (NSE), and the coefficient of determination (R^2). The hybrid CB-BOA model outperformed other approaches, achieving the highest accuracy (R^2 = 0.989) and the lowest prediction errors. SHAP analysis identified Distance (Di) as the most influential variable affecting PPV, while uncertainty analysis confirmed CB-BOA as the most reliable model, featuring the narrowest prediction interval. These findings highlight the effectiveness of hybrid machine learning models in refining PPV predictions, contributing to improved blast design strategies, enhanced structural safety, and reduced environmental impacts in mining and geotechnical engineering.

1. Introduction

Blasting operations in mining and civil engineering are crucial for rock fragmentation, which facilitates subsequent processes like loading, transportation, and excavation [1]. However, while blasting is an effective technique for breaking down rock masses, it also generates significant amounts of energy that do not contribute to fragmentation, leading to unintended environmental impacts [2,3]. Research indicates that only about 30% of the energy from an explosion is actually used for breaking rock, while the remaining energy results in environmental disturbances such as ground vibration, fly rock, and back break [4]. Among these, ground vibration (GV) is particularly concerning due to its potential to cause severe damage to nearby structures, compromise the stability of slopes and benches in open-pit mines, and threaten groundwater and other critical infrastructure [5].

The severity of ground vibrations produced during blasting is affected by various factors, including the type and quantity of explosives, the properties of the rock, the distance from the blast site, and the specific design parameters of the blast [6,7]. These vibrations can lead to significant environmental and structural

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: yewuhala@asahikawa-nct.ac.jp (Y. Fissha), hasanipanahmahdi@duytan.edu.vn (M. Hasanipanah), Refka.Ghodhbani@nbu.edu.sa (R. Ghodhbani).

Peer review under the responsibility of China Ordnance Society.

https://doi.org/10.1016/j.dt.2025.06.019

challenges, including damage to buildings, destabilization of mining slopes, and harmful effects on vegetation and surrounding communities. As a result, predicting and controlling GV has become a central focus of research, leading to the development of numerous experimental, statistical, and computational methods aimed at mitigating these effects [8,9]. Among the various indicators of blast-induced vibrations, peak particle velocity (PPV) is the most widely used metric, serving as a key measure of the potential damage that ground vibrations (GV) can cause [10,11]. Accurate prediction of PPV is critical for ensuring safety and minimizing the adverse impacts of blasting operations. Traditional approaches to predicting PPV typically rely on empirical formulas derived from regression analysis of vibration data. However, these methods can be limited by site-specific factors, such as rock mass properties and terrain characteristics, leading to discrepancies between predicted and actual vibration levels [12,13].

Recent advancements in artificial intelligence (AI) and machine learning (ML) have revolutionized the prediction of blast-induced vibrations, particularly in estimating PPV. These methods offer significant advantages, including higher accuracy, the ability to incorporate diverse rock properties and blasting parameters, cost-effectiveness, and time efficiency. As AI technologies and computational power continue to evolve, increasingly sophisticated models are being developed, providing a deeper understanding of the complex interactions that influence blasting outcomes [14–17]. The integration of AI into blasting engineering research has thus become a focal point, leading to more robust and adaptable solutions for mitigating the environmental impacts of blasting operations.

Nguyen et al. [18] combined an earthworm optimization algorithm with a neuro-fuzzy system for PPV prediction, demonstrating its effectiveness. Additionally, Bui et al. [6] employed a neuro-fuzzy system optimized with a flame optimization algorithm for the same purpose, and their results confirmed its accuracy. The use of neural networks (NNs) has also been highlighted by many scholars, particularly when combined with optimization algorithms. For instance, Zhao et al. [19] predicted PPV using a combination of NN and chaos game optimization, and their results confirmed the efficacy of this hybrid approach. In another study, Hosseini et al. [13] applied extreme gradient boosting (XGBoost) and NN for PPV prediction, with their findings showing that XGBoost achieved higher accuracy.

Nguyen et al. [20,21] have developed several AI-based methods for prediction in this field. For example, they combined two AI methods—extreme learning machine (ELM) and self-organizing NN—with optimization algorithms for predicting PPV. Their studies tested various algorithms and ultimately demonstrated the effectiveness of sparrow search optimization and manta ray foraging optimization algorithms, respectively. In another study, they optimized NN using a hunger games search optimization for PPV prediction [20]. Additionally, they employed a deep neural network (DNN) and combined it with different algorithms to enhance the performance of individual models [22]. According to their results, the combination of the Harris hawks algorithm and DNN showed the best performance in PPV prediction.

In the present study, CatBoost (CB)-based models were developed to predict PPV, leveraging the advantages of CB such as its ability to handle categorical data efficiently, robustness to overfitting, and superior performance with imbalanced datasets. CB's gradient boosting mechanism also ensures faster convergence and enhanced accuracy compared to traditional ML models. To further optimize CB's predictive performance, four nature-inspired metaheuristic algorithms were integrated: Grasshopper Optimization Algorithm (GOA), Sparrow Search Algorithm (SSA), Bat Algorithm (BAT), and Butterfly Optimization Algorithm (BOA). These algorithms were chosen based on several key considerations:

(i) Their diverse exploration–exploitation strategies, which are critical for effectively navigating complex, high-dimensional search spaces during hyperparameter tuning.
(ii) Their demonstrated robustness and efficiency in addressing nonlinear and stochastic optimization problems across various engineering and scientific applications.
(iii) Their population-based structure, which is well-suited to global optimization tasks like tuning the parameters of machine learning models such as CB.

Moreover, to the best of our knowledge, these algorithms have not previously been applied in the context of predicting blast-induced ground vibrations, particularly PPV. Thus, their use in this study represents a novel methodological contribution. The integration of these optimization algorithms with CB offers a new modeling framework aimed at improving the reliability and accuracy of PPV prediction in geotechnical and mining engineering practices.

2. Field investigation

The Sarcheshmeh Copper Mine in Kerman, Iran, ranks among the largest copper deposits in the world, with an estimated reserve of 1.7 billion tons (Fig. 1). The mine annually produces approximately 25 million tons of ore and 35 million tons of waste. Its operational layout includes 12.5 m-high benches with widths of 8.75 m, interspersed with wider benches of 23.75 m every four levels. The final wall slopes range between 34° and 36°, and mine roads are 30 m wide with an 8% slope. Located 160 km southwest of Kerman, the mine sits at an elevation of 2500 m above sea level and is subject to a cold desert climate, with average yearly precipitation of 550 mm. Temperature fluctuations range from −15 °C in winter to 32 °C in summer. Geological assessments indicate that the deposit contains over 1.2 billion tons of sulfuric copper ore with a 0.7% average grade, underlain by Eocene volcanic rocks, such as trachybasalt and andesite. Blasting operations rely on ANFO explosives and a staggered blast pattern, tackling common challenges like ground vibration, air blasts, and fly rock.

To evaluate the PPV, 111 blast events were monitored, with several key parameters measured, including the number of holes per delay (NHD), burden (B), spacing (S), number of rows (NR), powder factor (PF), ANFO charge weight (W), and point load index (PLI). The distance (Di) between the monitoring station and the blast point was determined using GPS. These eight parameters were utilized as input variables for predicting and assessing the PPV. The PPV values themselves were recorded by a calibrated seismograph installed at the monitoring station.

The selection of these input variables was informed by both engineering experience and theoretical foundations rooted in stress wave propagation. Specifically, B, S, and NR directly affect the distribution and reflection of stress waves within the rock mass. W and PF influence the magnitude and energy of the explosion, which are primary sources of ground vibration. The PLI provides a proxy for the rock's mechanical resistance to stress wave transmission. The NHD affects the simultaneous release of energy, and the Di governs attenuation effects, where PPV typically decreases with increasing distance due to energy dispersion. This combination of variables was chosen to capture both the source-related and path-related characteristics of blast-induced ground vibrations, enabling more robust and realistic predictive modeling.

The descriptive statistics of all the parameters are presented in Table 1. The data summarizes the minimum, maximum, quartiles, median, mean, and standard deviation values observed across 111

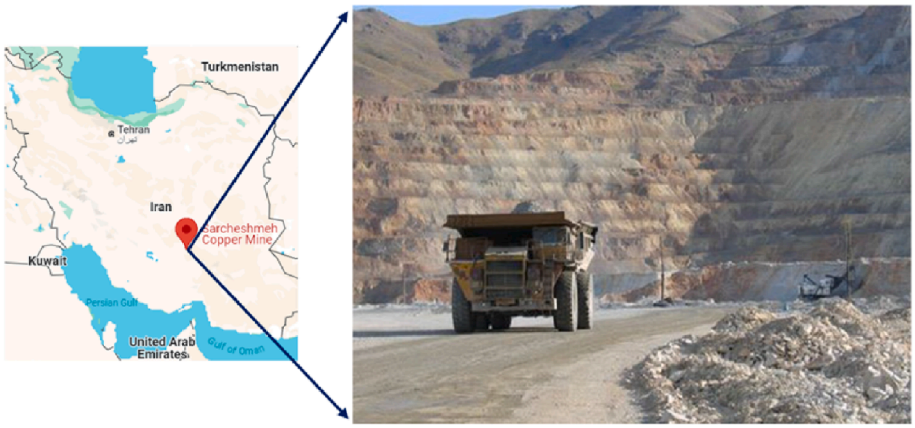

Fig. 1. A view of Sarcheshmeh Copper Mine [23].

blast events.

The violin plots for all parameters provide a comprehensive visual representation of the data distribution across the measured variables, including NHD, B, S, NR, PF, W, PLI, Di, and PPV, as shown in Fig. 2. These plots reveal not only the central tendency but also the distribution shape, variability, and presence of any potential outliers. In other words, the violin plots highlight the diversity within the dataset, with several parameters exhibiting skewness or bimodal distributions. These visualizations complement the descriptive statistics by offering deeper insights into the underlying patterns and variability across the measured parameters. The following observations can be made for each parameter based on Fig. 2:

- B: The distribution shows a wide spread with a relatively symmetrical distribution around the median. Clustering around the upper quartiles suggests that higher B values are more common in the dataset.
- S: Exhibits a bimodal distribution, indicating two distinct groups of values. The narrow range between the 25th and 75th percentiles show limited variation in S for most events.
- NHD: Displays a wider distribution with a long tail toward higher values, reflecting variability in NHD. There is a slight skewness towards larger values, suggesting that higher NHD occurrences are less frequent but still present.
- NR: The concentration of data around the middle quartiles with tapering on both sides suggests that NR clusters around a central range, with fewer extreme values.
- PF: Shows a moderately wide distribution with a slight right skewness. Most values are concentrated around the media, with a gradual decrease toward higher PF values.
- W: Exhibits a uniform distribution with a long tail on the higher end, indicating that larger charge weights are less frequent. The large interquartile range (IQR) suggests variability in charge weights across blasts.
- PLI: Displays a significantly narrower distribution compared to other parameters, indicating lower variability in rock strength values, with most data points falling within a narrow range, suggesting consistency in the rock's load-bearing capacity.
- Di: The distribution shows a widespread, with measurements concentrated in the lower to mid-range distances. Higher distances are less frequent.
- PPV: The distribution is slightly skewed towards lower values, indicating that most blast-induced vibrations measured fall in the lower range, with few instances of higher PPV values. This is typical of blast events, where excessive vibration is uncommon.

