electronics

Article

A Novel Hybrid Intelligent Approach to Assess Blasting-Induced Overbreak Incorporating Geological Conditions in Different Tunnel Sections

Jiang Yuan [1], Qing Wang [1], Jianglu Wang [2,*], Yongqiang Fan [1], Weining Jiao [1] and Ang Li [2,*]

[1] CCCC Second Highway Engineering Co., Ltd., Xi'an 710065, China; jyuan@shbcccc.com (J.Y.); qing12wang@shbcccc.com (Q.W.); fyq23754647@shbcccc.com (Y.F.); weining250@shbcccc.com (W.J.)
[2] School of Highway, Chang'an University, Xi'an 710064, China
* Correspondence: wjl18480668530@chd.edu.cn (J.W.); angli_613@chd.edu.cn (A.L.); Tel.: +86-18480668530 (J.W.); +86-18091881556 (A.L.)

Abstract: Overbreak induced by tunnel blasting is a harmful phenomenon. Accurate assessment of overbreak can effectively reduce investment and ensure operational safety. In this study, a hybrid intelligent model for assessing blasting-induced overbreak is proposed which can accurately predict overbreak and effectively evaluate the importance of feature parameters. To ensure accurate prediction of overbreak, hyperparameters of four machine learning algorithms are optimized using a whale optimization algorithm. Their performance is compared based on three regression metrics: R^2, RMSE, and VAF. Given the limitations of traditional feature importance analysis methods, the Shapley Additive Explanation method is used in conjunction with the random forest algorithm. After accurately predicting overbreak caused by different sections of the tunnel, the impact of each input parameter on overbreak is analyzed, and recommendations for design values of certain significant parameters are provided. The research indicates that the proposed method can accurately predict overbreak caused by actual engineering blasts and provide insights into the selection of design parameter values.

Keywords: overbreak; machine learning; feature importance; geological condition

check for updates

Citation: Yuan, J.; Wang, Q.; Wang, J.; Fan, Y.; Jiao, W.; Li, A. A Novel Hybrid Intelligent Approach to Assess Blasting-Induced Overbreak Incorporating Geological Conditions in Different Tunnel Sections. *Electronics* **2024**, *13*, 4755. https://doi.org/10.3390/electronics13234755

Academic Editor: Ping-Feng Pai

Received: 3 October 2024
Revised: 11 November 2024
Accepted: 14 November 2024
Published: 2 December 2024

1. Introduction

The drilling and blasting method has been widely applied in tunnel excavation due to its low cost and unique rock fragmentation mechanism [1]. However, issues of overbreak and under-excavation frequently occur during actual blasting operations. Overbreak is defined as the actual excavation boundary exceeding the design limits, while under-excavation refers to the actual excavation boundary not reaching the design limits [2]. In engineering practice, overbreak is more prevalent, leading to resource wastage and potentially compromising the stability of surrounding rock, thereby posing threats to tunnel construction and operational safety. Consequently, there is an urgent need to optimize the design parameters of tunnel blasting to mitigate overbreak caused by blasting [3].

Overbreak is influenced by multiple factors, and its complexity aligns with the capabilities of artificial intelligence in addressing nonlinear problems, granting AI significant advantages in overbreak prediction and optimization [4]. Key influencing factors include geological conditions, borehole parameters, and charge configurations. On a technical level, the research has shifted from single machine learning algorithms to comparative analysis and the selection of multiple algorithms, combined with meta-heuristic algorithms to optimize hyperparameters for enhanced prediction performance. With the advancement of deep learning, new pathways have been opened for the study of overbreak in tunnel blasting. For instance, He et al. [5] combined meta-heuristic algorithms (GWO, WOA, TSA) with the random forest model to develop a novel tunnel blasting overbreak prediction model (RF-TSA), which demonstrates superior prediction accuracy and generalization capabilities compared to existing models. A hybrid model based on Support Vector Regression and a novel Multi-Objective Golden Jackal Optimization algorithm has been proposed by Li et al. [6] for predicting and optimizing the adverse effects of highway tunnel blasting excavation. The prediction results are presented in real time through a visualization program, providing intuitive data support for the optimization of tunnel blasting design.

Despite the significant achievements of machine learning technologies across various fields, the complexity of their models often renders them as "black boxes", resulting in poor interpretability of the output results [7]. In contrast, certain algorithms, such as tree-based models, exhibit strong intrinsic interpretability, aiding in the elucidation of decision-making processes. For instance, researches have utilized the interpretability of random forests (RFs) to rank factors influencing overbreak, identify key parameters, and emphasize the importance of design parameters in controlling overbreak [8]. For machine learning algorithms lacking interpretability, additional methods are required to clarify decision logic. After employing an adaptive neuro-fuzzy inference system (ANFIS) to predict overbreak, Mottahedi analyzed the influence of five parameters using the cosine amplitude method (CAM), finding that specific drilling parameters had the most significant impact on overbreak [9]. Currently, the interpretation of overbreak prediction models primarily relies on interpretable algorithms, such as decision trees, or employs sensitivity analysis formulas to assess influencing factors. For instance, Wu et al. [10] innovatively combined orthogonal experimental design with the random forest algorithm to conduct a sensitivity analysis of 11 factors influencing tunnel construction and determined the optimal parameter combinations, providing a theoretical basis for tunnel construction safety. After accurately predicting overbreak caused by geological conditions and blasting design parameters, optimizing design parameters to align with current geological conditions is crucial [11]. Numerous studies have focused on using predictive models to adjust input parameters in real time [12]. Furthermore, research indicates that meta-heuristic algorithms can efficiently and precisely identify optimal blasting parameters. Koopialipoor combined artificial neural networks (ANNs) with the artificial bee colony algorithm (ABC) to predict and optimize blasting parameters, significantly reducing the overbreak area, with optimized parameters resulting in a 47% reduction compared to unoptimized conditions [13]. It is noteworthy that research on utilizing meta-heuristic algorithms to identify optimal blasting parameters remains insufficient. For instance, the research by Liu et al. [14] has proposed a novel hybrid model that combines MLP with meta-heuristic algorithms to optimize tunnel blasting performance. By leveraging machine learning interpretability techniques, the study has quantitatively assessed the impact of diverse parameters on overbreak, ultimately revealing that geological conditions are the predominant determinant of overbreak.

