Arabian Journal for Science and Engineering (2024) 49:5225–5241
https://doi.org/10.1007/s13369-023-08360-0

RESEARCH ARTICLE-CIVIL ENGINEERING

A novel Hybrid XGBoost Methodology in Predicting Penetration Rate of Rotary Based on Rock-Mass and Material Properties

Mohammad Mirzehi Kalate Kazemi[1] · Zohre Nabavi[1] · Danial Jahed Armaghani[2]

Received: 17 June 2023 / Accepted: 27 September 2023 / Published online: 28 October 2023

Abstract
Predicting the drill penetration rate is a fundamental requirement in mining operations, profoundly impacting both the cost-effectiveness of mining activities and strategic mine planning. Given the intricate web of factors influencing rotary drilling performance, the necessity for advanced modeling techniques becomes evident. To this end, the hybrid extreme gradient boosting (XGBoost) was utilized to gauge the penetration rate of rotary drilling machines, utilizing random search, grid search, Harris Hawk optimization (HHO), and the dragonfly algorithm (DA) as metaheuristic algorithms. Our research draws from extensive data collected in copper mine case studies, encompassing both field and investigational data. This dataset incorporates critical material properties, such as tensile strength (TS), uniaxial compressive strength (UCS), as well as vital rock-mass characteristics including joint direction (JD), joint spacing (JS), and bit diameter (D). Our investigation evaluates the reliability of these prediction methods using various performance indicators, including mean absolute error (MAE), root mean square error (RMSE), average absolute relative error (AARE), and coefficient of determination (R^2). The multivariate analysis reveals that the HHO-XGB model stands out, demonstrating superior prediction accuracy (MAE: 0.457; RMSE: 2.19; AARE: 2.29; R^2: 0.993) compared to alternative models. Furthermore, our sensitivity analysis emphasizes the substantial impact of uniaxial compressive strength and tensile strength on the drill penetration rate. This underlines the importance of considering these material properties in mining operations. In conclusion, our research offers robust models for forecasting the penetration rate of similar rock formations, providing invaluable insights that can significantly enhance mining operations and planning processes.

Keywords Extreme gradient boosting (XGBoost) · Penetration rate · Harris Hawks optimization (HHO) · Dragonfly algorithm (DA) · Random and grid search · Rotary drilling

List of Symbols

UCS	Uniaxial compressive strength (MPa)
SSH	Shore scleroscope hardness
WOB	Weight on bit (MPa)
P	Air pressure (MPa)
FM	Flushing media
Comp St	Compressive strength (MPa)
RA	Relative abrasiveness
RQD	Rock quality designation
W	Water content (%)
R_N	Schmidt rebound hardness
REC	Recovery (%)
ρ	Density (g/cm^3)
Sj	Sievers' J value
$Is_{(50)\downarrow}$	Axial point load strength index (MPa)
GA	Genetic algorithm
ANN	Artificial neural network
MLP	Multilayer perceptron
BP	Back propagation
IWO	Invasive weed optimization
HHO	Harris hawk optimization
AdaBoost	Adaptive boosting

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

Mohammad Mirzehi Kalate Kazemi
m.mirzehi@modares.ac.ir

Zohre Nabavi
n.zohre@modares.ac.ir

[1] Department of Mining Engineering, Faculty of Engineering, Tarbiat Modares University, Tehran, Iran

[2] School of Civil and Environmental Engineering, University of Technology Sydney, Ultimo, NSW 2007, Australia

FIS	Fuzzy inference system
GS	Grid search
PR	Penetration rate (cm/min)
D	Diameter (inch)
BTS	Brazilian tensile strength (MPa)
RPM	Rotation speed (RPM)
RD_i	Rock mass drillability index
DA	Dragonfly algorithm
σ_t	Tensile strength (MPa)
σ_c	Uniaxial compressive strength (MPa)
ANI	Anisotropy index
B	Brittleness
$Is_{(50)\rightarrow}$	Diametral point load strength index (MPa)
S_{20}	Brittleness value
CAI	Cerchar abrasivity index
M	Porosity (%)
ANFIS	Adaptive neuro-fuzzy inference system
NLMR	Non-linear multiple regression
FA	Firefly algorithm
SAA	Simulated annealing algorithm
DT	Decision tree
RF	Random forest
ALO	Antlion optimizer
RS	Random search
SFLA	Shuffled frog leaping algorithm
SRMD	Specific rock mass drillability

1 Introduction

The ability to predict drilling machines' performance and efficiency is critical in mining operations. Mining operations' planning, development, and economics are influenced by the penetration rate. In addition, the penetration rate of rotary drilling plays a crucial role in preventing mine drilling costs [1]. The penetration rate describes how several parameters affect the rate of drilling in mining and construction. In terms of mining operations, the selection of mining machinery and equipment, and the product's final price, rough estimations of the penetration rate are potentially risky [2]. Furthermore, drilling rate equations can estimate total drilling costs. Selecting the type of machine can be accomplished using these equations. Numerous factors, such as rock-mass and material properties and the specifications of the drilling rig, influence the rotary drilling's penetration rate [3]. Despite drilling equipment parameters, geological conditions and rock characteristics are not normally controllable by humans [1, 4].

Various researchers have developed different methodologies to predict rotary drilling's penetration rates in recent years. Based on the literature study, these methods are divided into experimental and empirical models, statistical models, and machine learning models. Akün and Karpuz [5] established an empirical penetration rate model that was based on rock quality designation (RQD), pressure loss, discontinuity frequency, specific energy, and depth of cut. Their results showed that the drilling-specific energy is the most significant parameter in predicting the drilling rate. Krúpa et al. [6] conducted an experimental study that evaluated the relationship between achievable penetration depth and wear in rotary drilling. They created a mathematical model based on thrust force and drill length, omitting other parameters, such as torque and vibrations. Kumar et al. [7] evaluated rotary drilling's penetration rate through coarseness index mapping and vibration, observing that PR increased with mean particle size (d) but decreased with vibration for varying pulldown force and torque at different rotational speeds. Adoko et al. [8] developed an empirical model for estimating drilling rates in hard rock mining that showed a strong correlation between actual and estimated drill rates.

As the next category of the available techniques, some other researchers used statistical approaches for predicting rotary drilling's penetration rate. Kahraman [9] created an NLMR model to forecast the penetration rate of rotary and percussive drilling systems. The study found that uniaxial compressive strength is the primary rock property that affects rotary drilling. Hoseinie et al. [10] established a rating classification for drillability (DRI) prediction by using six rock mass properties, including Mohs hardness, grain size, UCS, joint filling, joint spacing, and joint dipping. Yarali and Kahraman [11] proposed new relationships to predict DRI by utilizing the brittleness values of 32 different rocks. Cheniany et al. [12] developed linear and nonlinear multiple regression techniques to estimate the SRMD index for specific rock masses. Moein et al. [13] measured the DRI values of carbonate rock in the laboratory and found good relationships for predicting DRI using the alteration index and specific energy. Yenice [14] established a mathematical regression model that considered rock strength properties to predict DRI, concluding that there is a strong relationship between PR and rock strength.

The last category of available techniques that are flexible and applicable is referred to as machine learning (ML). Recently, these techniques have been applied to civil and mining engineering [15–27]. Kahraman [28] demonstrated that ANNs could determine diamond drilling penetration rates much more accurately than regression models. Darbor et al. [29] evaluated the penetration rate using MLP-ANN and nonlinear multiple regression (NLMR). MLP-ANN is more accurate at predicting performance than NLMR, according to the study. Fattahi and Bazdar [30] used five hybrid ANN models to estimate DRI. Algorithms used include the simulated annealing algorithm (SAA), backpropagation (BP), firefly algorithm (FA), invasive weed optimization (IWO), and shuffled frog leaping algorithm (SFLA). According to

Table 1 An overview of models for predicting penetration rates in rotary drilling

Research	Input	ML Method	R^2
Bhatnagar and Khandelwal [33]	Thrust, RPM, FM, Comp St	ANN	0.63
Ataei et al. [2]	WOB, RPM, D, P,RD_i	ANN	0.97
Khandelwal and Armaghani [34]	UCS, BTS	GA-ANN	0.94
Kahraman [28]	σ_c, σ_t, T, R, RA	ANFIS	0.71
Fattahi and Bazdar [30]	UCS, BTS, SSH, Is(50) $\rightarrow$, Is(50)$\downarrow$	ANN-SAA	0.97
		ANN-SFLA	0.92
		ANN-IWO	0.90
		ANN-FA	0.89
		ANN-BP	0.75
Darbor et al. [29]	RQD, REC, W, UCS, BTS, B1, B2, B3, B4, ANI, RN	MLP-ANN	0.86
		NLMR	0.58
Kamran [31]	UCS, BTS, $S20$, Sj	DT	0.96
		AdaBoost	0.97
		RF	0.98
Sakız et al. [32]	UCS, BTS, CAI	FIS	0.95
Lawal et al. [3]	ρ, μ, Is(50)	ALO-ANN	0.90

** Symbols are explained in the Abbreviation section

the results, the ANN-SSA model performed better than other models ($R^2 = 0.97$). Shad et al. [1] investigate the penetration rate by the chemical component of intact rock, in addition to rock mass properties and machine specifications. The iron oxide percentage has been used as a new parameter for the penetration rate of rotary drills. Kamran [31] developed a probabilistic approach using AdaBoost, random forest (RF), and decision tree (DT) for predicting the drilling rate index (DRI). The results and Monte Carlo simulations show that this approach is more reliable in predicting the probability distribution of DRI. Sakız et al. [32] estimated the drilling rate index by focusing on the abrasive properties and rock strength. For this purpose, they used the fuzzy inference system (FIS) method as a model for precise prediction. The results showed that using the proposed method, DRI prediction is very efficient and accurate.