Fig. 3 presents the Pearson correlation coefficient matrix, illustrating the relationships between the inputs and PPV. The correlation coefficients span from −1 to 1, with values approaching 1 signifying a strong positive correlation, while those nearing −1 represent a strong negative correlation. As shown in the heatmap, Di has the most substantial negative correlation with PPV (−0.63911), implying that an increase in distance leads to a significant decrease in PPV. Additionally, PF and NR show negative correlations with PPV, although these correlations are less pronounced compared to Di. Conversely, parameters such as NHD and PLI exhibit weak or minimal correlations with PPV, suggesting that these variables have a limited impact on the fluctuations of PPV. The overall weak negative or near-zero correlations of other parameters indicate that their influence on PPV prediction is relatively small compared to Di. These findings highlight Di as the dominant factor affecting PPV in blasting operations, with other input variables playing more secondary roles in influencing the outcome.

In comparing the results of the sensitivity analysis with the Pearson correlation analysis, there are some interesting patterns regarding the influence of input parameters on PPV.

The sensitivity analysis using the R_{ij} index ranks PLI as the most influential factor on PPV with a value of 0.791, followed by PF, B, S, NHD, and others. In contrast, Di is identified as the least influential

Table 1
Descriptive statistics of all parameters employed in this study.

Statistic	NHD	B/m	S/m	NR	PF/(kg·m^{-3})	W/kg	PLI	Di/m	PPV/(mm·s^{-1})
Minimum	6.000	3.000	4.000	2.000	0.116	1332	1.000	133.020	1.604
Maximum	32.000	7.500	11.000	7.000	0.226	10985	8.900	2845.020	42.530
1st Quartile	10.500	7.500	9.500	3.000	0.164	4318	7.160	626.440	4.303
Median	13.000	7.500	10.000	4.000	0.167	5122	7.470	934.940	6.403
3rd Quartile	16.000	7.500	10.500	5.000	0.180	7525	7.740	1333.620	9.207
Mean	14.441	7.050	9.369	4.027	0.171	5742.982	7.485	1050.579	7.702
St. deviation	5.579	1.134	1.724	1.179	0.020	2295.716	1.095	568.954	5.803

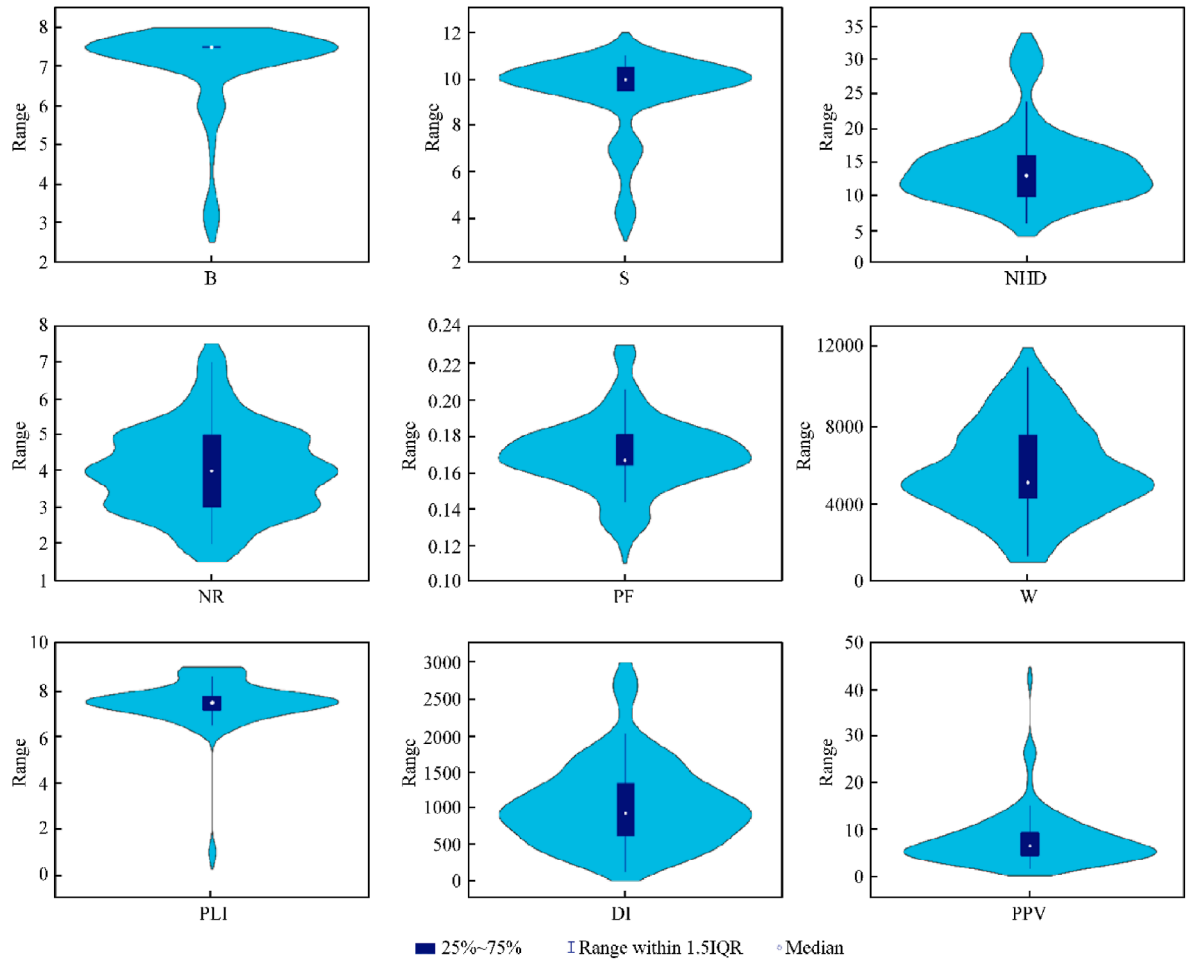

Fig. 2. Violin plots for all parameters employed in this study.

factor in the sensitivity analysis. However, the Pearson correlation matrix provides a somewhat different perspective. While PLI shows a relatively weak correlation with PPV (R = 0.00269), Di exhibits a stronger negative correlation with PPV (−0.63911), suggesting that the distance between the blast and the measurement point significantly affects PPV, in contrast to its minimal impact according to the sensitivity analysis.

This discrepancy between the sensitivity analysis and the correlation matrix could be due to the different methodologies used. The sensitivity analysis focuses on the overall effect of the parameters on the prediction model, while the Pearson correlation captures the linear relationships between the variables. Thus, while PLI may be more critical to the prediction model's performance, Di plays a crucial role in directly affecting the PPV values, as indicated by the correlation analysis. These complementary insights emphasize that multiple approaches are essential for understanding the complex relationships between input parameters and PPV in blasting operations.

3. Methods

In the present study, the CB methods is integrated with four optimization algorithms including GOA, BAT, BOA, and SSA. In the following sections the aforementioned method and algorithms are briefly explained.

3.1. CB

CB (Categorical Boosting) is a ML algorithm based on gradient boosting over decision trees, specifically designed to handle categorical features efficiently [24]. Unlike traditional tree-based methods, CB requires minimal data preprocessing, such as converting categorical data into numerical forms, making it highly efficient for datasets with numerous categorical features [25]. It operates by iteratively constructing decision trees to minimize the prediction error. The boosting process refines each tree by learning from the errors made by the previous tree, improving prediction accuracy over time [26].

One of the significant advantages of CB is its ability to handle noisy data, overfitting, and high-cardinality categorical features. It also offers computational efficiency, making it suitable for large-scale datasets [27,28]. The key control parameters in CB include the learning rate, depth of the trees, number of iterations, and regularization parameters to avoid overfitting [29]. Equation for updating the model's weights during training:

$$F_m(x) = F_{m-1}(x) + \gamma h_m(x) \tag{1}$$

where $F_m(x)$ is the prediction at the m-th iteration, $h_m(x)$ is the weak learner (decision tree), and γ is the learning rate [24]. Fig. 4 presents a diagram of CB structure.

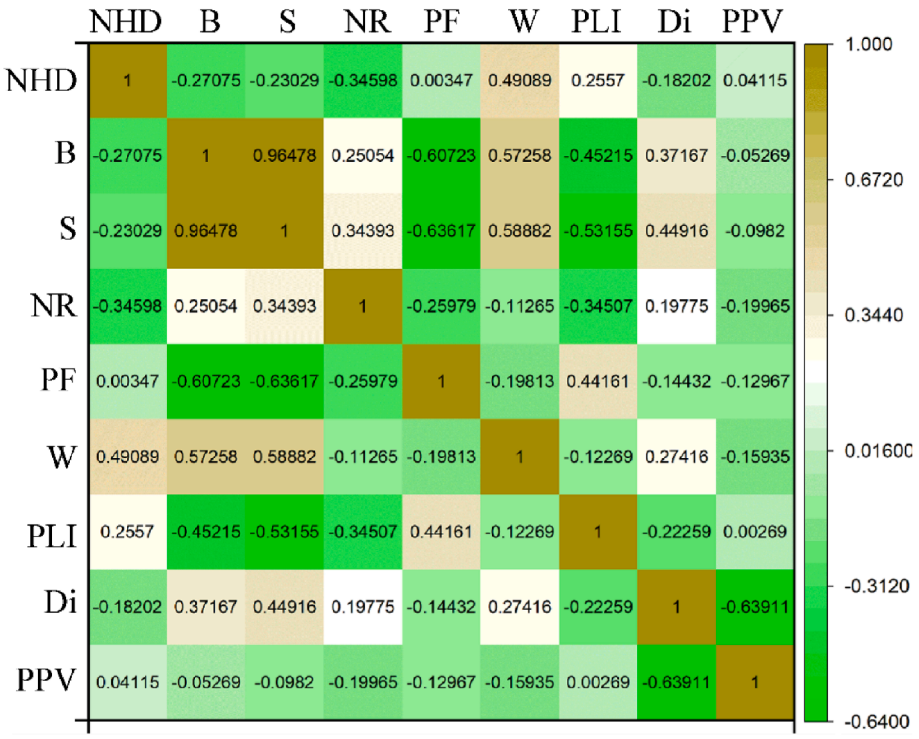

Fig. 3. Heatmap illustrating the relationships between the input parameters and PPV.