The literature review indicates that despite significant progress in the field of overbreak prediction and optimization, several challenges remain to be addressed. First, although advanced intelligent technologies such as machine learning demonstrate potential in overbreak prediction, their application in this domain is still insufficient, necessitating further exploration and research to enhance prediction accuracy and optimization effectiveness. Second, traditional analyses of input parameter importance primarily rely on the interpretability of tree models or sensitivity analysis formulas, which exhibit inadequate stability, as minor data variations can lead to significant discrepancies in analytical outcomes [15]. Current methods only identify key features without clarifying the specific impact mechanisms of these features on prediction results [16]. Therefore, the development of machine learning models that can comprehensively assess parameter importance is an urgent requirement in current research [17].

This study constructs a tunnel blasting overbreak assessment model based on a hybrid intelligent algorithm, which integrates the Whale Optimization Algorithm (WOA) with four machine learning algorithms. By optimizing hyperparameters and evaluating the importance of feature parameters, the model achieves precise prediction of blasting overbreak. Compared with traditional methods, the proposed model demonstrates significant advantages in prediction accuracy and generalization capability. The performance of the model is comprehensively evaluated using three regression indicators, and the Shapley Additive Explanation method is combined with the random forest algorithm to provide a deeper analysis of the importance of feature parameters. Building on the accurate prediction of overbreak for different tunnel cross-sections, this study further analyzes the influence of input parameters on overbreak and proposes optimization suggestions for key parameters, providing important theoretical and technical support for tunnel blasting design. The structure of this paper is as follows: the second section provides a brief introduction to the database used; the third and fourth sections overview the selected intelligent algorithms and evaluate the most effective predictive model through multi-criteria comparison; the fifth section employs the SHAP method for post hoc analysis of complex models to investigate the mechanisms influencing overbreak in detail.

2. Materials and Methods

2.1. Machine Learning Predictions of Blasting-Induced Overbreak

An in-depth review and analysis of the preliminary literature presented in Table 1 have revealed that at the methodological level, certain limitations in the selection of methods by some researchers have been observed, or thorough optimization of the hyperparameters for the algorithms has not been conducted. Additionally, the evaluation system employed in regression analysis metrics has been found to be insufficient. In response to these findings, the Whale Optimization Algorithm has been innovatively introduced in this paper to comprehensively optimize the hyperparameters of four machine learning algorithms, significantly enhancing the model's generalization capability. Concurrently, a three-dimensional evaluation metric system has been adopted to conduct an all-encompassing and comprehensive assessment of the performance of each predictive model, aiming to derive more precise and reliable conclusions.

Table 1. Research on intelligent methods for overbreak prediction.

Research	Method	Number of Inputs	Number of Outputs	Regression Metrics
[5]	RF-GWO, RF-WOA, RF-TSA	7	1	R^2, RMSE, VAF, A-20index
[6]	MOGJO-SVR	11	4	R^2, RMSE
[9]	FST	9	1	R^2, RMSE
[11]	BO-RF, BO-SVM, BO-XGBOOST	8	1	R^2, RMSE, MAE
[13]	ANN-ABC	8	1	R^2, RMSE
[18]	PSO-DBP	12	4	/
[14]	MLP-PSO, MLP-WOA, MLP-AOA	12	4	R^2, MSE, MAE

2.2. Data Description

The causes of overbreak phenomena in blasting operations represent a complex process influenced by the interaction of multiple factors, making it insufficient to focus solely on the optimization of specific locations or single variables. Given that overbreak is affected by a combination of numerous variables, it is essential to construct a comprehensive data framework that encompasses various structural components of the tunnel and associated influencing factors. Considering the significant impact of geological parameters on overbreak and the differences in the formation mechanisms of different regions of the tunnel face, it is necessary to adopt parameter optimization strategies tailored to these distinct areas.

This study collected 95 groups of data from Zhang's paper as the dataset for this research, with dataset a containing 48 data points from the upper region of the tunnel face (from crown to shoulder) and dataset b containing 47 data points from the lower region (from shoulder to haunch) [18]. The data provided originate from the Panlongshan Tunnel, a highway tunnel located in Shandong, China. The tunnel's rock body consists of moderately weathered limestone with a medium-thin stratified structure. The interlayer bonding is of general quality, and the joint fissures are relatively well developed, showing clear cutting by horizontal stratification. The construction of the tunnel was predominantly executed using the upper and lower bench method. In the structural analysis of the tunnel face, data were collected from the upper region (dataset a) and the lower region (dataset b). The datasets include 12 input parameters and 1 output parameter. The input parameters include the uniaxial compressive strength of the surrounding rock (UCS) (MPa), the surrounding rock classification (SRG), the degree of bedding plane cohesion (Jd), the depth of burial (D) (m), the total number of blast holes (N), the spacing between peripheral holes (S) (cm), the spacing between auxiliary holes (Sr) (cm), burden (B) (cm), the total charge weight (Q) (kg), the concentration of charge in peripheral holes (P_c) (kg/m), the charge structure of peripheral holes (P_s), and the maximum charge weight per hole for slot blasting (Qc) (kg). The output parameter is defined as the area of overbreak under the influence of specific factors (OA) (m^2).

Through the analysis of Figure 1, it is observed that the selected dataset in this study not only encompasses parameters from various sections of the tunnel face but also equally includes four parameters within each of the three key influencing categories: geological parameters, borehole parameters, and charge parameters. This ensures a comprehensive and balanced exploration of factors affecting overbreak. To train the blasting prediction model, the original dataset was divided into two parts: 80% for the training set, used for model training, and 20% for the test set, used to evaluate model prediction performance.