Table 1 summarizes some of the ML models that have been proposed by different researchers to predict drilling machine penetration rates. As shown, a relatively high level of accuracy can be obtained using ML techniques. ML methods can handle highly nonlinear relationships between predictor and response variables, which is often the case in drilling operations. Machine learning models are able to learn from large and complex datasets and generate highly accurate predictions, whereas statistical and empirical methods rely on assumptions such as the linearity and normality of data, which may not always hold in the context of drilling. Additionally, machine learning models can easily incorporate multiple sources of data, including real-time sensor data, to enhance their prediction accuracy, which is difficult to achieve using statistical or empirical methods.

Accurately predicting penetration rates in rotary drilling is of utmost importance in geomechanics, as it heavily relies on various rock-mass and material properties. In this article, we propose a groundbreaking hybrid XGBoost methodology to significantly enhance the accuracy of penetration rate predictions during drilling operations. This methodology finds wide-ranging applications in vital industries, such as oil and gas exploration, construction, and civil engineering. To achieve more precise predictions, a comprehensive understanding of rock-mass properties is essential. Notably, the mechanical behavior of crystalline rocks, as influenced by brittleness, has been studied extensively [35]. By optimizing drilling strategies and mitigating rock failures, this knowledge is instrumental in improving drilling performance. Furthermore, the significance of comprehending rock-mass heterogeneity has been highlighted through investigations into radionuclide transport in multi-scale fractured rocks [36]. Such understanding is crucial for achieving accurate predictions and ensuring efficient drilling operations.

In the context of solute transport monitoring in sedimentary media, integrated experimental design frameworks have been proposed [37]. Additionally, data-worth analysis using stochastic deep learning frameworks have been explored to identify subsurface structures [38], providing valuable insights into optimizing drilling practices. Ground-penetrating radar (GPR) has emerged as a valuable tool for subsurface exploration, with denoising methods being employed to enhance GPR data quality [39]. Moreover, deep learning models have been developed for pipeline recognition using GPR B-scans [40]. Radar technology's potential for remote sensing applications, such as discriminating between dry and water ices on Mars, has also been demonstrated [41]. The complexity of solute transport in naturally fractured media has spurred research on upscaling dispersivity for conservative transport analysis [42],with significant implications for groundwater management and contaminant assessments.

In engineering, various structures have been subject to advanced analyses and modeling. Shield tunnel linings, for instance, have been studied using finite element modeling [43], and seismic fragility analyses have considered soil property variability [44]. Ocean engineering studies have focused on the failure analysis of reinforced thermoplastic pipes [45] and the dynamic response of riserless rotating drill strings [46]. Moreover, quantitative determination of high-order crack fabric in rock planes contributes to rock mass stability assessment [47], while research on the effects

of carbonate minerals and exogenous acids on carbon flux addresses global and planetary change [48].

As far as the authors know, no study has applied hybrid XGBoost to predict penetration rate in rotary drilling. Hence, this study aims to fill this gap by proposing a novel approach to optimize XGBoost using various search algorithms, including random search, grid search, Harris Hawk optimization (HHO), and dragonfly algorithm (DA). The study was conducted using data collected from a copper mine in Iran, where the predictive models were developed by considering various rock properties. The authors then compared the developed models with the traditional XGB model to evaluate their effectiveness in predicting variations in the penetration rate. The proposed methodology and the role of HHO and DA in enhancing the predictive accuracy of XGB contribute to the field of mining engineering by introducing a practical approach to predict the penetration rate in rotary drilling.

2 Methodology

2.1 Extreme Gradient Boosting (XGBoost)

The XGBoost method is founded on gradient-boosting trees, which can be very useful for gradient enhancement [49]. A regression and classification problem can be very effectively solved using XGBoost based on the concept of regression and classification trees [21, 22]. Also, XGBoost combines the novel algorithm with the GBDT method to represent a soft computing library.

Two parts explain the XGBoost's objective function: first, the deviation from the model, and then the regular phrase to prevent overfitting. The data set represented by $D = \{(x_i, y_i)\}$ contains m features and n samples. Predictive variables are additive models consisting of k basic models. Equations (1) and (2) represent the results of the sample prediction.

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i),\ f_k \in \varphi, \tag{1}$$

$$\varphi = \{f(x) = w_s(x)\}\left(s : R^m \to T,\ w_s \in R^T\right) \tag{2}$$

In these Equations, $\hat{y}_i$ represents the prediction label, and one of the samples symbolizes by x_i. Also, φ represents the set of regression tree which is a tree structure parameter of s, and $f_k(x_i)$ represents the predicted score for that sample. In addition, $f(x)$ denotes the leave's value, and the number of leaves represents by w.

XGBoost's objective function contains both the complexity of the model and the traditional loss function. In this way, the algorithm can be evaluated in terms of its operational efficiency. A traditional loss function is represented by the first term in Eq. (3), while the complexity of the model is represented by the second term.

$$\text{Obj} = \sum_{i=1}^{m} l\left(y_i,\ \hat{y}_i^{(t-1)} + f_i(x_i)\right) + \Omega(f_k), \tag{3}$$

$$\Omega(f_k) = \gamma T + 1/2\lambda w^2 \tag{4}$$

In Eqs. (3) and (4), use m to specify how much data is imported into the *kth* tree and *i* to specify the number of samples in the dataset. The complexity of a tree can be adjusted using γ and λ. The final learning weight can be smoothed by adding regularization terms, and overfitting can be avoided [21, 22, 50, 51].

2.2 Hyperparameter Tuning

Based on existing data, tuning can be used to learn an algorithm that finds the optimal hyperparameters. A hyperparameter can determine an algorithm's optimum performance in supervised learning [52–55]. This research used three tuning methods to find the optimal hyperparameter: grid search, random search, and metaheuristic algorithms.

XGBoost is a machine-learning algorithm with great potential and many hyperparameters. The following are the hyperparameters that adjusted in this study:

learning_rate: This hyperparameter sets the step size for each boosting iteration. Smaller values may improve performance, but may also make training longer.

max_depth: This hyperparameter controls the maximum depth of the decision tree, which affects the model's complexity. Higher values may cause overfitting, while lower values may cause underfitting.

n_estimators: This hyperparameter determines the number of boosting iterations or trees to build.

2.2.1 Grid Search

The grid search is an exhaustive search using a set of subsets, with hyperparameters determined by a lower, an upper, and a number of steps [56]. The grid method creates a grid to find all possible outcomes. The most suitable grid will be chosen among all other grids, and all steps will be carried out systematically [57]. Using the grid search method, data can be processed with high accuracy [58]. In grid search, the following steps are taken:

1. The parameter values are all initialized
2. Combining all parameter values in a loop
3. Based on training data, XGBoost is used to conduct training
4. Analyzing test data with the resultant classifications

5. Analyzing the classification results to determine the most effective combination of parameter values

2.2.2 Random Search

Random search tries a number of predetermined combinations, evaluates hyperparameters, and selects the most promising ones [59]. Large volumes of data can be processed efficiently by random search [60]. The following are the steps involved in a random search:

1. Setting up the number of iterations for a parameter combination
2. The parameter values are all initialized
3. Combining parameter values randomly based on the iteration count
4. Based on training data, XGBoost is used to conduct training
5. Analyzing test data with the resultant classifications
6. Analyzing the classification results to determine the most effective combination of parameter values

In summary, random search is a method of selecting hyperparameters randomly from a given range. It is relatively simple to implement and can be used to quickly explore the parameter space. The main advantage of random search is its ability to find good solutions faster than grid search, as it does not require an exhaustive search of the parameter space. However, random search may not be able to find the best solution as it does not guarantee that all possible combinations are explored. Grid search is a method of systematically searching through a given range of hyperparameters in order to find the optimal combination. This method is more computationally expensive than random search, but it guarantees that all possible combinations are explored and thus can provide better results. The main disadvantage of grid search is its time-consuming nature, as it requires an exhaustive exploration of the parameter space [61].

2.2.3 Metaheuristic Algorithms

As an important branch of machine learning, the extreme gradient boosting model has been widely used in many areas, such as blasting, mining, energy, and geotechnics. By searching for the optimal combination of prediction models, prior literature shows that the use of metaheuristic optimization techniques can significantly improve prediction performance [62]. Hence, the Harris Hawks optimization (HHO) and dragonfly algorithm (DA) were tested in this paper.

Harris Hawks Optimization (HHO) The HHO algorithm's behavior cooperation in hunting has been used to illustrate a variety of issues requiring optimal solutions [63, 64]. According to Heidari et al. [65], it is useful for solving optimization problems in many different scientific and engineering fields. HHO is divided into three phases, as shown in Fig. 1: exploration, exploitation, and transition. Hawks search for and locate a prey animal, and then determine its position during the first phase, X_{rabit}. The hawks, using an iterative process, express their position relative to the prey in which they assign a random relevance to the prey, X_{rand}:

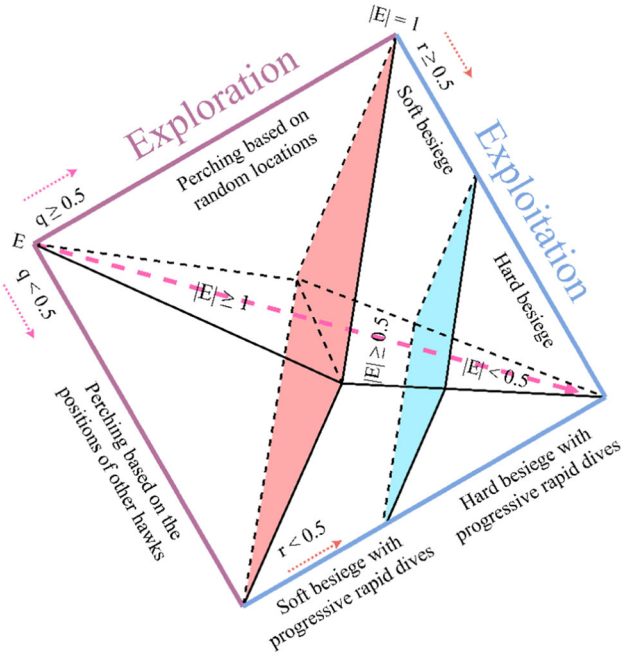


Fig. 1 An explanation of the HHO algorithm with three stages

$$X(\text{iter}+1) = \begin{cases} X_{\text{rand}}(\text{iter}) - r_1 X_{\text{rand}}(\text{iter}) - 2r_2 X_{\text{rand}}(\text{iter}), & q \geq 0.5 \\ X_{\text{rabit}}(iter) - X_m(\text{iter}) - r_3(\text{LB} + r_4(\text{UB} - \text{LB})), & q \geq 0.5 \end{cases} \tag{5}$$