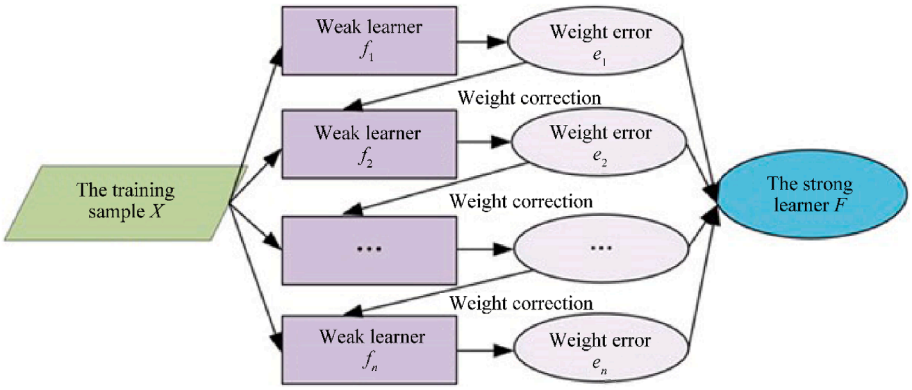

Fig. 4. A diagram of CB structure [30].

3.2. GOA

The GOA is inspired by the swarming behavior of grasshoppers, mimicking how they move between exploration and exploitation [31]. The main idea is to search globally for the optimal solution in the early stages and then refine the search locally as the solution converges [32,33]. Grasshoppers move in a multi-dimensional search space, adjusting their positions based on their distance from neighbors and global attractors.

GOA's strengths lie in its robust ability to balance exploration and exploitation, avoiding local optima while converging on a global solution [34]. Its key control parameters include the number of grasshoppers, the decreasing factor for step size, and the number of iterations [31,35].

Equation for grasshopper position update:

$$X_i = S\left(\sum_{j=1}^{n} \frac{d_{ij}}{\|d_{ij}\|} r_j - e\right) + T \tag{2}$$

where X_i is the position of the i-th grasshopper, d_{ij} is the distance between grasshoppers, and T is the target position [31]. Fig. 5 illustrates the behavior of grasshoppers in the GOA, showing the balance between attraction and repulsion forces. Grasshoppers outside the "comfort zone" are attracted toward others (green arrows), while those insides are repelled (yellow arrows), maintaining a dynamic equilibrium essential for the optimization process [36].

3.3. BAT

The BAT is a metaheuristic optimization algorithm inspired by the echolocation behavior of bats [37]. Bats emit sound pulses and

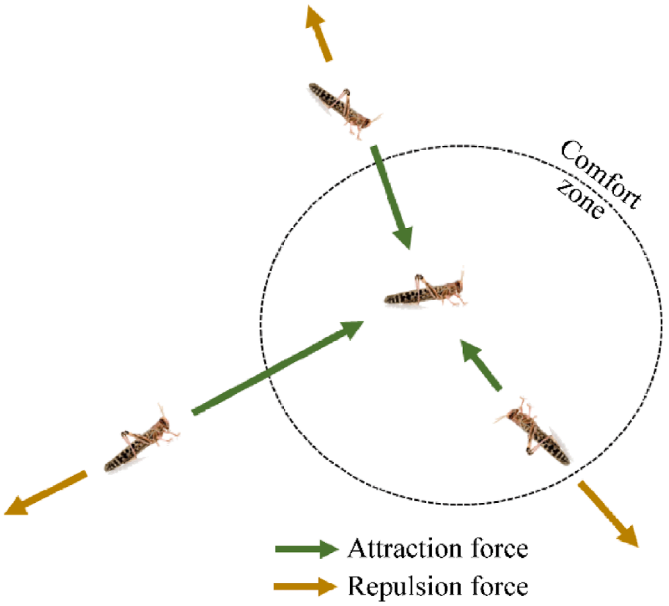

Fig. 5. Behavior of grasshoppers in the GOA.

use the echoes to detect objects and prey, which is analogous to the search for optimal solutions in a problem space. BAT balances exploration and exploitation by adjusting the frequency, loudness, and pulse rate of the bats based on their proximity to the target solution [38].

BAT's advantages include simplicity, flexibility, and strong convergence capabilities in both continuous and discrete optimization problems [39]. The key control parameters include frequency (f), pulse rate (r), loudness (A), and velocity (v) [37]. Equation for velocity and position updates:

$$v_i^{t+1} = v_i^t + (X_i - X_g) f_i \tag{3}$$

$$X_i^{t+1} = X_i^t + v_i^{t+1} \tag{4}$$

where v_i is the velocity, X_i is the position, f_i is the frequency, and X_g is the global best position [4].

3.4. BOA

The BOA is inspired by the foraging behavior of butterflies, utilizing their sense of smell to locate the most promising flowers (solutions) [40]. The search process involves a balance between local exploitation and global exploration, with butterflies attracted to global best solutions while maintaining diversity through random movements [41]. BOA's key strength is its ability to converge quickly while maintaining solution diversity to avoid local optima. Control parameters include sensory modality, stimulus intensity, and switching probability [40]. Equation for movement update:

$$X_i^{t+1} = X_i^t + C \cdot I \cdot (X_g - X_i) \tag{5}$$

where I is the stimulus intensity, C is the modality, and X_g is the global best position [40].

3.5. SSA

The SSA is a population-based optimization technique inspired by the foraging and anti-predator behavior of sparrows [42]. In SSA, sparrows either search for food (exploration) or guard against predators (exploitation). A portion of the population (scouts) searches for new areas, while the remaining population exploits known areas [43]. SSA's advantage lies in its ability to efficiently balance exploration and exploitation through adaptive behavior, making it highly effective for complex optimization tasks. Key control parameters include the number of sparrows, the vigilance rate, and the number of iterations [42]. Equation for position update:

$$X_i^{t+1} = X_i^t + v \cdot (X_{\text{best}} - X_i) \tag{6}$$

where X_i is the position of the i-th sparrow, and X_{best} is the best solution found [42]. Fig. 6 explains the SSA, where sparrows are categorized as producers (leaders) and scroungers (followers) that compete for resources. Producers guide the search direction (a), while scroungers follow or disperse when they cannot find resources (b). In aversion behavior, sparrows sense danger and move away from it, with producers being the first to react (c), and stragglers (outliers) lagging behind and adjusting their positions to avoid threats (d). This reflects the balance of exploration and exploitation in the SSA.

4. Development of hybrid methods

This study aims to evaluate the performance of optimized CB models, including CB, CB-BAT, CB-BOA, CB-GOA, and CB-SSA, for predicting PPV. A total of 111 datasets were used, featuring eight input parameters: NHD, B, S, NR, PF, W, PLI, and Di, with PPV as the target output parameter. Of the 111 datasets, 89 (80%) were allocated for training, while 22 (20%) were reserved for testing. In the initial data preprocessing step, we performed an outlier analysis using the Interquartile Range (IQR) method. Based on Table 1 (descriptive statistics of all parameters employed in this study), the lower and upper bounds were calculated for each parameter as Q1 – (1.5 × IQR) and Q3 + (1.5 × IQR), respectively. The results of this analysis, presented in Table 2, confirm that all data points fall within the calculated bounds, indicating the absence of significant outliers. This ensures the dataset remains robust and free from potential biases, thus facilitating reliable and unbiased model training.

Following outlier detection, normalization was applied to the datasets. Normalization is essential to ensure that all input parameters are on the same scale, preventing variables with larger magnitudes from dominating the model training process. This step improves model performance by enhancing numerical stability and optimizing convergence during training. Additionally, this study employs the Cross-Validation (CV) technique to ensure model robustness and prevent overfitting. Cross-validation is a method where the dataset is split into multiple subsets (folds), and the model is trained on a portion of the data while the remaining portion is used for testing [45]. This process is repeated several times to ensure that every data point is used for both training and validation. The advantage of cross-validation lies in its ability to provide a more reliable estimate of model performance, as it reduces bias and variance that can occur when the model is trained on a single dataset split. In this study, we employed a 5-fold cross-validation method, dividing the dataset into five equal sections. The model was trained on four of these sections and validated on the fifth, with this process being repeated five times. This method enhances the generalization capability of the model, leading to more accurate and reliable predictions. As outlined in previous sections, various control parameters must be determined when modeling using the CB and hybrid CB models. To optimize the performance of the models, several control parameters were

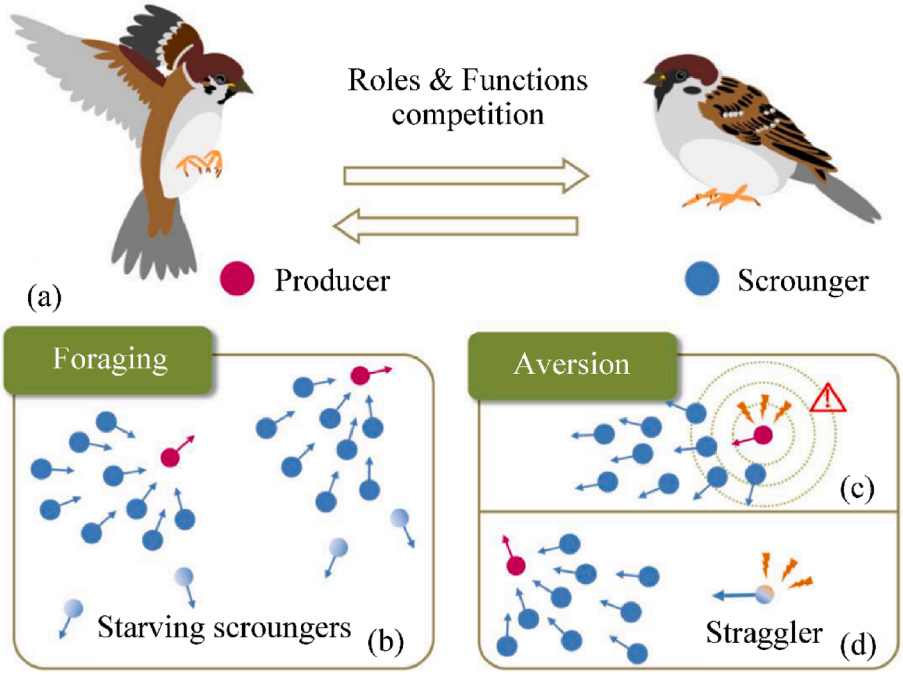

Fig. 6. Schematic representation of the SSA, illustrating the roles and behaviors of sparrows in the optimization process [44].

Table 2
Lower and Upper bounds for outlier detection based on IQR method.

Statistic	NHD	B	S	NR	PF	W	PLI	Di	PPV
IQR	5.500	0.000	1.000	2.000	0.016	3207.000	0.580	707.180	4.904
Lower bound	2.250	7.500	8.000	0.000	0.140	−492.500	6.290	−434.330	−3.054
Upper bound	24.250	7.500	12.000	8.000	0.204	12335.500	8.610	2394.390	16.564

tested, and their ranges are provided below:

CB Model:

- Learning rate: Tested in the range of 0.05–0.3.
- Depth: Tested in the range of 2–10.
- Iterations: Tested in the range of 100–1000.

Note: A smaller learning rate is recommended for a larger number of iterations to ensure stability. The depth parameter reflects the complexity of the problem: smaller values help avoid overfitting, while larger values capture more intricate patterns.