Tables 2 and 3 list the statistical characteristics of input parameters from the crown to the shoulder and from the shoulder to the haunch, respectively. These include the mean, maximum and minimum values, variance, and standard deviation. This clarifies the range of each parameter and reveals the degree of dispersion through variance and standard deviation. For SRG, Jd, Ps, Qc, and Pc, the variance is less than 1, indicating a relatively limited range of values, while other parameters exhibit a broader range.

Table 2. Descriptive statistical characteristics of dataset from crown to shoulder.

Parameter	Mean	Variance	Standard Deviation	Max	Min
UCS (MPa)	43.64	54.48	7.38	53.60	31.70
SRG	3.46	0.25	0.50	4.00	3.00
Jd	0.69	0.06	0.25	1.00	0.30
D (m)	93.19	323.28	17.98	148.00	62.00
N	51.27	46.03	6.78	68.00	43.00
S (cm)	65.44	3.97	1.99	68.30	60.50
Sr (cm)	129.25	13.69	3.70	136.60	120.90
B (cm)	69.52	4.30	2.07	72.60	65.10
Q (kg)	41.39	9.23	3.04	45.00	37.50
P_s	0.32	0.09	0.31	1.00	0.15
Qc (kg)	3.26	0.20	0.45	3.90	2.40
P_c (kg/m)	0.18	0.0015	0.04	0.23	0.15
$OA(\text{m}^2)$	5.30	3.80	1.95	9.50	2.90

Table 3. Descriptive statistical characteristics of dataset from shoulder to haunch.

Parameter	Mean	Variance	Standard Deviation	Max	Min
UCS (MPa)	43.44	53.81	7.34	53.60	31.70
SRG	3.47	0.25	0.50	4.00	3.00
Jd	0.74	0.04	0.20	1.00	0.30

Table 3. *Cont.*

Parameter	Mean	Variance	Standard Deviation	Max	Min
D (m)	92.74	320.74	17.91	148.00	62.00
N	91.38	171.90	13.11	113.00	74.00
S (cm)	55.09	6.64	2.58	57.80	48.60
Sr (cm)	83.44	51.53	7.18	92.40	68.80
B (cm)	83.26	35.98	6.00	93.20	71.80
Q (kg)	180.30	70.77	8.41	189.30	166.20
P_s	0.78	0.12	0.35	1.00	0.23
Qc (kg)	3.17	0.20	0.45	3.90	2.40
P_c (kg/m)	0.25	0.0017	0.04	0.30	0.20
$OA\,(\mathrm{m}^2)$	2.57	0.59	0.77	4.10	1.50

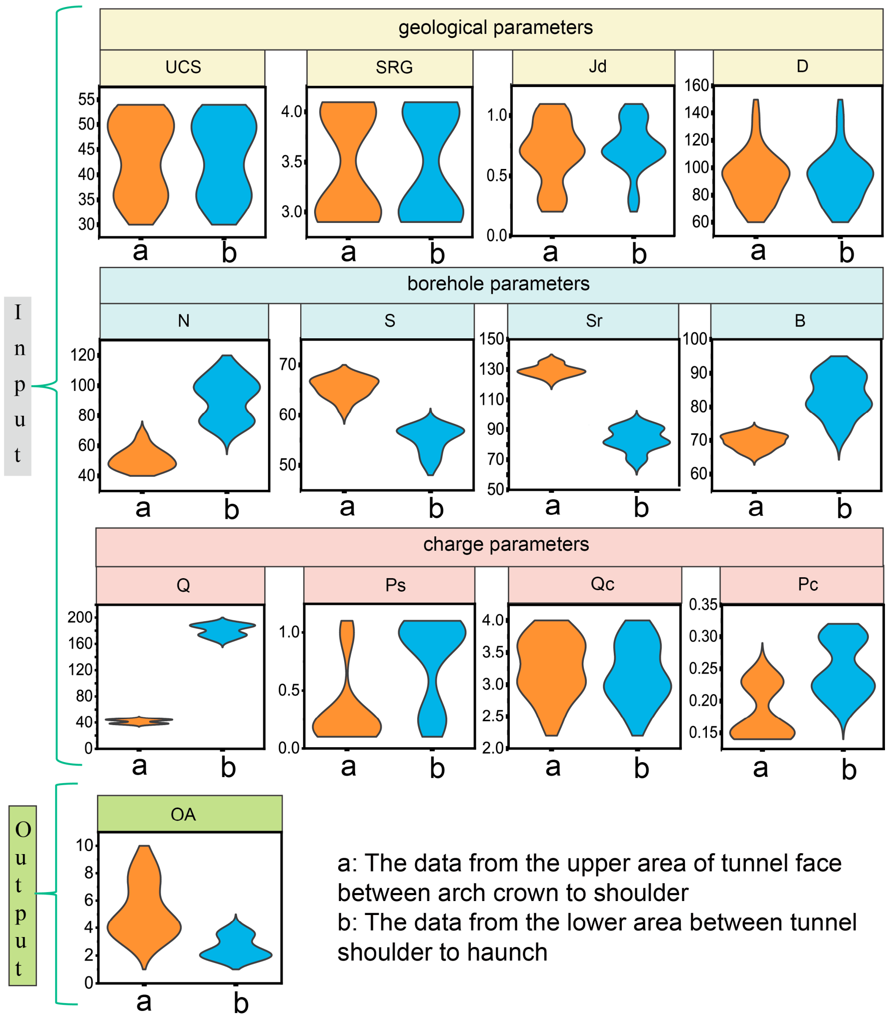

Figure 1. Distribution of various parameters across different sections of tunnel part.

2.3. Introduction to Machine Learning Algorithms

The widespread application of machine learning technologies in the field of civil engineering, including construction, tunnel, and bridge engineering, has yielded significant benefits [19]. For the model developed in this study, various predictive and optimization algorithms are available. A comparative analysis of four algorithms was conducted, and meta-heuristic methods were employed to optimize hyperparameters, aiming to enhance the performance of machine learning in predicting overbreak phenomena.

2.3.1. Random Forest

Random forest (RF) is an ensemble learning technique based on decision trees as its fundamental units [20].