Average position is determined by X_m, while r_i is determined by a random number i, which is a range of $(1 - q)$. A definition of m can be found in Eq. (6).

$$X_m(\text{iter}) = \frac{1}{N} \sum_{(i=1)}^{N} X_i(\text{iter}) \tag{6}$$

X_i indicates the location and N indicates the hawk's size. Hunting's escaping energy, E_{h}, comes from:

$$E_{\text{h}} = E_0 \left(1 - \frac{\text{iter}}{T}\right) \tag{7}$$

E_0 represents the initial energy and T represents the maximum number of repetitions. The following is noted that

$E_0 \in (-1, 1)$ and depending on the value of |E|, the exploration or exploitation phase will be initiated. The value of |E| indicates how the rabbit was captured during the exploitation phase. When $|E| \geq 0.5$, the catch is explained to be easy, but when $|E| < 0.5$, the catch is described as problematic [66, 67].

Dragonfly Algorithm (DA) Mirjalili [68] presents the DA as a novel method for optimizing systems. This approach is based on the swarm intelligence of dragonflies, which exhibits both static and dynamic behavior. Exploration and exploitation are the two main phases of DA. Dragonflies' dynamic or static searches for food or avoidance of enemies result in these two phases [69].

There are three distinct behaviors of swarms: alignment, cohesion, and separation [69]. Separation in this concept refers to avoiding a collision with an element in a swarm [Eq. (8)]. In Eq. (9), alignment is the speed at which factors adjust their positions to match their neighbors. In Eq. (10), cohesion is the inclination of elements toward the center. There are two different strategies in DA: approaching the food and avoiding enemies. Each swarm's main goal is survival, which is the reason for this addition. Thus, when all elements are moving toward food sources [Eq. (11)], they must stay away from the enemy [Eq. (12)].

$$S_i = -\sum_{j=1}^{N} X - X_i \tag{8}$$

$$A_i = \frac{\sum_{j=1}^{N} V_j}{N} \tag{9}$$

$$C_i = \frac{\sum_{j=1}^{N} X_j}{N} - X \tag{10}$$

$$F_i = X^+ - X \tag{11}$$

$$E_i = X - X \tag{12}$$

where X^+ and X^- represent food and enemies, respectively. V_j represents the speed of the jth neighbor element, and N denotes the number of neighbor elements. Also, X represents the momentary position of the element; X_j is the momentary position of the jth neighbor element [70, 71].

Finally, XGBoost predictions are optimized using DA and HHO algorithms in this study. XGBoost has been used to predict penetration rate in three forms; two usual forms of XGBoost (random search and grid search) and its combined form with the dragonfly algorithm and harris hawk optimization as XGB-DA and XGB-HHO, respectively. The schematic form of merging XGBoost with DA and HHO is shown in Fig. 2. Additionally, in the study approach, the prediction models were prepared using the XGBoost library, and the programming environment used for model development and evaluation was Python with the scikit-learn and XGBoost libraries.

2.3 Model Evaluation and Verification

Based on the correlation between predicted penetration rate and measured penetration rate value, several indicators are used in this research to assess the dependability of optimized models, including mean absolute error (MAE), root mean square error (RMSE), average absolute relative error (AARE), and coefficient of determination (R^2) [72–77].

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\mathrm{PR}}_i - \mathrm{PR}_i\right)^2} \tag{13}$$

$$\mathrm{AARE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\widehat{\mathrm{PR}}_i - \mathrm{PR}_i}{\mathrm{PR}_i}\right| \times 100 \tag{14}$$

$$R^2 = 1 - \frac{\sum_i \left(\mathrm{PR}_i - \widehat{\mathrm{PR}}_i\right)^2}{\sum_i \left(\mathrm{PR}_i - \overline{\mathrm{PR}}_i\right)^2} \tag{15}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{\mathrm{PR}}_i - \mathrm{PR}_i\right| \tag{16}$$

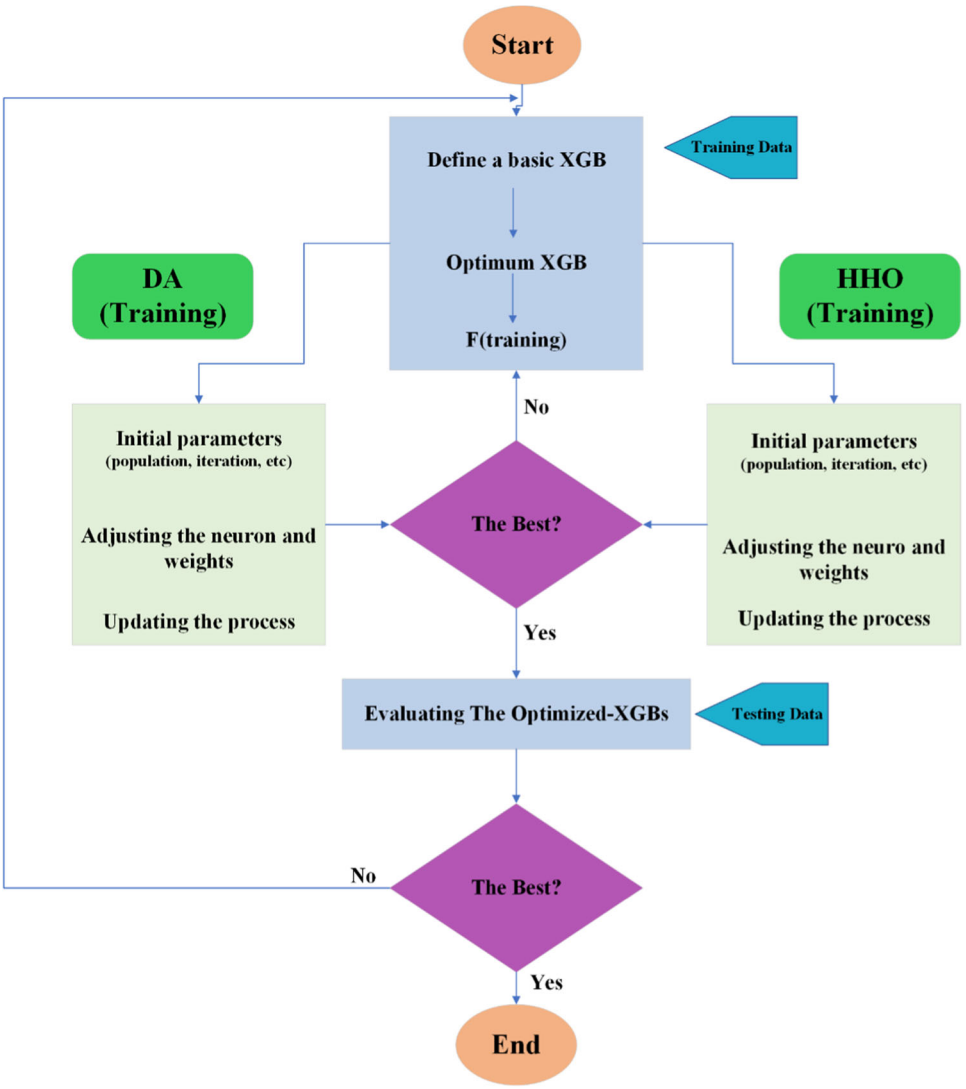
where $\widehat{\mathrm{PR}}_i$, PR_i, and $\overline{\mathrm{PR}}_i$ represents the predicted, measured, and average of the measured penetration rate values, respectively. In addition, n denotes the number of samples in the training and testing dataset.

3 Field Study and Data collection

3.1 Field Study

The copper mine Sarcheshmeh is located 50 and 160 km northwest of the Rafsanjan and Kerman cities, respectively. In terms of open pit mines in Iran, it is the largest (Fig. 3). A variety of rock types are found in the Sarcheshmeh porphyry deposits due to their complicated geology. Due to the high production rates and project planning requirements of large surface mining operations, large drilling rigs, such as rotary drills with triphonic bits, are commonly used. In recent years, rotary drills have become increasingly popular among mining contractors due to their ability to drill large diameters and deep blast holes. The disadvantages of utilizing this equipment are expensive maintenance and problematic transportation in areas with harsh topography. A list of all the rotary drill rig configurations is presented in Table 2.

Fig. 2 Copper mine Sarcheshmeh's location



3.2 Data Collection

Knowing the penetration rate of a project is crucial for determining drilling costs. The performance of drilling is affected by a variety of factors, including rock properties and the drilling equipment. Changing drilling equipment factors is possible, but changing rock parameters is not [4].

It is most helpful to provide a dataset with a wide geographical distribution when developing XGB-based methods to predict penetration rates in this paper. The first step in providing a dataset should also be to select the most appropriate input parameters. Therefore, a brief review of previous investigations is needed to determine the factors most influential on rotary drill performance. According to Bhatnagar and Khandelwal [33], thrust, RPM, flushing media, and compressive strength are the most effective rotary drill parameters. Numerous researchers believe that UCS and BTS play an important role in rotary drill performance [29, 30–32, 34, 78–80]. The joint specification, including joint spacing, joint aperture, joint fillings, and joint direction has a significant impact on the performance of rotary drills, according to Saeidi et al. [81] and Hoseinie et al. [10]. Lawal et al. [3] noted that density and porosity influence rotary drilling's penetration rate. Also, bit diameter has been used as an input parameter in the prediction of the penetration rate of rotary drilling models [2].