GOA:

- Number of grasshoppers (population size): Tested in the range of 5–100.
- Max iterations: Tested in the range of 100–1000.
- C_{max} and C_{min}: C_{max} was tested in the range of 1–5, while C_{min} was tested in the range of 0.001–0.01.

Note: Larger population sizes and more iterations generally yield better results but come with higher computational costs. The C_{max} and C_{min} parameters significantly impact convergence and should be fine-tuned through trial and error.

BA:

- Number of bats (population size): Tested in the range of 20–200.
- Max iterations: Tested in the range of 100–1000.
- Loudness (A): Tested in the range of 0–2.
- Pulse rate (r): Tested in the range of 0–1.
- Frequency (f): Tested in the range of 0–2.

Note: Pulse rate and loudness should be adjusted as the search progresses to prevent premature convergence. Higher pulse rates enhance global search capabilities, while higher loudness values explore broader areas in the solution space.

BOA:

- Number of butterflies (population size): Tested in the range of 20–200.
- Max iterations: Tested in the range of 100–1000.
- Switch probability: Tested in the range of 0–1.
- Sensory modality: Tested in the range of 0–0.1.
- Power exponent: Tested in the range of 0–1.

Note: The switch probability controls the balance between exploration and exploitation, where higher values promote global exploration. The sensory modality and power exponent should be tuned to refine local searches based on the problem characteristics.

SSA:

- Number of sparrows (population size): Tested in the range of 20–200.
- Max iterations: Tested in the range of 100–1000.
- Warning value: Tested in the range of 0.5–0.9.
- Alert threshold (PD): Tested in the range of 0.1–0.3.

Note: The warning value and alert threshold are key to balancing exploration (searching for new solutions) and exploitation (refining existing solutions). Proper tuning of these values is crucial to avoid excessive local optimization or premature convergence.

In this study, these control parameters were thoroughly tested, and the most optimal values were selected for each model. Achieving a balance between exploration and exploitation was vital, as adjusting parameters like the pulse rate in BA, C_{min} in GOA, and switch probability in BOA can help maintain this balance and improve the model's overall performance.

Based on the obtained results, the optimal control parameters for the CB model and its hybrid variations were determined as follows: for the CB model, the selected learning rate, depth, and iterations were 0.269, 4, and 680, respectively. For the hybrid models, the following parameters were chosen:

- CB-GOA: learning rate = 0.229, depth = 3, iterations = 680,
- CB-BAT: learning rate = 0.249, depth = 3, iterations = 680,
- CB-BOA: learning rate = 0.264, depth = 4, iterations = 680,
- CB-SSA: learning rate = 0.278, depth = 2, iterations = 680.

Additionally, the control parameters for each optimization algorithm associated with the CB-BAT, CB-BOA, CB-GOA, and CB-SSA models are presented in Table 3. These parameters were fine-tuned through extensive testing to achieve an optimal balance between model accuracy and computational efficiency. The selected values account for the complexity of the optimization problem and were carefully chosen to ensure robust performance across the models. The fine-tuning of these parameters significantly contributes to the improved prediction accuracy of the models.

It is important to emphasize that the selection of these parameters was guided not only by prior studies but also through systematic trial-and-error procedures and sensitivity analyses, ensuring a balance between prediction accuracy and computational efficiency. Moreover, the selection of learning rate, depth, and iteration values was carefully coordinated with the characteristics of each optimization algorithm to prevent overfitting while promoting model generalization. Future studies may benefit from incorporating more advanced hyperparameter optimization techniques, such as, Bayesian Optimization or Genetic Algorithms, to further enhance tuning robustness and model transferability.

Table 3
Values of control parameters associated with the CB-BAT, CB-BOA, CB-GOA, and CB-SSA models.

Model	Control parameters		Value
CB-GOA	Number of grasshoppers		10
	Max iterations		100
	C_{max}		4
	C_{min}		0.004
CB-BAT	Number of bats		20
	Max iterations		100
	Loudness		0.5
	Pulse rate		0.5
	Frequency	f_{max}	2
		f_{min}	0
CB-BOA	Number of butterflies		40
	Max iterations		200
	Switch probability		0.6
	Sensory modality		0.01
	Power exponent		0.2
CB-SSA	Number of sparrows		30
	Max iterations		200
	Warning value		0.7
	Alert threshold		0.2

For a clearer understanding, the development process of the predictive models is illustrated in Fig. 7. This diagram provides a step-by-step visualization of the workflow, beginning with data acquisition and pre-processing, followed by model development using CB and hybrid optimization algorithms (GOA, BA, BOA, and SSA). The process concludes with model evaluation, selection of the best-performing model, and conducting sensitivity analysis to identify the most significant input parameters influencing PPV prediction.

5. Results and discussion

5.1. Analysis of the results

In this study, five different ML models were developed to predict PPV. The initial model was based on the CB algorithm, and subsequently, four optimization algorithms, BAT, BOA, GOA, and SSA, were used to enhance the performance of the CB model. These optimizations resulted in the development of four additional models: CB-BAT, CB-BOA, CB-GOA, and CB-SSA. This section provides a comparative analysis of the performance of these five models using various evaluation metrics. The models were assessed using several performance indicators [46–58]: Index of Agreement (IoA), Root Mean Square Error (RMSE), Pearson Correlation (PC), Nash-Sutcliffe Efficiency (NSE), Relative Absolute Error (RAE), and Mean Absolute Percentage Error (MAPE). Table 4 presents a comprehensive comparison of these performance metrics for the five predictive models: CB, CB-BAT, CB-BOA, CB-GOA, and CB-SSA. Among the five models, CB-BOA consistently outperforms the others across most metrics. It exhibits the highest IoA at 0.997, the lowest RMSE of 0.347, and the best PC at 0.995, reflecting its excellent prediction accuracy and strong correlation with observed values. Furthermore, it achieves the highest NSE (0.988), the lowest RAE (0.086), and a MAPE of 5.0%, highlighting its minimal prediction error and robust performance. The CB-BAT model follows closely behind, with an IoA of 0.994, an RMSE of 0.538, and a PC of 0.988. It also demonstrates strong predictive capabilities, with an NSE of 0.975, an RAE of 0.119, and a slightly better MAPE (4.8%) than CB-BOA, indicating a marginally better fit in terms of absolute percentage errors but overall, slightly lower performance compared to CB-BOA. On the other hand, the base CB model performs the worst across all metrics, with an IoA of 0.953, a relatively high RMSE of 2.161, and the lowest PC of 0.915. Its MAPE of 11.8% indicates substantial errors in its predictions, making it the least reliable model in the study.

The CB-GOA and CB-SSA models exhibit moderate performance, with CB-GOA slightly outperforming CB-SSA in most categories. Both models provide more accurate predictions than the base CB model, but they fall short of the high standards set by CB-BOA and CB-BAT. As a result, CB-BOA stands out as the most reliable and accurate model, followed closely by CB-BAT, with both models significantly outperforming the other models in terms of prediction accuracy and consistency. The results presented in Table 4 also highlight the ranking of each model based on its performance metrics. According to the table, CB-BOA achieved the highest accuracy, with a total rank score of 29, making it the most accurate model. It was followed by CB-BAT, CB-GOA, CB-SSA, and CB, which had total rank scores of 25, 16, 14, and 5, respectively.

For a clearer visual representation, Fig. 8 illustrates the rank assigned to each model concerning their performance metrics, including MAPE, RAE, Nash-Sutcliffe, PC, RMSE, and IoA. The different colored sections for each model represent the performance across these metrics. From the figure, it is evident that CB-BOA has the highest rank, followed closely by CB-BAT. This visual confirms that CB-BOA is the most optimal model, offering the best

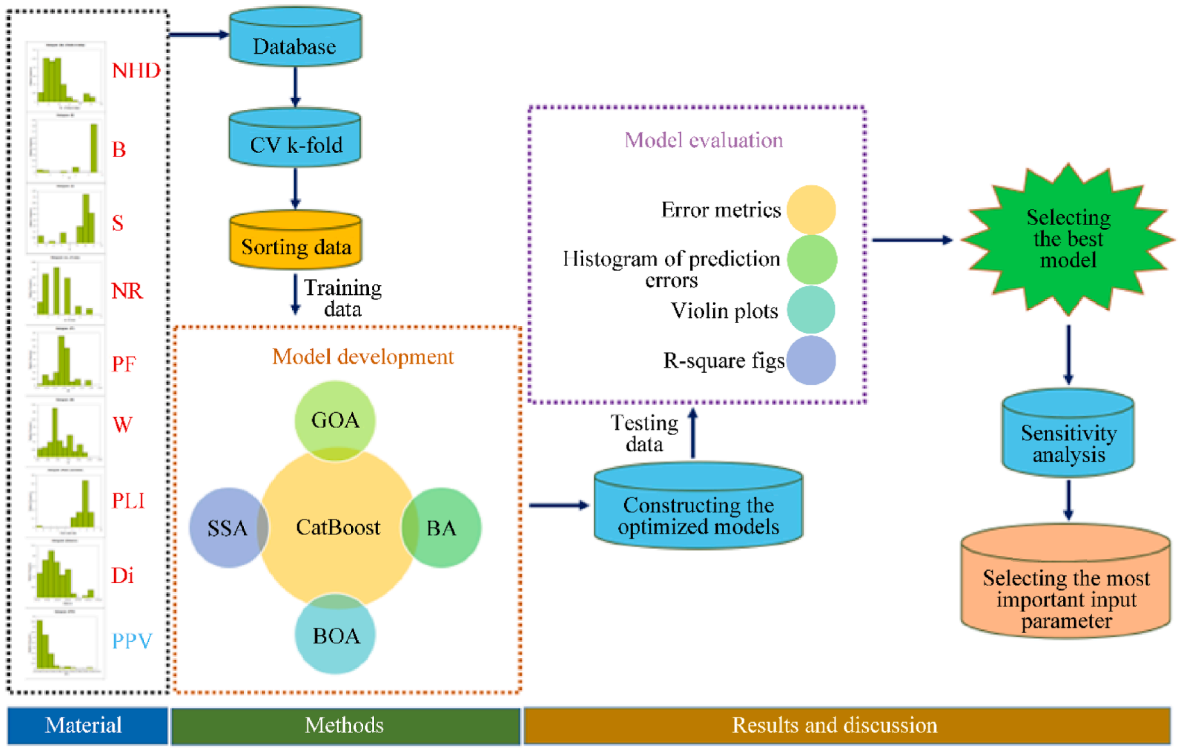

Fig. 7. A systematic view of the applied process for PPV prediction.

Table 4
Performance comparison of developed models based on key evaluation metrics.

Performance metrics	Prediction Models									
	CB		CB-BAT		CB-BOA		CB-GOA		CB-SSA	
	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank
IoA	0.953	1	0.994	4	0.997	5	0.975	2	0.986	3
RMSE	2.161	1	0.538	4	0.347	5	1.324	3	1.398	2
PC	0.915	1	0.988	4	0.995	5	0.977	3	0.974	2
NSE	0.837	1	0.975	4	0.988	5	0.921	2	0.947	3
RAE	0.253	1	0.119	4	0.086	5	0.179	3	0.198	2
MAPE	11.8	1	4.8	5	5	4	5.8	3	10.3	2

predictive performance in comparison to the other models.