Random forest was proposed by Breiman in 2001, integrating Bagging ensemble learning with random subspace techniques [21]. As illustrated in Figure 2, a random forest consists of multiple decision trees trained using Bagging. In classification tasks, the random forest aggregates the outputs of individual trees through a majority voting mechanism to determine the final classification, effectively enhancing the performance of a single decision tree and increasing the model's robustness against noise and outliers. Additionally, it demonstrates excellent scalability and parallel computing capabilities when handling high-dimensional data.

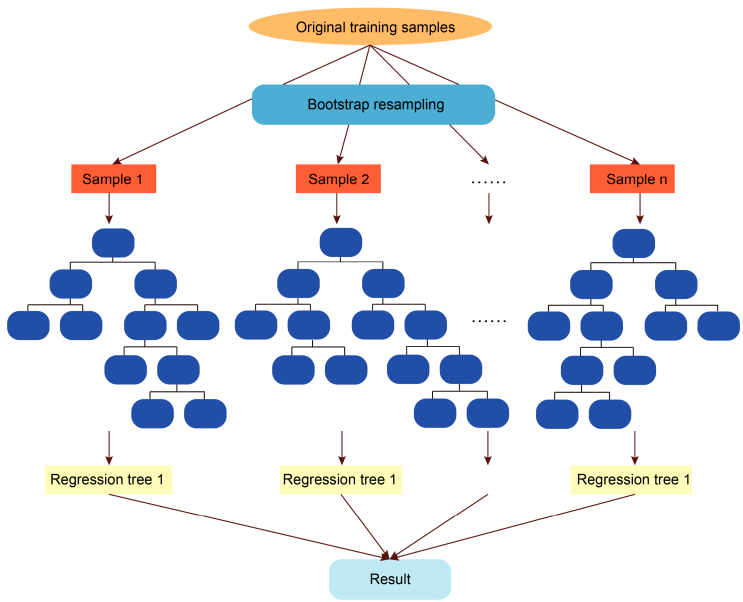

Figure 2. Structure of random forest regression prediction.

2.3.2. Gaussian Process

A Gaussian Process (GP) is a collection of random variables characterized by the property that any finite subset of these random variables follows a joint Gaussian distribution [22]. The random variables in a GP are uniquely determined by a mean function and a covariance function. Formally, a Gaussian process is defined as follows:

$$g(x) \sim GP(M(x), K(x, x_0)) \tag{1}$$

where:

$$\begin{cases} M(x) = E[g(x)] \\ K(x, x_0) = E\{\{g[x - M(x)]\}\{g(x_0 - M(x_0))\}\} \end{cases} \tag{2}$$

where x represents the exponentially weighted input.

2.3.3. SVM

A Support Vector Machine (SVM) is a machine learning algorithm constructed based on the VC dimension theory and the principle of structural risk minimization [23]. To address datasets that are challenging to manage in low-dimensional space, a SVM employs nonlinear mapping to transform the data from the input space into a high-dimensional feature space, where the optimal classification hyperplane is established [24]. This hyperplane not only accurately distinguishes training samples but also maximizes the classification margin, thereby enhancing prediction accuracy. The specific solution steps are outlined as follows.

The Lagrange function is introduced to transform the minimization problem into the corresponding dual problem, which is expressed as follows:

$$\begin{cases} \max L(a) = \sum\limits_{i=1}^{n} a_i - \dfrac{1}{2}\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} a_i a_j \beta_m \beta_n (\gamma_m^T \gamma_n) \\ S.T. \sum\limits_{i=1}^{n} a_i y_i = 0, a_i \geq 0 \end{cases} \tag{3}$$

where a_i and a_j are the Lagrange multipliers; γ_m and γ_n are the feature vectors of the m-th and n-th samples; β_m and β_n are the class labels of the m-th and n-th samples.

2.3.4. BP Neural Network

The BP (Backpropagation) neural network is a computational model that employs a multi-layer structure, proposed by Rumelhart and McClelland in the late 20th century, and is one of the most widely used neural networks today [25]. The BP network is essentially a mathematical simulation implemented through computer programs. Its core structure includes an input layer, at least one hidden layer, and an output layer, with the input layer located at the start of the network and the output layer at the end. The number of hidden layers can be configured based on research needs. A schematic diagram of the network structure is shown in Figure 3.

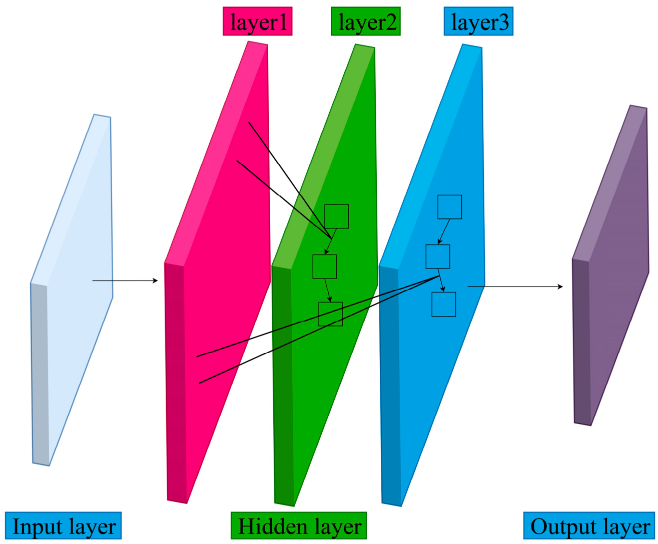

Figure 3. Structure of BP neural network model.

2.3.5. Whale Optimization Algorithm

Within the realm of meta-heuristic algorithms, the Whale Optimization Algorithm (WOA) demonstrates remarkable performance, inspired by the foraging behavior of humpback whales [26]. The WOA solves problems by simulating the random or optimal search behavior of humpback whales, along with their spiral bubble net feeding strategy. The algorithm is recognized for its simplicity, minimal parameter requirements, and strong optimization capabilities. The WOA mimics the search and encirclement process of humpback whales, which consists of three main stages: encircling prey, using bubble nets to capture prey, and searching for prey. During the implementation of the WOA, the position of each whale represents a candidate solution in the potential solution space, and the algorithm explores the global optimal solution by continuously updating the positions of the whales [27]. The computational flow of the standard WOA is illustrated in Figure 4.