Observations and laboratory tests were conducted at the Sarcheshmeh copper mine to develop a database for hybrid intelligent techniques. A database consisting of 116 metamorphic, sedimentary, and igneous rock samples was examined m (the small version of database is available in Table 8). After discussing the matter and conducting field observations and laboratory tests, five parameters were chosen as inputs for predicting the PR of rotary drilling. These parameters include rock material properties such as uniaxial compressive strength (UCS) and tensile strength (TS), as

Fig. 3 The flowchart of the proposed process of optimized XGBs

Table 2 Various configurations of rotary drills were used in this study

Model	Trust (MPa)	Rotation press (MPa)	Bit diameter (m)	Air-line pressure (MPa)	Rotary speed (RPM)
DMH-IR-XL	0–21	0–34	0.25	0–2.75	0–200
Bucyrus 45-R-135490	0–17.23	0–31	0.23	0–2.41	0–150
DMH-Ingersoll-Rand	0–24.13	0–35	0.25	0–2.76	0–200

Table 3 Output and inputs data with symbols, details, and statistical descriptions

Type of data	Variables	Symbol	Unit	Minimum	Maximum	Mean	Std. deviation
Inputs	Bit diameter	D	inch	9	9.8	9.27	0.2
	Uniaxial compressive strength	UCS	MPa	50	226	157.93	4.26
	Tensile strength	TS	MPa	6	28	19.81	0.54
	Joint spacing	JS	cm	5	115	41.77	2.63
	Joint direction	JD	deg	10	90	36.03	1.49
Output	Penetration rate	PR	cm/min	8	100	40.36	2.28

well as rock mass properties like joint direction (JD) and joint spacing (JS), along with bit diameter (D). Although some studies use other parameters like RPM and WOB, they were not included in this study to keep the predictive models simple. Jahed Armaghani et al. [82] and Armaghani et al. [83] have suggested that models with fewer input parameters are better since they are less complex. Additionally, equipment parameters are typically selected based on the rock properties being drilled, so including them in the prediction model would be redundant and could make it unnecessarily complicated. Details and descriptions of the datasets are shown in Table 3. The data have been randomly divided into two parts: training data (92 cases) were allocated for model development, while test data (24 cases) were allocated to assess model reliability.

The Pearson correlation coefficients were computed as shown in Fig. 4 to determine the most suitable features of the predictive models. In terms of Pearson correlation coefficients, bit diameter (D) and joint direction (JD) with the strongest correlations ($r = -0.64$) and ($r = 0.27$), respectively, have the strongest linear relationship with penetration rate. However, after performing a sensitivity analysis, it is

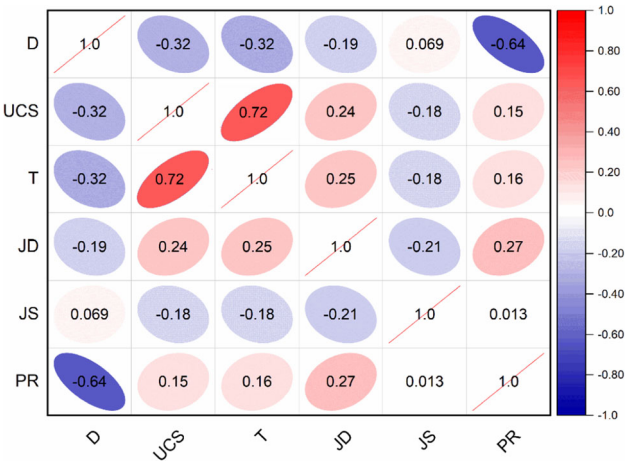


Fig. 4 An overview of the correlation matrix for all data samples (inputs and output)

Table 4 Model's optimal parameters

Model	Tuning method	Optimal parameters
RS–XGB	Random search	learning_rate = 0.15
		maximum_depth = 4
		n_estimator = 150
GS–XGB	Grid search	learning_rate = 0.09
		maximum_depth = 4
		n_estimator = 150
DA–XGB	Dragonfly algorithm	learning_rate = 0.15
		maximum_depth = 4
		n_estimator = 160
HHO–XGB	Harris hawk optimization	learning_rate = 0.109
		maximum_depth = 4
		n_estimator = 162

possible to identify the most influential parameters in the simulation of the penetration rate.

4 Results and Discussion

4.1 Comparison Analysis of Optimized Models

In order to estimate drilling penetration rates, a database must be prepared. All models were trained on 80% of the data randomly selected from the database and tested on 20% of the data. Several performance indicators (RMSE, AARE, R^2, and MAE) in Eqs. (13)–(16) were applied to evaluate the model's accuracy. All prediction models were trained and tested using the same set of data. The optimized XGB models was carried out based on Sect. 2 and the method shown in Fig. 2. Firstly, the relevant parameters of XGB models were initialized. After that, each optimization algorithm's appropriate parameters were determined. Table 4 includes the optimal parameters obtained from the optimization process.

On the training set in this paper, varieties of XGB-tuned models were trained, and their prediction performances varied. Figure 5 illustrates the correlation between the actual and predicted values of the training data set. These optimized models exhibit relatively good training effects, with the training data points distributed close to the perfect fit line. In terms of RMSE, AARE, R^2, and MAE, the HHO-XGB optimized model has a slight advantage, with values of 0.1384, 0.1384, 99, and 3.9882, respectively.

In this paper, four optimized XGB techniques are presented that can achieve high training effects, with R^2 values generally above 0.96. Following training, the models are evaluated and verified on the testing data set. Figure 6 illustrates how the test data set is essentially distributed close to the perfectly fitted line when examining the relationship and error between measured and predicted values.

These four optimized models have been compared and analyzed based on their predicted performance, as shown in Table 5 and Figs. 7, 8, 9 and 10. According to Table 5, the performance ranking system and index results of four models (DA–XGB, HHO–XGB, RS–XGB, and GS–XGB) predict rotary drilling penetration rates. The stacked graph in Fig. 7 presents the overall rankings more intuitively. Comparing all optimized models, the results of the comprehensive analysis indicate that the HHO-XGB optimized model is the most precise predictive model.

Using the testing dataset, we compare the predicted penetration rate accuracy of the selected models, as shown in Fig. 8. The XGBoost technique-based metaheuristic algorithm gives the most accurate and consistent results in penetration rate prediction, as shown in Fig. 8.

This subsection illustrates how optimal predictive models perform with the Taylor diagram. This mathematical diagram is used to illustrate which of the developed models is the most realistic [84]. RMSE, standard deviation, and Pearson correlation are used to assess the degree of agreement between modeled and observed behavior [85–87]. Figure 9 depicts the Taylor diagram of this study's models generated for testing and training datasets. According to the results, the HHO–XGB hybrid model is more accurate in predicting penetration rate than other predictive models.

The distribution of predicted values is one way to evaluate predictive models. Box plots in Fig. 10 show the distribution functions for predicted and measured penetration rates. Due to its similar probability distribution to observational results, the HHO-XGB approach performed better than other models.

Another way to evaluate ML approaches is by calculating the cumulative frequency of absolute relative errors (ARE, %). Based on this Fig. 11, more than 90% of predicted

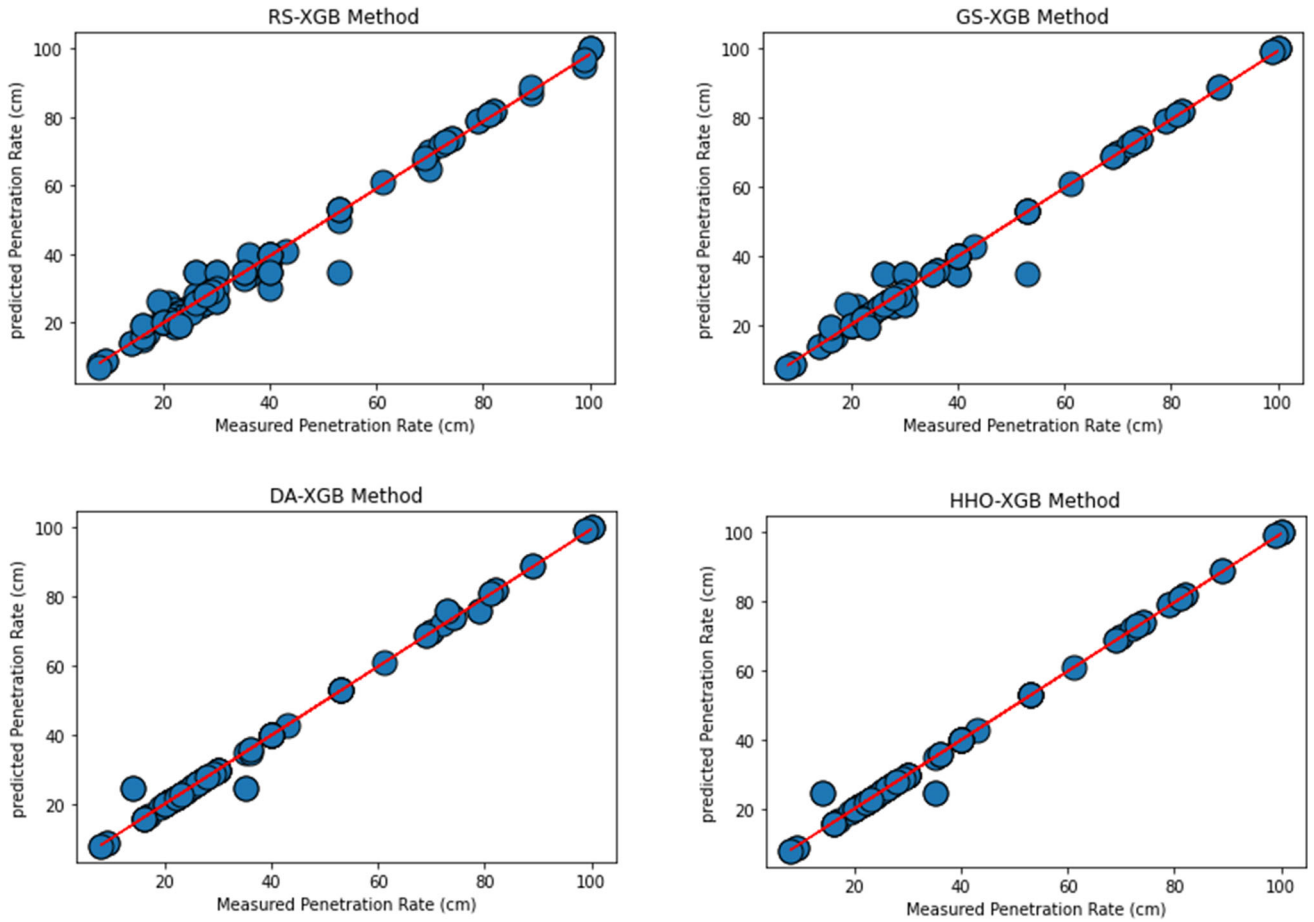


Fig. 5 Analyses the correlation between measured and predictive values from the training dataset

penetration rates for HHO–XGB and DA–XGB models are below 5%, as opposed to less than 85%, 60%, and 30% for GS–XGB, RS–XGB, and XGB models. The HHO–XGB and DA–XGB correlations with actual penetration rates are excellent, as demonstrated by the presented figures.