Fig. 9 presents regression plots comparing the measured and predicted PPV using five models: CB, CB-BAT, CB-BOA, CB-GOA, and CB-SSA. Each plot shows the predicted values on the x-axis and the measured values on the y-axis, with fitted regression lines for the model predictions along with 95% confidence intervals for both the mean and the observations.

- CB Model (R^2 = 0.837): The first plot shows a moderate correlation between the measured and predicted PPV values, indicated by an R^2 value of 0.837. The confidence intervals are broader compared to the other models, reflecting less accurate predictions.
- CB-BAT Model (R^2 = 0.976): This model shows a much stronger fit with an R^2 value of 0.976, suggesting highly accurate predictions. The confidence intervals are narrower, indicating greater precision in the model's performance.
- CB-BOA Model (R^2 = 0.989): This model exhibits the highest correlation (R^2 = 0.989) with an almost perfect fit between measured and predicted values. The narrow confidence intervals further support its robustness in predicting PPV.
- CB-GOA Model (R^2 = 0.955): While the fit is strong, with an R^2 of 0.955, the CB-GOA model does not perform as well as CB-BOA or CB-BAT. The confidence intervals are wider than in the CB-BAT model but still relatively narrow, showing good prediction reliability.
- CB-SSA Model (R^2 = 0.949): This model also demonstrates a strong correlation (R^2 = 0.949), although slightly weaker compared to CB-BOA and CB-BAT. The performance is still highly acceptable, with relatively narrow confidence intervals.

As a comparison, the CB-BOA and CB-BAT models are the best performers, with R^2 values close to 1, suggesting that these models are nearly perfect in predicting PPV. The CB model is the least accurate, as reflected by the lower R^2 value and broader confidence intervals. CB-GOA and CB-SSA models perform well but do not match the exceptional accuracy of CB-BOA and CB-BAT.

Another form of evaluation is provided by the histogram in Fig. 10, which illustrates the distribution of errors between the predicted and measured values for the five model paradigms: CB and its hybrid variations (CB-BAT, CB-BOA, CB-GOA, and CB-SSA). The error distribution patterns reveal that CB-BAT and CB-GOA

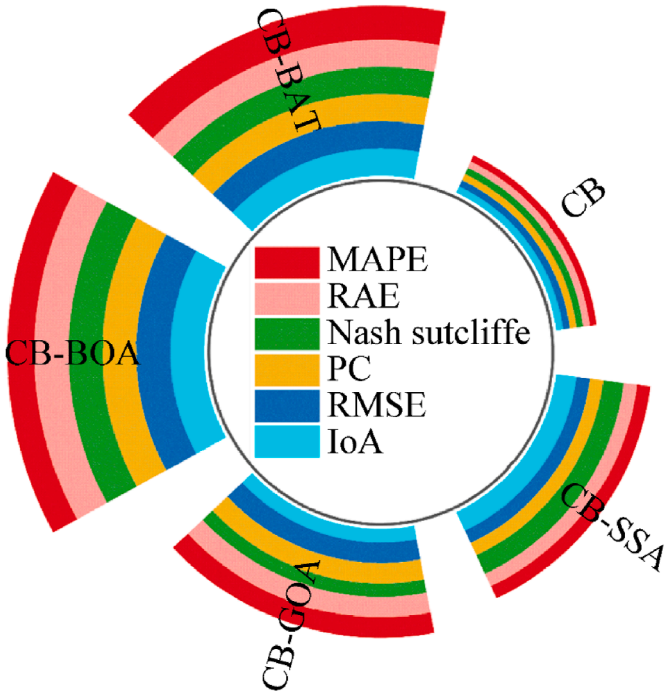

Fig. 8. Comparing the performance of models based on their ranks.

outperform the others by having errors tightly clustered around zero, indicating their predictions are more consistent and accurate. This concentration reflects their superior ability to minimize deviations. Conversely, CB and CB-SSA display wider error distributions, with noticeable outliers, suggesting that their predictions are less reliable and prone to greater inaccuracies. The performance of CB-BOA falls between the two extremes, showing moderate distribution error, performing better than CB and CB-SSA but not as effectively as CB-BAT and CB-GOA.

Fig. 11 presents violin plots with embedded box plots to visualize and compare the prediction performance of different models

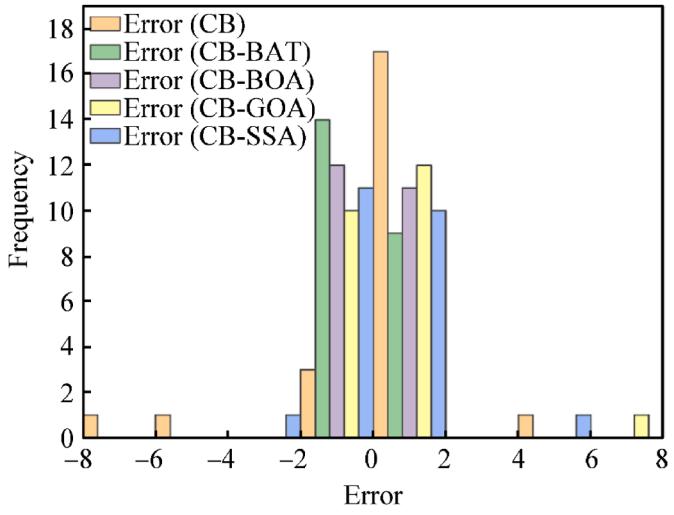

Fig. 10. Histogram of prediction errors for the CB and its hybrid models (CB-BAT, CB-BOA, CB-GOA, and CB-SSA).

(CB, CB-BAT, CB-BOA, CB-GOA, and CB-SSA) against the measured (observed) values. The violin plots highlight the distribution of the data, while the box plots provide details about the interquartile range (IQR), median, and whiskers (range within 1.5 times the IQR). According to this figure, CB-BOA and CB-GOA stand out as the most optimal models due to their tighter distribution and closer alignment with the measured values, reflecting higher accuracy and consistency in predictions. Between the two, CB-BOA slightly edges out in performance. However, CB-BAT and CB-SSA, while performing moderately well, shows a wider distribution than CB-BOA and CB-GOA, indicating some variability in their predictions. CB display broader distributions with notable deviations between measured and predicted values, making them the least reliable models in this comparison.

To provide a more holistic comparison of model performance, a Taylor diagram was employed (Fig. 12). The Taylor diagram is a powerful statistical visualization tool that simultaneously displays

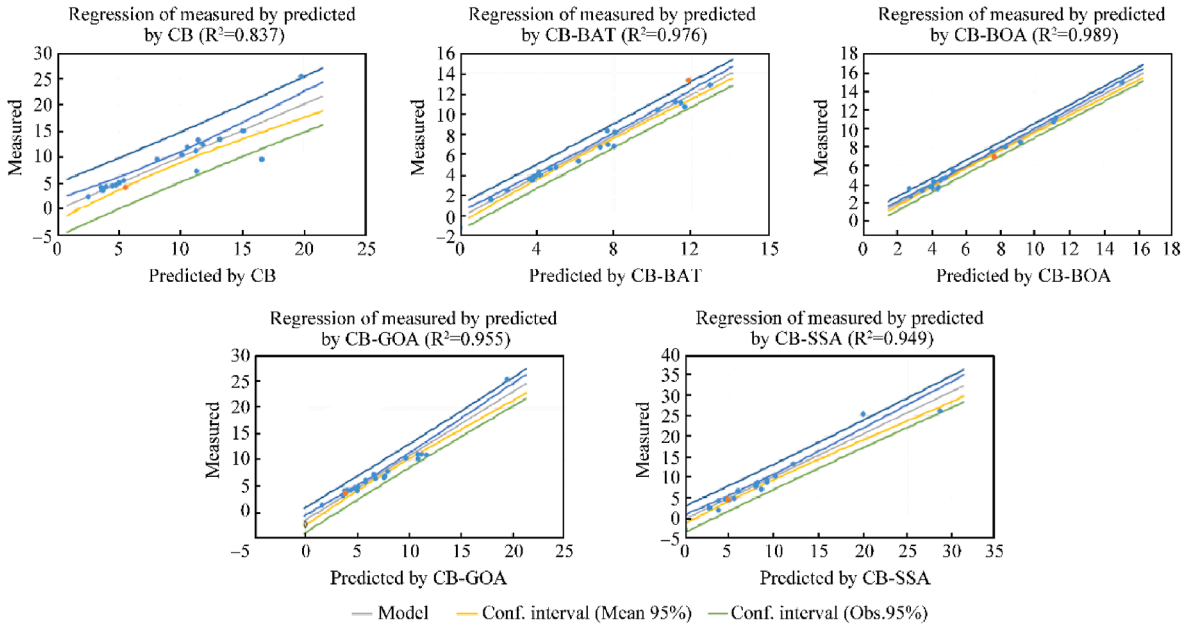

Fig. 9. Regression plots comparing measured and predicted PPV values using five models: CB, CB-BAT, CB-BOA, CB-GOA, and CB-SSA.

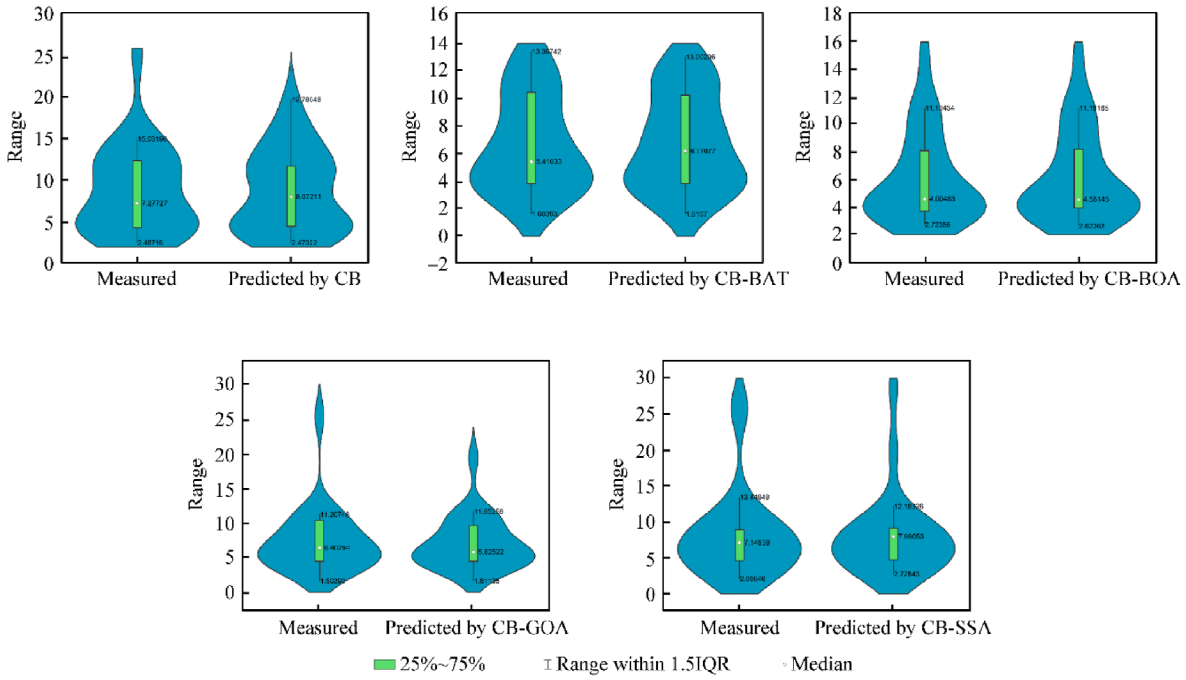

Fig. 11. Violin plots with embedded box plots comparing the prediction performance of hybrid models.

the correlation coefficient, standard deviation, and centered root mean square error (RMSE) between model predictions and actual values. It is particularly useful in environmental, geoscientific, and machine learning applications where multiple models must be compared in a compact form. As illustrated, the CB-BOA model is positioned closest to the reference point labeled "Actual" indicating superior performance across all three criteria. It is followed by CB-BAT, CB-GOA, CB-SSA, and CB, in descending order of predictive reliability. This further validates the effectiveness of optimization algorithms in enhancing the baseline CB model's accuracy.