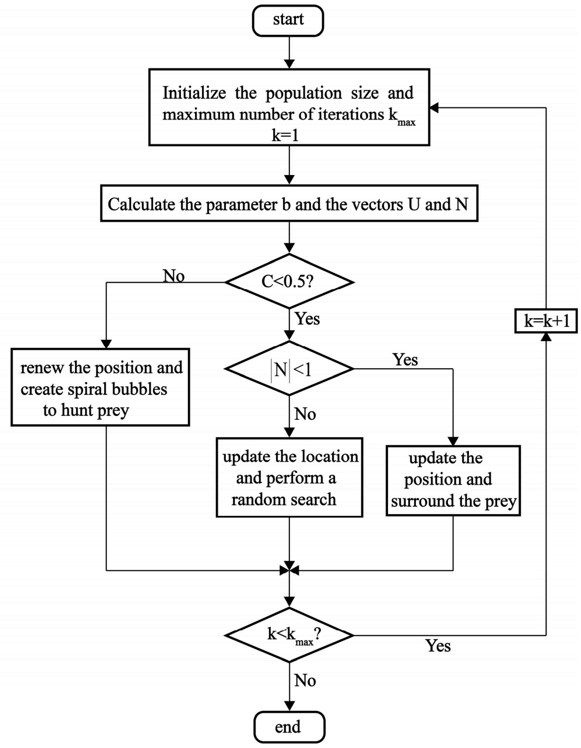

Figure 4. Computational process of WOA [28].

The Whale Optimization Algorithm (WOA) is recognized as an efficient meta-heuristic algorithm whose unique search mechanism and global optimization capabilities confer it with a pronounced advantage in the field of hyperparameter tuning [29]. In the present study, the WOA was employed to optimize the key hyperparameters of the random forest (RF) model, such as the number of trees and tree depth, thereby enhancing the model's classification accuracy while maintaining its complexity. For the Gaussian Process (GP) model, the parameters of the kernel function were effectively adjusted by the WOA, resulting in a reduction in the prediction error and an enhancement of the model's robustness. In the case of the Support Vector Machine (SVM) model, the WOA was utilized to finely tune the penalty and kernel function parameters, significantly improving the model's performance in handling nonlinear problems. The SVM model optimized by the WOA managed to maintain high classification accuracy while mitigating the risk of overfitting. For the Backpropagation (BP) neural network, the WOA played a pivotal role in adjusting hyperparameters such as the learning rate and the number of hidden layer nodes, enabling the network to achieve higher prediction accuracy within a shorter training time.

2.4. Comparison of Evaluation and Prediction Models

In the practice of machine learning, the appropriate configuration of hyperparameters is deemed crucial for enhancing the predictive accuracy of models. In this study, the Whale Optimization Algorithm (WOA) was employed during the construction of predictive models to improve their performance. Differences exist in hyperparameter optimization among various predictive models, and the impact of different hyperparameters on algorithm performance varies. This study referenced the findings of Zhang (2022) in the field of hyperparameter optimization for machine learning algorithms, and adjustments were made to those key hyperparameters within suitable ranges, aiming to optimize the performance of predictive algorithms and further enhance the accuracy of overbreak phenomenon predictions [30].

2.4.1. Generalization and Cross-Validation

In the field of machine learning, the evaluation of model performance focuses on its generalization ability, emphasizing generalization error rather than solely relying on the empirical error derived from training samples [31]. K-fold cross-validation is a widely accepted performance assessment method. As illustrated in Figure 5, this technique involves randomly partitioning the dataset into K mutually exclusive subsets, with K-1 subsets utilized for training and the remaining subset for validation. This process is repeated K times to ensure that each subset serves as the validation set in turn. The final score of model performance is derived from the average generalization error across the K iterations. In practical applications, K is typically set to 5 or 10.

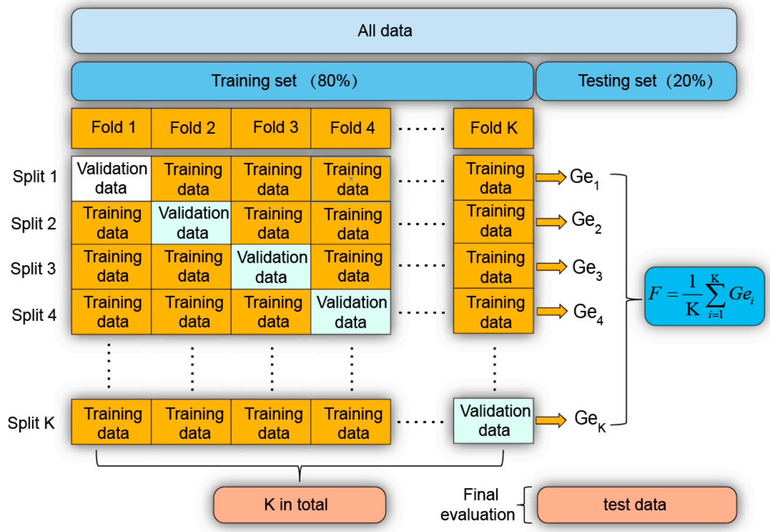

Figure 5. Schematic diagram of K-fold cross-validation.