The significance of R^2 can be assessed through a 't' test, under the assumption that both variables follow a normal distribution and the observations are randomly selected. The 't' tests involve the comparison of the calculated 't' value and the tabulated 't' value under the null hypothesis [88, 89]. A 95% confidence level is employed for this test. If the computed value exceeds the tabulated value, the null hypothesis is rejected, indicating the significance of '*r*'. By applying a 95% confidence level, we obtain the corresponding critical value of 1.98. The data presented in Table 6 demonstrates that all the computed 't' values surpass the tabulated 't' values, signifying a statistically significant correlation between D, T, JD, JS, UCS, and the penetration rate.

To enhance prediction confidence, the *F*-test was conducted, as illustrated in Table 7. The *F*-test is employed to compare standard deviations [88]. Employing a 95% confidence level yielded a corresponding critical value of 3.92. The information presented in Table 7 reveals that all the calculated F values exceed the tabulated F values, providing additional evidence of a meaningful connection between *D*, *T*, JD, JS, UCS, and the penetration rate. As the computed *F* value surpasses the tabulated *F* value in each case, the null hypothesis is rejected. This supports the assertion that the datasets originate from distinct populations of measurements.

Statistical significance, in essence, determines whether a given independent variable has an impact on the model. When the 'signif. value' of an independent variable exceeds (α), it does not contribute significantly to the model. However, when its 'signif. value' falls below (α), it assumes a meaningful role in the prediction process.

4.2 Sensitivity Analysis

Performing sensitivity analysis is a beneficial tool for evaluating the influence of variables on penetration rate predictions. In this study, the importance of variables was determined from the HHO–XGB model by comparing the results of different XGB models. The cosine amplitude method is

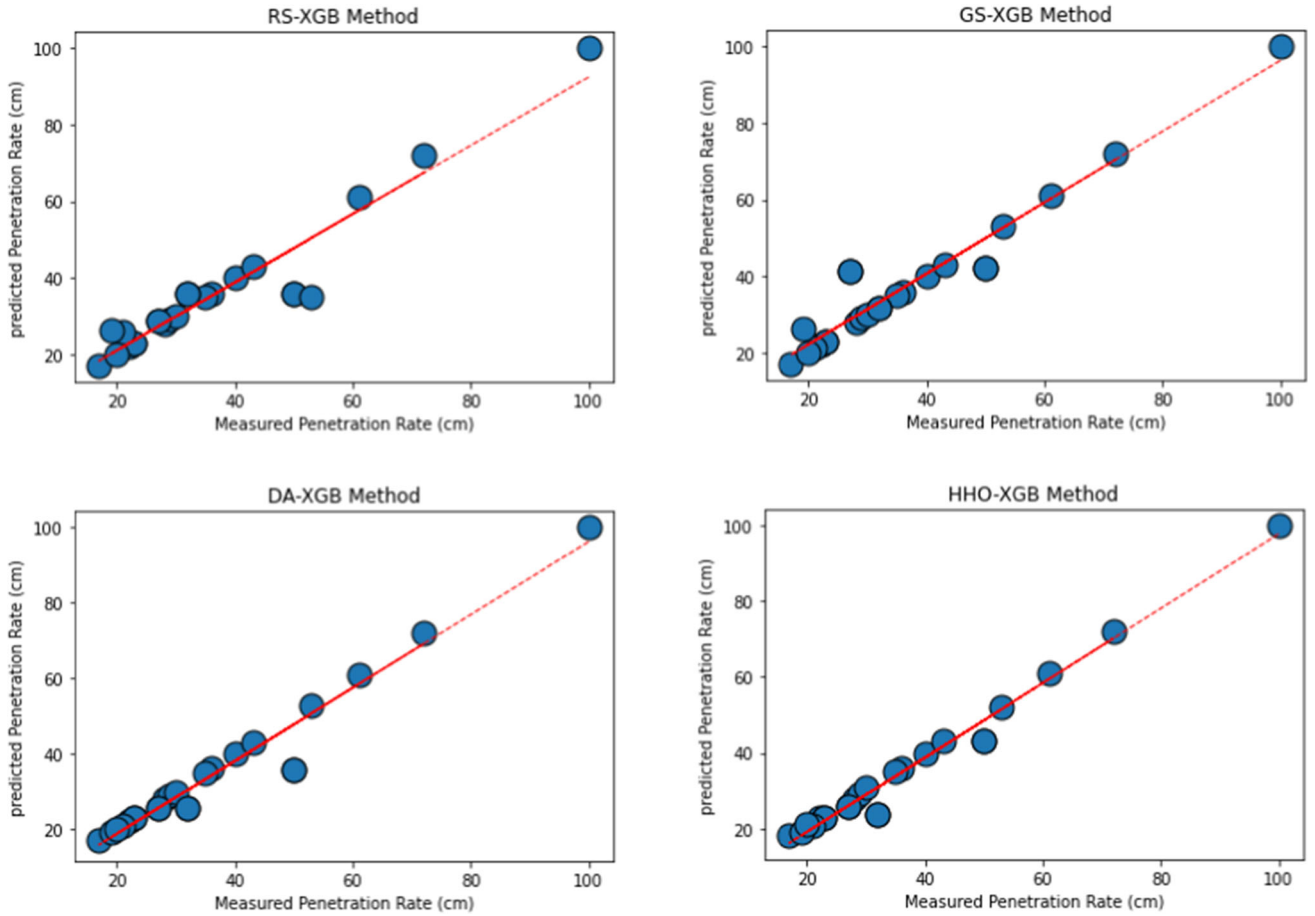


Fig. 6 Analyses the correlation between measured and predictive values from the testing dataset

Table 5 Performance comparison of the optimized XGB models

Data set	Model	Model performance evaluation criteria								Rank
		RMSE	Score	AARE	Score	R^2	Score	MAE	Score	
Training	XGBoost	5.64	1	12.79	1	0.9652	1	4.004	1	4
	RS–XGB	3.32	2	5.85	2	0.984	2	1.76	2	8
	GS–XGB	2.85	3	3.67	3	0.988	3	1.02	3	12
	HHO–XGB	2.19	5	2.29	5	0.993	5	0.457	5	20
	DA–XGB	2.28	4	2.49	4	0.992	4	0.599	4	16
Testing	XGBoost	5.12	2	10	1	0.928	2	2.36	2	7
	RS–XGB	5.8	1	7.9	2	0.908	1	2.94	1	5
	GS–XGB	5	3	7.48	3	0.933	3	2.2	3	12
	HHO–XGB	4.53	4	4.45	5	0.953	4	1.83	4	17
	DA–XGB	3.12	5	4.52	4	0.978	5	1.55	5	19

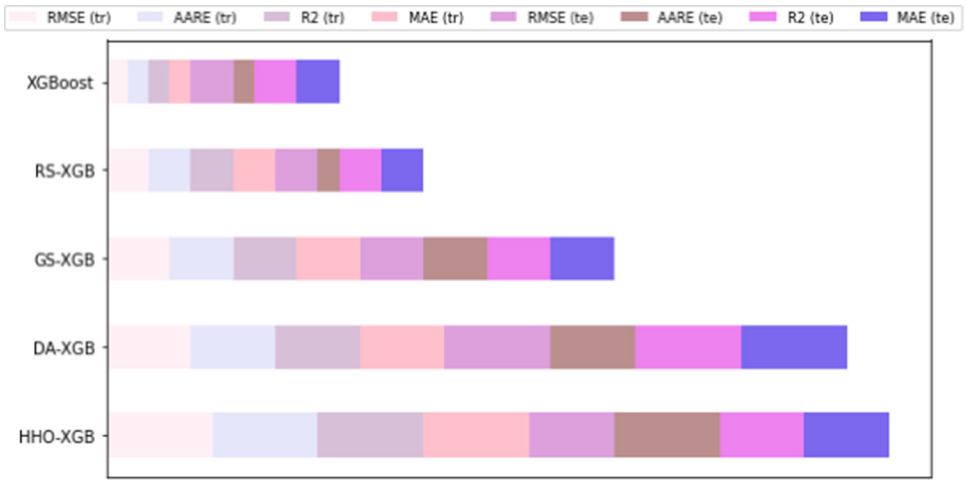


Fig. 7 Display comprehensive rankings of optimized models in an intuitive manner

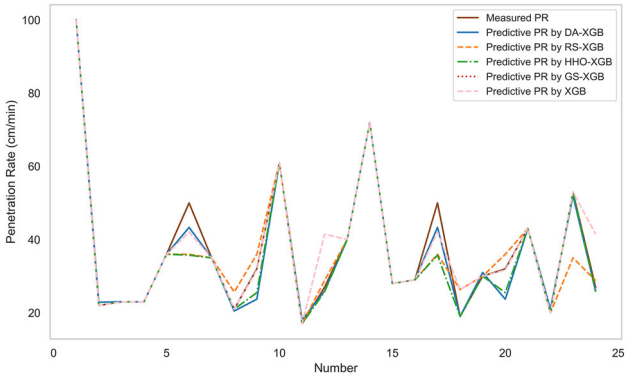


Fig. 8 A comparison of the prediction values of predictive models with testing dataset

employed for this purpose Shirani Faradonbeh et al. [90], considering the data set of the Copper Mine and the mathematical Eq. (17).

$$R_{ij} = \frac{\sum_{k=1}^{n}\left(x_{ik} \times x_{jk}\right)}{\sqrt{\sum_{k=1}^{n} x_{ik}^{2} \times \sum_{k=1}^{n} x_{jk}^{2}}} \tag{17}$$

The parameters x_i represent the inputs, x_j represents the outputs, and n represents the number of data sets. R_{ij} represents the strength of the relationship between the HHO–XGB and the independent variables. Figure 12 shows the strength of the relationship between penetration rate values and input data. Among the parameters affecting penetration rate, uniaxial compressive strength and tensile strength are the most influential.