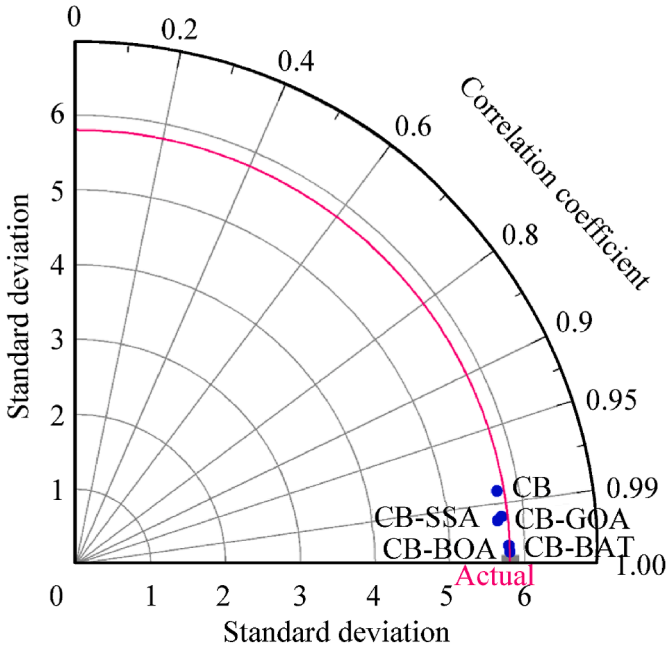

Fig. 12. Taylor diagram comparing the performance of CB-based models with actual data.

5.2. External validation using an independent dataset

To evaluate the generalization ability of the proposed CB-BOA model beyond the dataset from the Sarcheshmeh Copper Mine, we employed an independent dataset obtained from a granite quarry in Malaysia, as presented by Keshtegar et al. [59]. This external dataset includes blast design parameters and corresponding PPV values recorded under geological conditions different from those of the original study area.

In the original study, Keshtegar et al. [59] utilized a hybrid model combining Support Vector Regression (SVR) with Response Surface Methodology (RSM), which achieved an R^2 value of 0.896 for PPV prediction. To assess the CB-BOA model's transferability, we applied it to the Malaysian dataset. While the model structure remained unchanged, the BOA parameters were fine-tuned to better suit the characteristics of the new dataset. The parameter settings used were: number of butterflies = 50, maximum iterations = 200, switch probability = 0.6, sensory modality = 0.01, and power exponent = 0.2.

Under these settings, the CB-BOA model yielded an R^2 value of 0.960, significantly surpassing the performance of the SVR-RSM model. This result highlights the model's strong predictive capabilities and adaptability across different geological and operational settings.

This external validation underscores the robustness and practical applicability of the CB-BOA model, confirming its potential for reliable PPV prediction in varied mining environments.

5.3. Comparison with a traditional regression model

To provide a more comprehensive evaluation of the proposed hybrid CB models, we compared their performance with that of a traditional predictive method, Multiple Linear Regression (MLR). This comparison aims to highlight the added value and superiority of the developed hybrid models over conventional approaches in predicting PPV.

In the MLR model, the following input variables were

considered as independent parameters: NHD, B, S, NR, PF, W, PLI, and Di. PPV was set as the dependent variable. The resulting regression equation is as follows:

$$PPV = 10.266 + (0.123 \times NHD) - (1.276 \times B) + (2.682 \times S) - (1.374 \times NR) - (6.514 \times PF) - (0.0009 \times W) - (0.064 \times PLI) - (0.007 \times Di) \quad (7)$$

In addition to the MLR model, three widely used empirical models for PPV prediction were also considered:

- Ambraseys and Hendron [60].
- U.S. Bureau of Mines (USBM) [61].
- Langefors and Kihlstrom [62].

The general forms of these empirical equations are given as follows:

Ambraseys and Hendron [60]:

$$PPV = K \times \left(\frac{Di}{W^{1/3}}\right)^{K1} \quad (8)$$

USBM:

$$PPV = K \times \left(\frac{Di}{W^{0.5}}\right)^{K1} \quad (9)$$

Langefors and Kihlstrom [62]:

$$PPV = K \times \left(\frac{W^{0.5}}{Di^{3/4}}\right)^{K1} \quad (10)$$

Here, K and $K1$ are site-specific constants, which were determined using Excel through curve-fitting techniques. The derived empirical equations for the case study are:

Ambraseys and Hendron [60]:

$$PPV = 273.489 \times \left(\frac{Di}{W^{1/3}}\right)^{-0.936} \quad (11)$$

USBM:

$$PPV = 78.083 \times \left(\frac{Di}{W^{0.5}}\right)^{-0.973} \quad (12)$$

Langefors and Kihlstrom [62]:

$$PPV = 19.826 \times \left(\frac{W^{0.5}}{Di^{3/4}}\right)^{1.338} \quad (13)$$

Table 5 summarizes the statistical performance metrics obtained from the MLR and empirical models. When comparing the results in Table 5 with those in Table 4 (hybrid CB models), it becomes evident that both the regression and empirical models provide significantly lower predictive accuracy. For example, the RMSE of the MLR model was 3.981, while the hybrid CB models achieved RMSE values ranging from 0.347 to 1.398, a considerable improvement. All other performance indicators (R^2, IoA, PC, NSE, and MAPE) consistently confirmed the superiority of the hybrid models. These findings clearly demonstrate the robustness and reliability of the hybrid CB-based models for accurate PPV prediction in blasting operations, making them more effective than traditional regression and empirical techniques.

5.4. Uncertainty analysis

Uncertainty analysis is a crucial tool for enhancing model reliability and supporting informed decision-making. In this study, uncertainty analysis was conducted to evaluate the models' predictive uncertainty for PPV. Initially, the error for each sample ($PPV_{measured}$ - $PPV_{predicted}$) was calculated, followed by the determination of the mean error ($\overline{e}$) and the standard deviation of the prediction error (Se) using the following equations:

$$\overline{e} = \sum_{i=1}^{n} e_i \quad (14)$$

$$Se = \sqrt{\frac{\sum_{i=1}^{n} (e_i - \overline{e})^2}{n-1}} \quad (15)$$

A positive $\overline{e}$ indicates model overestimation, whereas a negative $\overline{e}$ suggests underestimation. Higher Se values reflect greater variability in error, implying increased uncertainty in the model's predictions. The 95% confidence interval can be calculated using $1.96 \times Se$ [63,64]. Table 6 presents the mean error ($\overline{e}$), standard deviation (Se), and the 95% confidence interval for error prediction across all models: CB, CB-BAT, CB-BOA, CB-GOA, and CB-SSA.

The uncertainty analysis highlights notable differences in the predictive performance of the various CB-based methods. Among them, CB-BOA exhibits the best performance, with the lowest mean error (−1.729), smallest standard error (1.726), and the narrowest uncertainty band (±3.382), indicating higher precision and reduced variability. In contrast, CB-GOA is the least reliable, demonstrating the highest mean error (6.716), largest standard error (6.699), and the widest uncertainty band (±13.130), which suggests a tendency for substantial overestimation and significant uncertainty. CB-BAT and CB-SSA show moderate performance, with CB-BAT slightly underestimating and CB-SSA slightly overestimating PPV, both offering narrower uncertainty bands than the original CB model. However, they do not surpass the predictive accuracy and consistency of CB-BOA. The original CB model, despite its positive mean error (0.854), suffers from relatively high uncertainty, with a standard error of 2.361 and an uncertainty band of ±4.627. In conclusion, CB-BOA is the most reliable and accurate model for predicting PPV, while CB-GOA shows the greatest uncertainty. The remaining models, CB, CB-BAT, and CB-SSA, demonstrate intermediate levels of performance and uncertainty.

5.5. Sensitivity analysis

To investigate the effects of NHD, B, S, NR, PF, W, PLI, and Di on PPV intensity, and to enhance the interpretability of machine learning models, the SHAP (SHapley Additive Explanations) method was employed. SHAP, grounded in game theory, assigns Shapley values to each feature, providing a fair and transparent representation of their contributions to model predictions. This methodology offers both local and global insights, highlighting feature importance and interactions, making it a crucial tool for interpreting complex models in rock engineering. Fig. 13 presents the SHAP summary plot for PPV prediction obtained from the most accurate model (CB-BOA). The horizontal axis represents SHAP values, indicating each feature's influence on the model's output, while the color gradient (blue to red) denotes feature values, with blue corresponding to lower values and red to higher values.

Table 5
Statistical performance of the regression and empirical models for PPV prediction.

Statistical metrics	Regression model	Ambraseys and Hendron	USBM	Langefors and Kihlstrom
R^2	0.525	0.736	0.746	0.737
RMSE	3.981	2.973	2.913	2.96
IoA	0.816	0.916	0.92	0.918
PC	0.725	0.858	0.864	0.859
NSE	0.525	0.735	0.746	0.737
MAPE	36.1%	23.6%	23.7%	25.6%

Table 6
Uncertainty predicts of PPV based on various CB methods.