This study aims to enhance prediction accuracy by integrating various predictive algorithms with the Whale Optimization Algorithm (WOA). If cross-validation is not employed, the resulting prediction accuracy will lack a solid scientific foundation. Therefore, a 10-fold cross-validation method has been implemented in this study to increase data validity and address the challenge posed by the limited volume of raw data. During the construction of the predictive models, the generalization error of the machine learning models effectively serves as the *fitness* function during the training phase. As shown in Figure 5, each iteration of cross-validation produces a value of generalization error, and the average generalization error from the K-fold cross-validation is utilized to evaluate the generalization performance of the model. The generalization error is associated with three primary parameters: *bias*, variance, and noise. *Bias* measures the deviation between the predicted and actual values during the training process, while variance quantifies the variability of results as the learning data and validation data change. Given the unpredictability of noise, this study focuses on the errors arising from bias and variance. The relevant formulas are expressed as follows:

$$fitness = bias^2 + var \tag{4}$$

In this study, the deviation between the model's predicted data and the actual data during the training phase is quantified using Mean Squared Error (MSE). Furthermore, the symbol *var* denotes the variance of the MSE values obtained throughout the cross-validation iterations. The following formulas are presented to evaluate the model's generalization capability:

$$F = \frac{1}{10}\sum_{i=1}^{10} Ge_i \tag{5}$$

2.4.2. Prediction Performance Evaluation Metrics

This study aims to develop an efficient, accurate predictive model with strong generalization capability. To achieve this, multiple machine learning algorithms were selected, and a unified optimization strategy was implemented across them. Considering that different algorithms may exhibit variability in generalization error, it is insufficient to determine the optimal model solely based on generalization error. Therefore, this study conducted detailed screening and further optimization of candidate predictive models based on evaluation metrics for prediction performance.

The variable R^2 reveals the degree of consistency between the actual values and the predicted values. Root Mean Squared Error (RMSE), widely utilized as a performance metric in regression analysis, quantifies the magnitude of bias in model predictions and assigns greater weight to larger errors. Therefore, a lower RMSE value indicates higher predictive accuracy of the model. An increase in the Variance Accounted For (VAF) enhances the model's explanatory power concerning real data, thereby making the model's predictions more closely aligned with actual observed data.

In this study, the performance evaluation of the predictive models incorporates the statistical metrics R^2, RMSE, and VAF. The corresponding formulas are expressed as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(OA_i - \hat{OA}_i\right)^2}{\sum_{i=1}^{N}\left(OA_i - \overline{OA_i}\right)^2} \tag{6}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(OA_i - \hat{OA}_i\right)^2} \tag{7}$$

$$VAF = \left(1 - \frac{\mathrm{var}(OA_i - \hat{OA}_i)}{\mathrm{var}(OA_i)}\right) \times 100 \tag{8}$$

where OA_i represents the original values of the overbreak area; $\hat{OA}_i$ denotes the predicted values of the overbreak area; $\overline{OA_i}$ signifies the average value of the overbreak area; and N indicates the data volume.

2.4.3. Initial Parameter Configuration

In this study, key parameters of the Whale Optimization Algorithm (WOA) were predetermined to facilitate hyperparameter optimization within the machine learning framework. The selection of these parameters significantly influences algorithm performance, including population size, the number of iterations, and random weights, which collectively determine the convergence rate and predictive accuracy of the algorithm. Within the WOA, the random weight parameter is particularly critical, as it regulates the exploration and exploitation phases of the algorithm. During the exploration phase, whale agents randomly search for prey, while concentrated searching occurs during the exploitation phase. The value of parameter a decreases linearly from 2 to 0 throughout the iterations, enabling a smooth transition from exploration to exploitation.

Regarding the selection of population size, this study draws on previous research indicating that the number of particles used in most optimization models typically does not exceed 100 [32]. This finding provides a reference for determining population size. Therefore, considering the diversity of variables, this study opted for 50 particles. The data from Tables 3 and 4 indicate that models utilizing this particle count exhibit high accuracy in predictions.

Table 4. The performance of each predicted model from crown to shoulder.

	Training Set				Testing Set			
	RF	GP	SVM	BP	RF	GP	SVM	BP
R^2	0.982	0.967	0.996	0.987	0.768	0.423	0.816	0.815
VAF	98.36%	96.74%	100%	98.69%	77.78%	44.49%	87.75%	82.52%
RMSE	0.210	0.362	0.004	0.230	0.831	1.311	0.659	0.742

The setting of the number of iterations requires a balance between optimization effectiveness and computational cost. While increasing the number of iterations may enhance optimization results, it concurrently leads to increased computation time. Thus, the number of iterations should be reasonably controlled to ensure predictive accuracy while managing training duration. By analyzing the fitness function variation curve (as shown in Figure 6), an appropriate number of iterations can be determined. When the curve stabilizes, it indicates that model training has reached a steady state. In this study, a threshold was established at which the curve showed stability, and to ensure the stability of model training, the number of iterations was appropriately increased, ultimately setting it at 160 iterations.

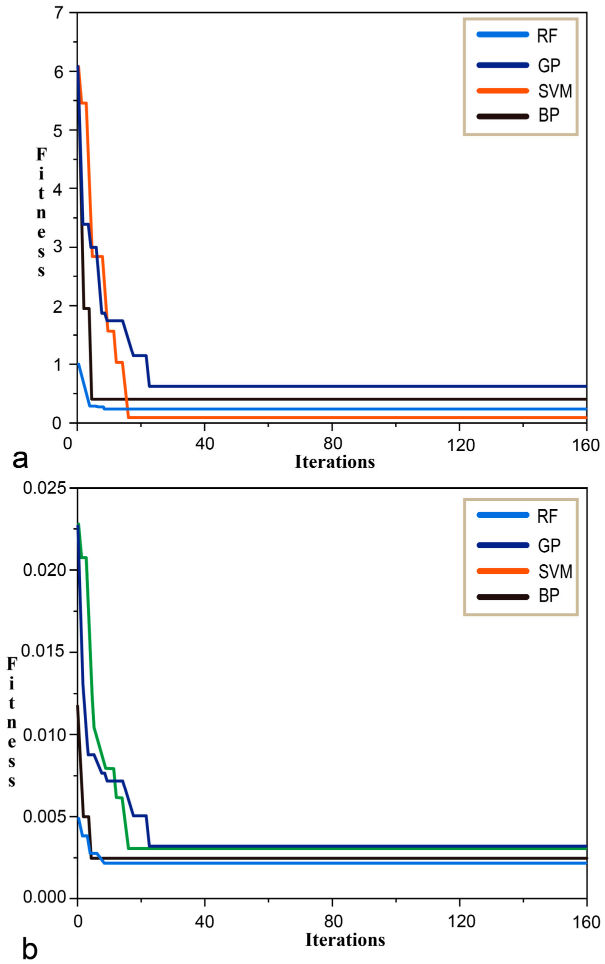

Figure 6. Visualization of the WOA search process for optimal hyperparameters: (**a**) from crown to shoulder; (**b**) from shoulder to haunch.