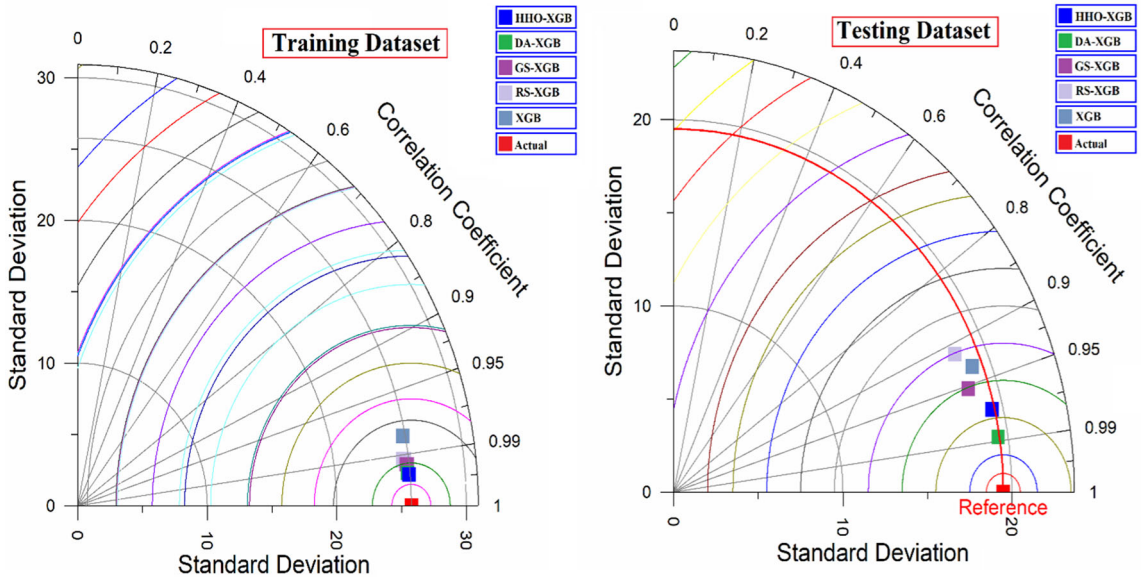
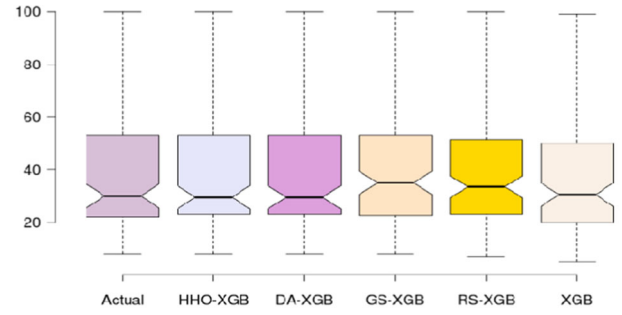


Fig. 9 Taylor diagram of predictive models for training and testing datasets



	Actual	HHO-XGB	DA-XGB	GS-XGB	RS-XGB	XGB
Upper whisker	100.00	100.00	100.00	100.00	100.00	99.10
3rd quartile	53.00	53.00	53.00	53.00	51.50	50.05
Median	30.00	29.50	29.50	35.00	33.50	30.50
1st quartile	22.00	23.00	23.00	22.50	23.00	20.00
Lower whisker	8.00	8.00	8.00	8.00	7.10	5.00
Nr. of data points	116.00	116.00	116.00	116.00	116.00	116.00

Fig. 10 An overview of the distribution function for all datasets in the developed models

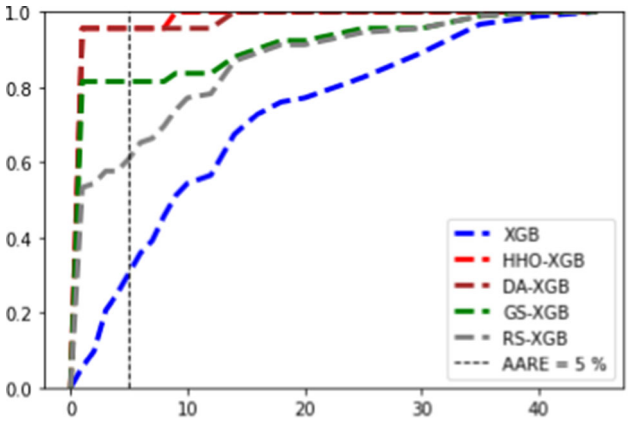


Fig. 11 Various models' cumulative absolute relative error

5 Limitations and Future Works

The current approach proposed in this study offers accurate predictions, but there are limitations that need to be addressed in the future. The proposed model is based on a database of 116 data samples from a copper mine in Iran, which limits its applicability to similar rock-mass and material properties. To develop a more generalized ML model, a comprehensive database with various types of parameters, including environmental conditions and different rock properties such as hardness, strength, and abrasiveness, can be collected. This would enable the use of a wider range of input parameters, making the ML model more reliable and flexible for researchers and designers. Moreover, future studies can explore the application of other ML methodologies or hybrid intelligence to compare their ability to predict the rotary drill penetration rate or other important properties. Such studies could improve the accuracy and robustness of the models developed for rotary drilling applications.

Potential biases may emerge due to site-specific conditions or sampling techniques, while uncertainties can stem from measurement errors, data interpolation, and temporal variations. These challenges collectively can compromise the reliability of predictive models, potentially resulting in inaccuracies during assessments and suboptimal decision-making. It's paramount to comprehend these limitations to effectively interpret results and employ robust modeling techniques that address biases and uncertainties.

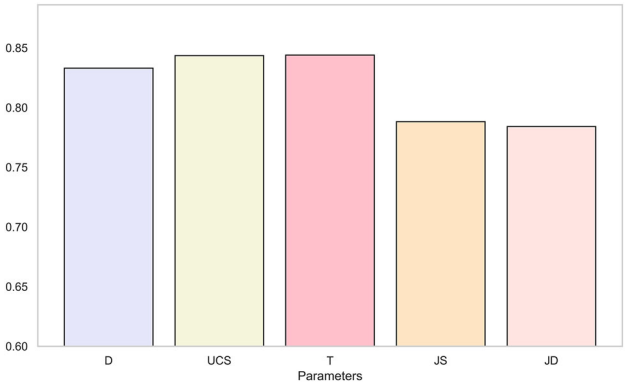


Fig. 12 Analyzing the sensitivity of the input variables and penetration rate

Table 6 Student's *t*-test

Parameter versus PR	*t* test	
	Calculated value	Tabulated value
D and PR	13.42	1.98
T and PR	26.03	1.98
JD and PR	9.09	1.98
JS and PR	28.74	1.98
UCS and PR	30.45	1.98

Table 7 Student's *F* test

Parameter versus PR	*F* test	
	Calculated value	Tabulated value
D and PR	185	3.92
T and PR	76.9	3.92
JD and PR	34.5	3.92
JS and PR	42.4	3.92
UCS and PR	591	3.92

6 Conclusions

This article carefully examines and compares different optimized XGB models to predict the penetration rate of rotary drilling. These models were created by combining XGB with four hyperparameter tuning methods including random search, grid search, and intelligent optimization algorithms like HHO and DA. Taking into account various factors that influence the penetration rate, a PR data set was used to train and test these XGB hybrid models. The models' performance was assessed using metrics, such as MAE, RMSE, AARE, and R^2. Lastly, the cosine amplitude method was employed to evaluate the significance of each input variable.

To sum up, the hybrid XGB models suggested in this study show promise in predicting rotary drilling penetration rates and can effectively aid XGB in adjusting hyperparameters. Based on the prediction result, HHO-XGB hybrid model demonstrates superior overall performance compared to the other three models. Additionally, an ordinary XGB model was developed to forecast the rotary drilling penetration rate for comparative purposes. The findings demonstrated that the XGB-based optimization methods had superior predictive accuracy compared to the ordinary model. The optimized HHO-XGB model was identified as the most effective model in predicting PR. To determine the input variable's importance, a sensitivity analysis technique called cosine amplitude was employed. Uniaxial compressive strength and tensile strength were found to be the most significant parameters affecting the penetration rate. Through the use of these developed models, the penetration rate of equivalent rocks can be accurately predicted.

It is important to highlight that the models developed during the PR prediction effort are specific to the current rock engineering challenge and cannot be easily applicable to other rock engineering issues. However, the offered created methods should be considered a foundation and should be re-evaluated, re-analyzed, and even re-addressed in order to take on the alternative rock designing and planning tasks. In predicting geological factors like PR, it's evident that incorporating additional rock parameters renders predictions more significant and reliable. In future research, it's advisable to enhance the quantity of rocks in the study to yield more credible prediction models. Additionally, evaluating rocks based on their source is recommended for robust evaluation within prediction models.