Method	Mean Error ($\bar{e}$)	Standard Error (Se)	Width of uncertainty band
CB	0.854	2.361	±4.627
CB-BAT	−2.259	2.274	±4.457
CB-BOA	−1.729	1.726	±3.382
CB-GOA	6.716	6.699	±13.130
CB-SSA	2.613	2.926	±5.734

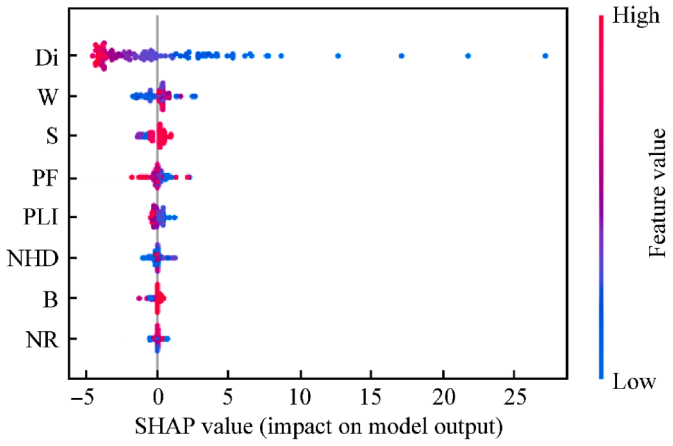

Fig. 13. SHAP summary plots: feature impacts on PPV.

According to Fig. 13, Di is the most influential factor, with higher values (blue) reducing PPV and lower values (red) increasing it, underscoring its pivotal role in controlling ground vibration. W and S also significantly impact PPV, with higher values generally correlating with increased vibration intensity. Moderate contributions are observed for PF and PLI, whereas NHD, B, and NR show relatively minor impacts. The color gradient in the plot highlights the interactions between feature values and their contributions to model predictions, emphasizing the importance of optimizing blasting parameters to effectively control vibrations. Additionally, the study utilized the importance values function from the CB-BOA model to further evaluate each parameter's contribution, as shown in Fig. 14. Among the factors influencing PPV predictions, Di was identified as the most critical, followed by W, S, PF, PLI, NHD, B, and NR in descending order of importance. This analysis emphasizes the need to prioritize parameters such as Di, W, and S in blasting design strategies to minimize PPV and mitigate ground vibrations effectively.

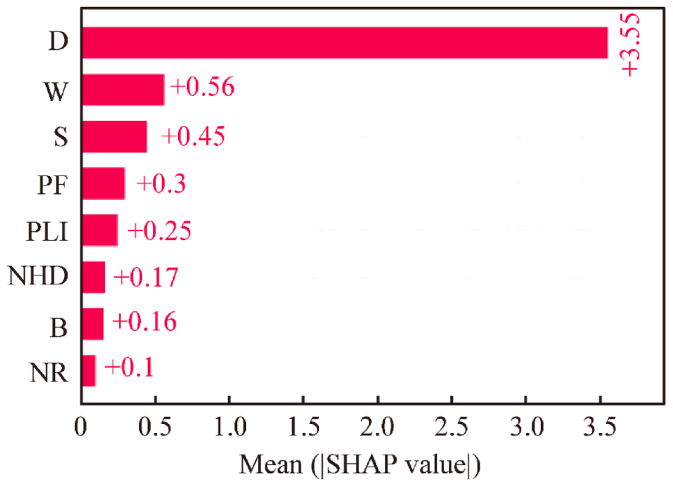

Fig. 14. Feature importance outputs by the CB-BOA.

6. Conclusions

Blast-induced ground vibration remains a critical concern in open-pit mining operations due to its impact on operational safety, structural integrity, and environmental sustainability. Accurately predicting PPV, a primary measure of ground vibration intensity, is essential for effective blast design and risk mitigation. In this study, five machine learning models were developed and evaluated: the standalone CB model and four hybrid models (CB-BAT, CB-BOA, CB-GOA, and CB-SSA), each combining CatBoost with a nature-inspired metaheuristic optimization algorithm to enhance prediction accuracy through efficient hyperparameter tuning.

- Among the models, CB-BOA consistently outperformed the others, achieving the highest R^2 value of 0.989, alongside the lowest RMSE, RAE, and MAPE. This indicates a near-perfect match between predicted and observed PPV values, highlighting the effectiveness of integrating the Butterfly Optimization Algorithm with CatBoost. CB-BAT and CB-GOA also demonstrated strong performance, while CB and CB-SSA exhibited wider error spreads and more significant outliers. These results were further validated by error distribution analysis and violin plots, which revealed a tighter clustering of predictions for CB-BOA and CB-GOA around the actual values.
- Effectiveness of Hybridization: The results clearly confirmed that hybridizing CB with nature-inspired optimization algorithms substantially enhanced predictive performance. While the standalone CB model achieved an R^2 of 0.837, the hybrid models improved significantly, reaching R^2 values from 0.949 to 0.989. This validates the capability of GOA, SSA, BAT, and BOA to boost machine learning models through effective global optimization strategies. Importantly, this study marks the first application of these metaheuristic algorithms in the context of PPV prediction, introducing a novel and robust framework for addressing blast-induced ground vibration in mining.
- To strengthen the robustness of the findings, the study also includes external validation and a comparative analysis with a classical regression model, thereby offering a broader perspective on model performance and applicability.

- Uncertainty Analysis: CB-BOA was found to be the most reliable model, exhibiting the narrowest uncertainty band and highest prediction stability. In contrast, CB-GOA showed the widest uncertainty, reflecting a higher risk of overestimation. The remaining models fell between these extremes, with moderate uncertainty levels and performance consistency.
- SHAP Analysis: The SHAP-based interpretability analysis revealed that Di was the most significant predictor of PPV, followed by W and S. These features captured both linear and non-linear interactions, providing critical insights for practical blast design aimed at minimizing vibration impact.
- Study Limitations: A noted limitation of this research is the use of fixed control parameters for the optimization algorithms, which may restrict generalizability across different datasets or mining conditions. Future work should explore adaptive or self-tuning optimization methods to further boost model robustness and adaptability.
- Future Directions: Expanding the dataset to include blast events from multiple geological and operational settings will improve model generalization. Furthermore, extending the proposed hybrid framework to predict other blast-induced outcomes, such as flyrock, air overpressure, or fragmentation, can enhance its utility and promote safer, more sustainable mining practices.

CRediT authorship contribution statement

Lihua Chen: Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Yewuhalashet Fissha:** Writing – review & editing, Validation, Project administration, Methodology, Data curation. **Mahdi Hasanipanah:** Writing – review & editing, Writing – original draft, Validation, Methodology, Data curation, Conceptualization. **Refka Ghodhbani:** Software, Project administration, Funding acquisition, Formal analysis, Data curation. **Hesam Dehghani:** Visualization, Supervision, Project administration, Methodology, Investigation. **Jitendra Khatti:** Visualization, Supervision, Project administration, Investigation, Funding acquisition.

Availability of data and materials

All data generated or analyzed during this study are available on reasonable request from the corresponding authors.

Declaration of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number "NBU-FFMRA-2025-2461-09".

References

[1] Jahed Armaghani D, Hajihassani M, Mohamad ET, Marto A, Noorani SA. Blasting-induced flyrock and ground vibration prediction through an expert artificial neural network based on particle swarm optimization. Arabian J Geosci 2024;7(12):5383–96. https://doi.org/10.1007/s12517-013-1174-0.

[2] Jamei M, Hasanipanah M, Karbasi M, Ahmadianfar I, Taherifar S. Prediction of flyrock induced by mine blasting using a novel kernel-based extreme learning machine. J Rock Mech Geotech Eng 2021;13:1438–51.

[3] Nguyen H, Bui XN, Tran QH, Nguyen HA, Nguyen DA, Hoa LTT, Le QT. Prediction of ground vibration intensity in mine blasting using the novel hybrid MARS–PSO–MLP model. Eng Comput Germany 2022;38:4007–25. https://doi.org/10.1007/s00366-021-01332-8.

[4] Yang H, Hasanipanah M, Tahir MM, Bui DT. Intelligent prediction of blasting-induced ground vibration using ANFIS optimized by GA and PSO. Nat Resour Res 2020;29:739–50. https://doi.org/10.1007/s11053-019-09515-3.

[5] Zhu C, Xu Y, Wu Y, He M, Zhu C, Meng Q, Lin Y. A hybrid artificial bee colony algorithm and support vector machine for predicting blast-induced ground vibration. Earthq Eng Eng Vib 2022;21:861–76.

[6] Bui XN, Nguyen H, Tran QH, Nguyen DA, Bui HB. Predicting blast-induced ground vibration in quarries using adaptive fuzzy inference neural network and moth–flame optimization. Nat Resour Res 2021;30:4719–34.

[7] Zhang Y, He H, Khandelwal M, Du K, Zhou J. Knowledge mapping of research progress in blast-induced ground vibration from 1990 to 2022 using CiteSpace-based scientometric analysis. Environ Sci Pollut Res 2023;30:103534–55. https://doi.org/10.1007/s11356-023-29712-1.

[8] Agrawal A, Choudhary BS, Murthy VMSR, Murmu S. Impact of bedding planes, delay interval and firing orientation on blast induced ground vibration in production blasting with controlling strategies. Measurement 2022;202:111887.

[9] Bayat P, Monjezi M, Mehrdanesh A, Khandelwal M. Blasting pattern optimization using gene expression programming and grasshopper optimization algorithm to minimise blast-induced ground vibrations. Eng Comput Germany 2022;38:3341–50.

[10] Gou Y, Shi X, Huo X, Zhou J, Yu Z, Qiu X. Motion parameter estimation and measured data correction derived from blastinduced vibration: new insights. Measurement 2019;135:213–30.

[11] Ke B, Nguyen H, Bui XN, Costache R. Estimation of ground vibration intensity induced by mine blasting using a state-of-the-art hybrid autoencoder neural network and support vector regression model. Nat Resour Res 2021;30:3853–64. https://doi.org/10.1007/s11053-021-09890-w.

[12] Hosseini S, Khatti J, Taiwo BO, et al. Assessment of the ground vibration during blasting in mining projects using different computational approaches. Sci Rep Uk 2023;13:18582. https://doi.org/10.1038/s41598-023-46064-5.

[13] Hosseini S, Pourmirzaee R, Armaghani DJ, et al. Prediction of ground vibration due to mine blasting in a surface lead–zinc mine using machine learning ensemble techniques. Sci Rep Uk 2023;13:6591. https://doi.org/10.1038/s41598-023-33796-7.

[14] Alshehri F, Shahfahad Rahman A. Optimizing machine learning models with Bayesian techniques for prediction of groundwater quality index in southwest Saudi Arabia. Earth Syst Environ 2024;8:1417–36. https://doi.org/10.1007/s41748-024-00475-0.

[15] Khan I, Umar R. Machine learning-driven optimization of water quality index: a synergistic ENTROPY-CRITIC approach using spatio-temporal data. Earth Syst Environ 2024;8:1453–75. https://doi.org/10.1007/s41748-024-00500-2.

[16] Metwaly MM, AbdelRahman MAE, Mohamed SA. A machine learning model and multi-temporal remote sensing for sustainable soil management in Egypt's Western nile delta. Earth Syst Environ 2024. https://doi.org/10.1007/s41748-024-00499-6.

[17] Swain KC, Singha C, Pradhan B. Estimating total rice biomass and crop yield at field scale using PlanetScope imagery through hybrid machine learning models. Earth Syst Environ 2024;8:1713–31. https://doi.org/10.1007/s41748-024-00481-2.