3. Experimental Results

3.1. Loss Function

To establish the most accurate predictive model for overbreak, this study conducted a detailed comparison and selection of models based on three predictive evaluation metrics. To ensure fairness and consistency in the comparisons, all calculations were performed on the same computational platform. The 12 input parameters illustrated in Figure 1 were used as benchmarks for model comparison, while output parameters served as the basis for evaluation. Ultimately, the predictive accuracy of four machine learning models optimized by the WOA was compared, resulting in the development of various overbreak prediction strategies by integrating data from the apex to the shoulder and from the shoulder to the waist.

Figure 6 visualizes the process of the WOA in searching for optimal hyperparameters across different predictive evaluation algorithms. It was observed that the influence of data has a significant effect on convergence outcomes. As shown in Figure 6a, the final convergence value of the GP algorithm was the highest, whereas in Figure 6b, the convergence value of the SVM algorithm surpassed that of the GP model. The overall convergence speed of the curve in Figure 6a was superior to that in Figure 6b. In both datasets, the BP algorithm exhibited the most significant performance improvement through hyperparameter optimization; in dataset a, the fitness value rapidly decreased from 6.138 to 0.424 within fewer than 10 iterations, and in dataset b, it declined from 0.0226 to 0.0054 within 10 iterations, demonstrating a change magnitude that exceeded that of other algorithms during the iterative process. Moreover, in both datasets, the BP algorithm consistently showed the fastest convergence speed among the tested algorithms.

3.2. Performance of Each WOA Machine Learning Model

After 160 iterations, this study selected the optimal hyperparameters for the random forest (RF), Gaussian Process (GP), Support Vector Machine (SVM), and Backpropagation (BP) algorithms, thereby constructing WOA-optimized models for overbreak prediction. The predictive accuracy of the models was assessed using the three evaluations metrics described in Section 2.4.2, which comprehensively utilized these performance indicators to evaluate the models from multiple perspectives.

Tables 4 and 5 present the evaluation metrics corresponding to all models. By comparing the data in Tables 4 and 5, it can be observed that the metrics for both datasets exhibit remarkable similarity on the training set, indicating that the selected models possess good robustness. It is important to note that not all metrics benefit from maximization; for the R^2 and VAF metrics, higher values indicate better model performance, whereas for the RMSE metric, lower values signify superior model performance.

Table 5. The performance of each predicted model from shoulder to haunch.

	Training Set				**Testing Set**			
	RF	GP	SVM	BP	RF	GP	SVM	BP
R^2	0.962	0.939	0.999	0.966	0.794	0.696	0.847	0.789
VAF	96.24%	93.87%	99.99%	96.63%	80.09%	74.43%	83.24%	88.56%
RMSE	0.154	0.194	0.007	0.145	0.311	0.378	0.268	0.315

Based on the data analysis from Tables 4 and 5, the GP algorithm exhibited lower R^2 and VAF values compared to the other algorithms, while its RMSE value was higher, indicating that the GP algorithm performs the worst among the four algorithms and is therefore excluded first. The BP algorithm showed relatively stable R^2 and VAF values; however, in the training set from the apex to the shoulder, its RMSE value varied significantly from 0.230 to 0.742, representing the largest fluctuation among the remaining three algorithms, leading to the exclusion of BP as well. In contrast, the RF and SVM algorithms displayed more stable R^2 values, with relatively stable variations in VAF and RMSE; however, between the two, the SVM algorithm demonstrated relatively greater stability in performance.

3.3. Quantifying the Impact of Input Parameters on Overbreak

After accurately predicting the overbreak area following blasting excavation, it is essential to further analyze the influence of each input parameter on the overbreak results. To this end, this study conducted a parameter importance analysis of factors affecting overbreak. Although traditional feature importance analyses can identify which features are significant, they do not provide in-depth explanations of how these features specifically affect the predictive output. The intrinsic interpretability of tree-based models, a method that is subject to considerable instability, has been discussed, with attention drawn to the fact that minor data variations can lead to entirely different outcomes [33]. The primary advantage of SHAP (SHapley Additive exPlanations) values lies in their ability to reveal the impact magnitude of each feature within each sample, clearly indicating whether these effects are positive or negative [34]. For post hoc model interpretation, SHAP values must satisfy three fundamental properties: local accuracy, missingness, and consistency. The marginal contribution of features to the output results is primarily assessed based on the magnitude of the SHAP values.

The overall model predictions can be represented by the following equation:

$$f(x) = \gamma_0 + \sum_{i=1}^{m} \gamma_i \tag{9}$$

where γ_i represents the influence of each feature on the prediction, and the output of the black-box model can be decomposed into the sum of the contributions from each feature γ_i.

The traditional KernelSHAP method is computationally expensive; however, TreeSHAP, as an efficient variant of SHAP, not only significantly improves computational speed but also accurately estimates Shapley values. TreeSHAP, introduced by Lundberg et al. in 2018, is specifically designed for tree-based machine learning models [35]. Given that the random forest (RF) model employed in this study is also tree-based and demonstrates excellent performance in data fitting according to the data in Tables 3 and 4, the RF model was utilized during the data training phase, followed by an analysis of the parameters affecting overbreak using TreeSHAP. After training the WOA-RF model with 50 particles over 160 iterations, TreeSHAP was introduced for post hoc analysis, providing targeted insights into the parameters influencing tunnel overbreak from the perspective of model interpretability.