Acknowledgements The authors express their gratitude to the staff of Sarcheshmeh copper mine, with a special acknowledgment to Mr. Tavakoli, for their generous cooperation and assistance throughout the course of this research.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Appendix

See Table 8.

Table 8 The small version of dataset

No	Bit diameter (inch)	Uniaxial compressive strength (MPa)	Tensile strength (MPa)	Joint spacing (cm)	Joint direction (deg)	Penetration rate (cm/min)
1	9.8	149	19	25	65	24
2	9.8	122	15	25	20	26
3	9.8	213	27	30	50	27
4	9.8	181	23	50	30	21
5	9.8	56	7	7.5	40	17
6	9.8	147	18	5	40	30
7	9	142	18	25	15	32
8	9	137	17	35	50	27
9	9	204	26	25	25	25
10	9	151	19	50	50	35

Table 8 (continued)

No	Bit diameter (inch)	Uniaxial compressive strength (MPa)	Tensile strength (MPa)	Joint spacing (cm)	Joint direction (deg)	Penetration rate (cm/min)
11	9	124	16	40	45	40
12	9	161	20	5	45	30
13	9.8	50	6	25	45	23
14	9.8	83	10	25	10	16
15	9.8	166	21	40	45	22
16	9.8	63	8	5	35	9
17	9.8	168	21	15	35	8
18	9	159	20	35	50	19
19	9	215	27	115	20	14
20	9	144	18	55	30	50
21	9	99	12	40	50	40
22	9	144	18	5	60	36

References

1. Shad, H.I.A., et al.: Prediction of rotary drilling penetration rate in iron ore oxides using rock engineering system. Int. J. Min. Sci. Technol. **28**(3), 407–413 (2018)
2. Ataei, M., et al.: Drilling rate prediction of an open pit mine using the rock mass drillability index. Int. J. Rock Mech. Min. Sci. **73**, 130–138 (2015)
3. Lawal, A.I.; Kwon, S.; Onifade, M.: Prediction of rock penetration rate using a novel antlion optimized ANN and statistical modelling. J. Afr. Earth Sc. **182**, 104287 (2021)
4. Yarali, O.; Soyer, E.: Assessment of relationships between drilling rate index and mechanical properties of rocks. Tunn. Undergr. Space Technol. **33**, 46–53 (2013)
5. Akün, M.; Karpuz, C.: Drillability studies of surface-set diamond drilling in Zonguldak region sandstones from Turkey. Int. J. Rock Mech. Min. Sci. **42**(3), 473–479 (2005)
6. Krúpa, V., et al.: Measurement, modeling and prediction of penetration depth in rotary drilling of rocks. Measurement **117**, 165–175 (2018)
7. Kumar, R.; Murthy, V.; Kumaraswamidhas, L.: Performance analysis of rotary blast-hole drills through machine vibration and coarseness index mapping–a novel approach. Measurement **165**, 108148 (2020)
8. Adoko, A., Moesi, D., Sharipov, A.: Empirical relationship for drilling rate in hard rock underground mines. In: IOP Conference Series: Earth and Environmental Science. IOP Publishing (2021)
9. Kahraman, S.: Rotary and percussive drilling prediction using regression analysis. Int. J. Rock Mech. Min. Sci. **36**(7), 981–989 (1999)
10. Hoseinie, S.; Ataei, M.; Osanloo, M.: A new classification system for evaluating rock penetrability. Int. J. Rock Mech. Min. Sci. **46**(8), 1329–1340 (2009)
11. Yarali, O.; Kahraman, S.: The drillability assessment of rocks using the different brittleness values. Tunn. Undergr. Space Technol. **26**(2), 406–414 (2011)
12. Cheniany, A., et al.: An estimation of the penetration rate of rotary drills using the specific rock mass drillability index. Int. J. Min. Sci. Technol. **22**(2), 187–193 (2012)
13. Moein, M.J.A.; Shaabani, E.; Rezaeian, M.: Experimental evaluation of hardness models by drillability tests for carbonate rocks. J. Petrol. Sci. Eng. **113**, 104–108 (2014)
14. Yenice, H.: Determination of drilling rate index based on rock strength using regression analysis. An. Acad. Bras. Ciênc. (2019). https://doi.org/10.1590/0001-3765201920181095
15. Cavaleri, L., et al.: Convolution-based ensemble learning algorithms to estimate the bond strength of the corroded reinforced concrete. Constr. Build. Mater. **359**, 129504 (2022)
16. He, B.; Armaghani, D.J.; Lai, S.H.: Assessment of tunnel blasting-induced overbreak: a novel metaheuristic-based random forest approach. Tunn. Undergr. Space Technol. **133**, 104979 (2023)
17. Ghanizadeh, A.R., et al.: Developing predictive models of collapse settlement and coefficient of stress release of sandy-gravel soil via evolutionary polynomial regression. Appl. Sci. **12**(19), 9986 (2022)
18. Shan, F., et al.: Success and challenges in predicting TBM penetration rate using recurrent neural networks. Tunn. Undergr. Space Technol. **130**, 104728 (2022)
19. Skentou, A.D., et al.: Closed-form equation for estimating unconfined compressive strength of granite from three non-destructive tests using soft computing models. Rock Mech. Rock Eng. **56**(1), 487–514 (2023)
20. Barkhordari, M.; Armaghani, D.; Fakharian, P.: Ensemble machine learning models for prediction of flyrock due to quarry blasting. Int. J. Environ. Sci. Technol. **19**(9), 8661–8676 (2022)
21. Nabavi, Z., et al.: A hybrid model for Backbreak prediction using XGBoost machine learning and metaheuristic algorithms in Chadormalu iron mine. J. Min. Environ. (2023). https://doi.org/10.22044/jme.2023.12796.2323
22. Kazemi, M.M.K.; Nabavi, Z.; Khandelwal, M.: Prediction of blast-induced air overpressure using a hybrid machine learning model and gene expression programming (GEP): a case study from an iron ore mine. AIMS Geosci. **9**(2), 357–381 (2023)
23. Pradeep, T., et al.: Reliability and prediction of embedment depth of sheet pile walls using hybrid ANN with optimization techniques. Arab. J. Sci. Eng. **47**(10), 12853–12871 (2022)
24. Cakiroglu, C., et al.: Explainable ensemble learning data-driven modeling of mechanical properties of fiber-reinforced rubberized recycled aggregate concrete. J. Build. Eng. (2023). https://doi.org/10.1016/j.jobe.2023.107279