[18] Nguyen H, Choi Y, Monjezi M, Van Thieu N, Tran TT. Predicting different components of blast-induced ground vibration using earthworm optimisation-based adaptive neuro-fuzzy inference system. Int J Min Reclamat Environ 2023;38(2):99–126. https://doi.org/10.1080/17480930.2023.2254147.

[19] Zhao S, Wang L, Cao M. Chaos game optimization-hybridized artificial neural network for predicting blast-induced ground vibration. Appl Sci 2024;14(9):3759. https://doi.org/10.3390/app14093759.

[20] Nguyen H, Bui XN, Topal E. Enhancing predictions of blast-induced ground vibration in open-pit mines: comparing swarm-based optimization algorithms to optimize self-organizing neural networks. Int J Coal Geol 2023;275:104294.

[21] Nguyen H, Bui XN, Topal E. Reliability and availability artificial intelligence models for predicting blast-induced ground vibration intensity in open-pit mines to ensure the safety of the surroundings. Reliab Eng Syst Saf 2023;231:109032.

[22] Nguyen H, Bui XN, Tran QH, Nguyen DA, Hoa LTT, Le QT, Giang LH. Predicting blast-induced ground vibration in open-pit mines using different nature-inspired optimization algorithms and deep neural network. Nat Resour Res 2021;30(6):4695–717. https://doi.org/10.1007/s11053-021-09896-4.

[23] Sheykhi H, Bagherpour R, Ghasemi E, Kalhori H. Forecasting ground vibration due to rock blasting: a hybrid intelligent approach using support vector regression and fuzzy C-means clustering. Eng Comput Germany 2018;34:357–65. https://doi.org/10.1007/s00366-017-0546-6.

[24] Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. 6638–6648, https://arxiv.org/abs/1810.11363v1; 2018.

[25] Vishwakarma DK, Kumar P, Yadav KK, Ali R, Markuna S, Chauhan S, Heddam S, Kuriqi A, Srivastava A, Alam M, Vinayak V. Evaluation of CatBoost

method for predicting weekly Pan evaporation in subtropical and sub-humid regions. Pure Appl Geophys 2024;181:719–47. https://doi.org/10.1007/s00024-023-03426-4.
[26] Demir S, Sahin EK. Predicting occurrence of liquefaction-induced lateral spreading using gradient boosting algorithms integrated with particle swarm optimization: PSO-XGBoost, PSO-LightGBM, and PSO-CatBoost. Acta Geotech 2023;18:3403–19.
[27] Altaf I, Kaul A. Classifying collisions in road accidents using XGBOOST, CATBOOST and SALP SWARM based optimization algorithms. Multimed Tool Appl 2024;83:38387–410. https://doi.org/10.1007/s11042-023-16969-4.
[28] Kounlavong K, Sadik L, Keawsawasvong S, Jamsawang P. Optimized CatBoost-based soft-computing models for prediction of the ultimate bearing capacity of T-shaped footings subjected to eccentric load. Arabian J Sci Eng 2024. https://doi.org/10.1007/s13369-024-09379-7.
[29] Wang YX, Wang RR, Wang JW, Li NB, Gao HY. A rock mass strength prediction method integrating wave velocity and operational parameters based on the bayesian optimization catboost algorithm. KSCE J Civ Eng 2023;27:3148–62. https://doi.org/10.1007/s12205-023-2475-9.
[30] Chen R, Zhou L, Xiong C, Xu H, Zhang Z, He X, Dong Q, Wang C. Islanding detection method for microgrids based on CatBoost. Front Energy Res 2023;10:1016754.
[31] Saremi S, Mirjalili S, Lewis A. Grasshopper optimisation algorithm: theory and application. Adv Eng Software 2017;105:30–47.
[32] Ewees AA, Abd Elaziz M, Houssein EH. Improved grasshopper optimization algorithm using opposition-based learning. Expert Syst Appl 2018;112:156–72.
[33] Luo J, Chen H, Zhang Q, Xu YT, Huang H, Zhao XH. An improved grasshopper optimization algorithm with application to financial stress prediction. Appl Math Model 2018;64:654–68.
[34] Yildiz BS, Pholdee N, Bureerat S, Yildiz AR, Sadiq M. Sait Enhanced grasshopper optimization algorithm using elite opposition-based learning for solving real-world engineering problems. Eng Comput Germany 2022;38:4207–19.
[35] Wu L, Wu J, Wang T. Enhancing grasshopper optimization algorithm (Goa) with levy flight for engineering applications. Sci Rep Uk 2023;13:124.
[36] Melekoodappattu JG, Kadan AB, Anoop V. Early detection of breast malignancy using wavelet features and optimized classifier. Int J Imag Syst Technol 2021;31:1551–63.
[37] Yang X. A new metaheuristic bat-inspired algorithm. In: Gonzlez J, Pelta D, Cruz C, Terrazas G, Krasnogor N, editors. Nature inspired cooperative strategies for optimization (NICSO 2010). Berlin: Springer; 2010. p. 65–74.
[38] Huang GQ, Zhao WJ, Lu QQ. Bat algorithm with global convergence for solving large-scale optimization problem. Appl Res Comput 2013;30:1323–8.
[39] Dehghani H, Bogdanovic D. Copper price estimation using bat algorithm. Resour Policy 2018;55:55–61. https://doi.org/10.1016/j.resourpol.2017.10.0154.
[40] Arora S, Singh S. Butterfly optimization algorithm: a novel approach for global optimization. Soft Comput 2019;23:715–34.
[41] Tiwari A. A hybrid feature selection method using an improved binary butterfly optimization algorithm and adaptive β–hill climbing. IEEE Access 2023;11:93511–37.
[42] Xue J, Shen B. A novel swarm intelligence optimization approach: sparrow search algorithm. Syst Sci Control Eng 2020;8(1):22–34.
[43] Song C, Yao L, Hua C, Ni Q. A water quality prediction model based on variational mode decomposition and the least squares support vector machine optimized by the sparrow search algorithm (VMD-SSA-LSSVM) of the Yangtze River, China. Environ Monit Assess 2021;193:363.
[44] Zhou J, Wang Z, Li C, Qiu Y, Tao M. Metaheuristic-hybridized MLS-SVR: a multi-output model for predicting continuous shear displacements of rock fractures. PREPRINT (Version 1) available at Research Square 2023. https://doi.org/10.21203/rs.3.rs-3450476/v1.
[45] Refaeilzadeh P, Tang L, Liu H. Cross-validation. In: LIU L, ÖZSU MT, editors. Encyclopedia of database systems. Boston, MA: Springer; 2009. https://doi.org/10.1007/978-0-387-39940-9_565.
[46] Ding X, Amiri M, Hasanipanah M. Enhancing shear strength predictions of rocks using a hierarchical ensemble model. Sci Rep Uk 2024;14(1):20268.
[47] Ding X, Hasanipanah M, Rezaei M. Assessment of mechanical properties of rock using deep learning approaches. Measurement 2025;250:117180.
[48] Ding X, Jamei M, Hasanipanah M, Abdullah RA, Le BN. Optimized data-driven models for prediction of flyrock due to blasting in surface mines. Sustainability 2023;15(10):8424.
[49] Hasanipanah M, Monjezi M, Shahnazar A, Armaghani DJ, Farazmand A. Feasibility of indirect determination of blast induced ground vibration based on support vector machine. Measurement 2015;75:289–97.
[50] Hosseini S, Poormirzaee R, Hajihassani M, Kalatehjari R. An ANN-fuzzy cognitive map-based Z-number theory to predict flyrock induced by blasting in open-pit mines. Rock Mech Rock Eng 2022;55(7):4373–90. https://doi.org/10.1007/s00603-022-02866-z.
[51] Murlidhar BR, Nguyen H, Rostami J, Bui XN, Armaghani DJ, Ragam P, Mohamad ET. Prediction of flyrock distance induced by mine blasting using a novel Harris Hawks optimization-based multi-layer perceptron neural network. J Rock Mech Geotech Eng 2021;13(6):1413–27. https://doi.org/10.1016/j.jrmge.2021.08.005.
[52] Qiu Y, Zhou J. Methodology for constructing explicit stability formulas for hard rock pillars: integrating data-driven approaches and interpretability techniques. Rock Mech Rock Eng 2025;58:5475–502. https://doi.org/10.1007/s00603-025-04387-x.
[53] Samadi H, Hassanpour J, Rostami J. Assessment of the relationship between rock mass classification systems using hybrid XGBoost algorithms: insights from the mechanized tunnelling project database. Rock Mech Rock Eng 2024. https://doi.org/10.1007/s00603-024-04227-4.
[54] Trivedi R, Singh TN, Raina AK. Prediction of blast-induced flyrock in Indian limestone mines using neural networks. J Rock Mech Geotech Eng 2014;6:447–54.
[55] Yu Z, Du LF, Liu JX, Zhou J, Li CQ. Feasibility of a hybrid AHA-GPR model for predicting blasting fragmention in surface mines. Earth Sci Inform 2025;18:278. https://doi.org/10.1007/s12145-025-01817-w.
[56] Yu Z, Shi XZ, Zhang ZX, Huo XF, Zhou J, Li EM, Cai XQ. Machine-learning-aided blasted muckpile analysis: prospects for reducing ore and profit losses through developing blast techniques. Mining Metall Explor 2025;42:115–31. https://doi.org/10.1007/s42461-024-01158-0.
[57] Zhou J, Liu Z, Li C, Du K, Yang H. Cutting-edge approaches to specific energy prediction in TBM disc cutters: integrating COSSA-RF model with three interpretative techniques. Undergr Space 2025;22:241–62.
[58] Zhou J, Lu Y, Tian Q, Liu H, Hasanipanah M, Huang J. Advanced machine learning methods for prediction of blast-induced flyrock using hybrid SVR methods. Cmes-Comp Model Eng 2024;140(2):1595–617.
[59] Keshtegar B, Piri J, Abdullah RA, Hasanipanah M, Sabri MM, Le BN. Intelligent ground vibration prediction in surface mines using an efficient soft computing method based on field data. Front Public Health 2023;10:1094771.
[60] Ambraseys NR, Hendron AJ. Dynamic behavior of rock masses. In: Stagg KG, Zienkiewicz OC, editors. Rock mechanics in engineering practices. New York: Wiley; 1968. p. 203–7.
[61] Duvall WI, Petkof B. Spherical propagation of explosion generated strain pulses in rock, vol. 5483. US Bureau of Mines Report of Investigation; 1959.
[62] Langefors U, Kihlstrom B. The modern technique of rock blasting. New York: Wiley; 1963.
[63] Newcombe RG. Improved confidence intervals for the difference between binomial proportions based on paired data. Stat Med 1998;17(22):2635–50.
[64] Hasanipanah M, Bakhshandeh Amnieh H. Enhancing ground vibration prediction in mine blasting: a committee machine intelligent system optimized with metaheuristic algorithms. Nat Resour Res 2025. https://doi.org/10.1007/s11053-025-10518-6.