25. Cakiroglu, C., et al.: Data-driven ensemble learning approach for optimal design of cantilever soldier pile retaining walls. Structures (2023). https://doi.org/10.1016/j.istruc.2023.03.109
26. Cakiroglu, C., et al.: Explainable machine learning models for predicting the axial compression capacity of concrete filled steel tubular columns. Constr. Build. Mater. **356**, 129227 (2022)
27. Sari Ahmed, B., et al.: Best-fit models for predicting the geotechnical properties of FA–stabilised problematic soils used as materials for earth structures. Int. J. Pavement Eng. **21**(7), 939–953 (2020)
28. Kahraman, S.: Estimating the penetration rate in diamond drilling in laboratory works using the regression and artificial neural network analysis. Neural. Process. Lett. **43**, 523–535 (2016)
29. Darbor, M.; Faramarzi, L.; Sharifzadeh, M.: Performance assessment of rotary drilling using non-linear multiple regression analysis and multilayer perceptron neural network. Bull. Eng. Geol. Env. **78**, 1501–1513 (2019)
30. Fattahi, H.; Bazdar, H.: Applying improved artificial neural network models to evaluate drilling rate index. Tunn. Undergr. Space Technol. **70**, 114–124 (2017)
31. Kamran, M.: A probabilistic approach for prediction of drilling rate index using ensemble learning technique. J. Min. Environ. **12**(2), 327–337 (2021)
32. Sakız, U.; Kaya, G.U.; Yaralı, O.: Prediction of drilling rate index from rock strength and cerchar abrasivity index properties using fuzzy inference system. Arab. J. Geosci. **14**(5), 354 (2021)
33. Bhatnagar, A.; Khandelwal, M.: An intelligent approach to evaluate drilling performance. Neural Comput. Appl. **21**, 763–770 (2012)
34. Khandelwal, M.; Armaghani, D.J.: Prediction of drillability of rocks with strength properties using a hybrid GA-ANN technique. Geotech. Geol. Eng. **34**(2), 605–620 (2016)
35. Peng, J., et al.: Numerical investigation of brittleness effect on strength and microcracking behavior of crystalline rock. Int. J. Geomech. **22**(10), 04022178 (2022)
36. Zhang, X., et al.: Radionuclide transport in multi-scale fractured rocks: a review. J. Hazard. Mater. **424**, 127550 (2022)
37. Dai, Z., et al.: An integrated experimental design framework for optimizing solute transport monitoring locations in heterogeneous sedimentary media. J. Hydrol. **614**, 128541 (2022)
38. Zhan, C., et al.: Data-worth analysis for heterogeneous subsurface structure identification with a stochastic deep learning framework. Water Resour. Res. **58**(11), e2022WR033241 (2022)
39. Li, R., et al.: Denoising method of ground-penetrating radar signal based on independent component analysis with multifractal spectrum. Measurement **192**, 110886 (2022)
40. Liu, H., et al.: Automatic recognition and localization of underground pipelines in GPR B-scans using a deep learning model. Tunn. Undergr. Space Technol. **134**, 104861 (2023)
41. Liu, H., et al.: Discrimination between dry and water ices by full polarimetric radar: implications for China's first martian exploration. IEEE Trans. Geosci. Remote Sens. **61**, 1–11 (2022)
42. Jia, S., et al.: Upscaling dispersivity for conservative solute transport in naturally fractured media. Water Res. **235**, 119844 (2023)
43. Liu, C., et al.: Development of crack and damage in shield tunnel lining under seismic loading: refined 3D finite element modeling and analyses. Thin Walled Struct. **185**, 110647 (2023)
44. Wang, W., et al.: Seismic fragility and demand hazard analyses for earth slopes incorporating soil property variability. Soil Dyn. Earthq. Eng. **173**, 108088 (2023)
45. Wang, Y.-Y., et al.: Stochastic failure analysis of reinforced thermoplastic pipes under axial loading and internal pressure. China Ocean Eng. **36**(4), 614–628 (2022)
46. Wang, Y., et al.: Experimental investigation of the effect of rotation rate and current speed on the dynamic response of riserless rotating drill string. Ocean Eng. **280**, 114542 (2023)
47. Li, X., et al.: Quantitative determination of high-order crack fabric in rock plane. Rock Mech. Rock Eng. (2023). https://doi.org/10.1007/s00603-023-03319-x
48. Li, C., et al.: Effects of carbonate minerals and exogenous acids on carbon flux from the chemical weathering of granite and basalt. Global Planet. Change **221**, 104053 (2023)
49. Chen, T., et al.: Xgboost: extreme gradient boosting. R Package Version 0.4-2 **1**(4), 1–4 (2015)
50. Çakıroğlu, M.A., et al.: Experimental examination of the behavior of shotcrete-reinforced masonry walls and XgBoost neural network prediction model. Arab. J. Sci. Eng. **46**(11), 10613–10630 (2021)
51. Kavzoglu, T.; Teke, A.: Predictive Performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (XGBoost) and natural gradient boosting (NGBoost). Arab. J. Sci. Eng. **47**(6), 7367–7385 (2022)
52. Probst, P.; Wright, M.N.; Boulesteix, A.L.: Hyperparameters and tuning strategies for random forest. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. **9**(3), e1301 (2019)
53. Khari, M.; Jahed Armaghani, D.; Dehghanbanadaki, A.: Prediction of lateral deflection of small-scale piles using hybrid PSO–ANN model. Arab. J. Sci. Eng. **45**(5), 3499–3509 (2020)
54. Mahadeva, R., et al.: A Novel AGPSO3-based ANN prediction approach: application to the RO desalination plant. Arab. J. Sci. Eng. (2023). https://doi.org/10.1007/s13369-023-07631-0
55. Mohamad, E.T., et al.: A new hybrid method for predicting ripping production in different weathering zones through in situ tests. Measurement **147**, 106826 (2019)
56. Syarif, I.; Prugel-Bennett, A.; Wills, G.: SVM parameter optimization using grid search and genetic algorithm to improve classification performance. TELKOMNIKA Telecommun. Comput. Electron. Control **14**(4), 1502–1509 (2016)
57. Ataei, M.; Osanloo, M.: Using a combination of genetic algorithm and the grid search method to determine optimum cutoff grades of multiple metal deposits. Int. J. Surf. Min. Reclam. Environ. **18**(1), 60–78 (2004)
58. Xiao, T., et al.: Based on grid-search and PSO parameter optimization for support vector machine. In: Proceeding of the 11th World Congress on Intelligent Control and Automation. IEEE (2014)
59. Zhang, L., Zhan, C.: Machine learning in rock facies classification: An application of XGBoost. In: International Geophysical Conference, Qingdao, China, 17–20 April 2017. Society of Exploration Geophysicists and Chinese Petroleum Society (2017)
60. Putatunda, S., Rama, K.: A comparative analysis of hyperopt as against other approaches for hyper-parameter optimization of XGBoost. In: Proceedings of the 2018 International Conference on Signal Processing and Machine Learning (2018)
61. Anggoro, D.A.; Mukti, S.S.: Performance comparison of grid search and random search methods for hyperparameter tuning in extreme gradient boosting algorithm to predict chronic kidney failure. Int. J. Intell. Eng. Syst **14**(6), 198–207 (2021)
62. Qiu, Y., et al.: Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration. Eng. Comput. (2021). https://doi.org/10.1007/s00366-021-01393-9
63. Bui, D.T., et al.: A novel swarm intelligence—Harris hawks optimization for spatial assessment of landslide susceptibility. Sensors **19**(16), 3590 (2019)
64. Aleem, S.H.A., et al.: Harmonic overloading minimization of frequency-dependent components in harmonics polluted distribution systems using harris hawks optimization algorithm. IEEE Access **7**, 100824–100837 (2019)
65. Heidari, A.A., et al.: Harris hawks optimization: Algorithm and applications. Future Gener. Comput. Syst. **97**, 849–872 (2019)

66. Zhang, Y.; Zhou, X.; Shih, P.-C.: Modified Harris Hawks optimization algorithm for global optimization problems. Arab. J. Sci. Eng. **45**, 10949–10974 (2020)
67. Issa, M.: Enhanced arithmetic optimization algorithm for parameter estimation of PID controller. Arab. J. Sci. Eng. **48**(2), 2191–2205 (2023)
68. Mirjalili, S.: Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. Neural Comput. Appl. **27**, 1053–1073 (2016)
69. Acı, Ç.İ; Gülcan, H.: A modified dragonfly optimization algorithm for single-and multiobjective problems using Brownian motion. Comput. Intell. Neurosci. (2019). https://doi.org/10.1155/2019/6871298
70. Joshi, M., et al.: A conceptual comparison of dragonfly algorithm variants for CEC-2021 global optimization problems. Arab. J. Sci. Eng. **48**(2), 1563–1593 (2023)
71. Xu, J.; Yan, F.: Hybrid Nelder–Mead algorithm and dragonfly algorithm for function optimization and the training of a multilayer perceptron. Arab. J. Sci. Eng. **44**, 3473–3487 (2019)
72. Momeni, E., et al.: Gaussian process regression technique to estimate the pile bearing capacity. Arab. J. Sci. Eng. **45**, 8255–8267 (2020)
73. Fayyazi, A.; Doostmohammadi, R.: Investigation of the effective parameters of travertine stones healing using bio-grouting. J. Min. Sci. **58**(6), 1069–1083 (2022)
74. Adnan, R. M.; Mostafa, R. R.; Dai, H. L.; Heddam, S.; Kuriqi, A.; Kisi, O.: Pan evaporation estimation by relevance vector machine tuned with new metaheuristic algorithms using limited climatic data. Eng. Appl. Comput. Fluid Mech. **17**(1), 2192258 (2023)
75. Adnan, R. M.; Meshram, S. G.; Mostafa, R. R.; Islam, A. R. M. T.; Abba, S. I.; Andorful, F.; Chen, Z.: Application of advanced optimized soft computing models for atmospheric variable forecasting. Mathematics **11**(5), 1213 (2023)
76. Ikram, R. M. A.; Mostafa, R. R.; Chen, Z.; Islam, A. R. M. T.; Kisi, O.; Kuriqi, A.; Zounemat-Kermani, M.: Advanced hybrid metaheuristic machine learning models application for reference crop evapotranspiration prediction. Agronomy **13**(1), 98 (2022)
77. Anvari, K., et al.: Automatic detection of rock boundaries using a hybrid recurrence quantification analysis and machine learning techniques. Bull. Eng. Geol. Env. **81**(10), 398 (2022)
78. Song, K.; Yang, H.; Xie, J.; Karekal, S.: An optimization methodology of cutter-spacing for efficient mechanical breaking of jointed rock mass. Rock Mech. Rock Eng. **55**(6), 3301–3316 (2022)
79. Liu, B.; Yang, H.; Karekal, S.: Effect of water content on argillization of mudstone during the tunnelling process. Rock Mech. Rock Eng. **53**, 799–813 (2020)
80. Yang, H.; He, H.; Chen, C.: Effect of cutterhead rotational speed on mudstone argillization during the tunneling process. Bull. Eng. Geol. Environ. **82**(2), 45 (2023)
81. Saeidi, O., et al.: A stochastic penetration rate model for rotary drilling in surface mines. Int. J. Rock Mech. Min. Sci. **68**, 55–65 (2014)
82. Jahed Armaghani, D., et al.: An adaptive neuro-fuzzy inference system for predicting unconfined compressive strength and Young's modulus: a study on main range granite. Bull. Eng. Geol. Env. **74**, 1301–1319 (2015)
83. Armaghani, D.J., et al.: Development of hybrid intelligent models for predicting TBM penetration rate in hard rock condition. Tunn. Undergr. Space Technol. **63**, 29–43 (2017)
84. Taylor, K.E.: Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. Atmos. **106**(D7), 7183–7192 (2001)
85. Ikram, R. M. A.; Dai, H. L.; Al-Bahrani, M.; Mamlooki, M.: Prediction of the FRP reinforced concrete beam shear capacity by using ELM-CRFOA. Measurement **205**, 112230 (2022)
86. Ikram, R. M. A.; Hazarika, B. B.; Gupta, D.; Heddam, S.; Kisi, O.: Streamflow prediction in mountainous region using new machine learning and data preprocessing methods: a case study. Neural Comput. Appl. **35**(12), 9053–9070 (2023)
87. Ikram, R. M. A.; Mostafa, R. R.; Chen, Z.; Parmar, K. S.; Kisi, O.; Zounemat-Kermani, M.: Water temperature prediction using improved deep learning methods through reptile search algorithm and weighted mean of vectors optimizer. J. Mar. Sci. Eng. **11**(2), 259 (2023)
88. Verma, A.; Singh, T.: Prediction of water quality from simple field parameters. Environ. Earth Sci. **69**, 821–829 (2013)
89. Jafer, H.M., et al.: A statistical model for the geotechnical parameters of cement-stabilised hightown's soft soil: a case study of Liverpool, UK. Int. J. Civ. Environ. Struct. Constr. Archit. Eng. **10**(7), 885–890 (2016)
90. Shirani Faradonbeh, R.; Monjezi, M.; Jahed Armaghani, D.: Genetic programing and non-linear multiple regression techniques to predict backbreak in blasting operation. Eng. Comput. **32**, 123–133 (2016)