4. Discussion

Figure 7 presents the results of the SHAP feature importance analysis, revealing that geological parameters are identified as the most critical influencing factors in both dataset a and dataset b. This finding further underscores the necessity of parameter optimization for different sections of the tunnel. Since the rock mass properties are fixed and cannot be altered, understanding them can only be achieved through investigation methods. Therefore, the selection of blasting hole parameters and charge parameters that align with the geological parameters becomes particularly crucial during the blasting design process. Customizing specific blasting parameters for different surrounding rock sections is indispensable. While traditional analysis methods can indicate the importance of a parameter, they often struggle to reveal its interactions with other parameters. By leveraging the advantages of SHAP, which combines parameter importance with its effects, we obtained a summary plot of SHAP values for each feature, providing deeper insights into understanding the overall model.

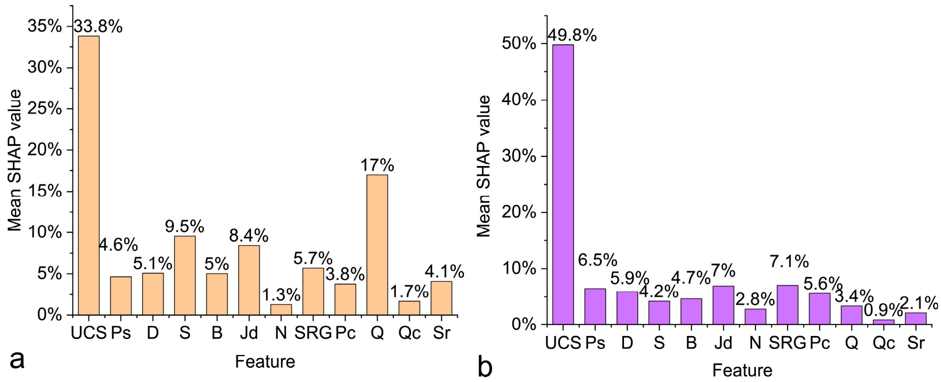

Figure 7. The percentage of importance for each parameter obtained through SHAP analysis: (**a**) from the apex to the shoulder; (**b**) from the shoulder to the waist.

In Figure 8a,b, the color gradient from blue to red indicates the variation in feature values from low to high, with overlapping points exhibiting fluctuations on the vertical axis. The analysis results reveal a negative correlation between the uniaxial compressive strength (UCS) of the rock and the predicted overbreak area; specifically, as the UCS increases, the predicted overbreak area decreases, indicating that higher rock hardness correlates with reduced overbreak. Conversely, when the UCS is lower, overbreak tends to increase. However, the hardness of the rock is generally unchangeable in engineering practice, thus underscoring the importance of designing other blasting parameters.

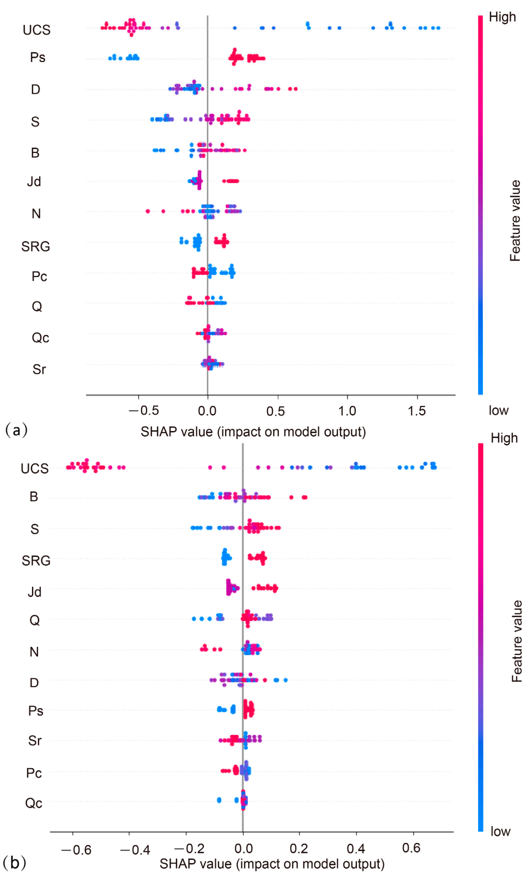

Figure 8. The SHAP analysis: (**a**) from the apex to the shoulder; (**b**) from the shoulder to the waist.

In Figure 8a, the relationship between the decoupling coefficient of the explosive charge's axial distribution (Ps) and overbreak is particularly clear. Especially in the region from the apex to the shoulder, lower decoupling coefficients contribute to reduced overbreak. A similar trend is observed in Figure 8b, but the impact of Ps in the apex region is more pronounced. The degree of structural surface bonding should not be excessively high at the apex; rocks with a high degree of bonding are difficult to fragment during blasting, potentially leading to overbreak. Parameters B and S should be analyzed together due to their interdependence. The symbol S represents the spacing between peripheral holes (cm), and B denotes the burden (cm). Figure 8a indicates that a larger S value combined with a smaller B value can reduce overbreak from the apex to the shoulder. In contrast, Figure 8b suggests that a combination of smaller S and B values yields better blasting results. According to the study by Liu (2024), for level III surrounding rock, the S/B ratio should be within 0.8 to 1.0, while for level IV surrounding rock, a smaller value enhances blasting effectiveness [14].

The effect of burial depth (D) on overbreak at the apex should not be overlooked; smaller D values are beneficial in reducing overbreak. D is related to in situ stress, and when underground excavation approaches high-stress zones, the influence of in situ stress on overbreak becomes significant. Thus, a smaller burial depth helps to mitigate apex overbreak.

It is noteworthy that the lower SRG values in Figure 8 actually reflect higher surrounding rock grades. Since the hardness of level III surrounding rock is greater than that of level IV, the numerical value of 3 is less than 4, which manifests in the figure as lower SRG values corresponding to smaller overbreak. In reality, the higher the surrounding rock grade, the smaller the overbreak.

The influence of auxiliary hole spacing is particularly significant in the waist section. Other parameters have a relatively minor overall impact on overbreak. Notably, for slot holes, their effect on under-extraction is mainly indirect, as they influence the formation of blasting-free surfaces, thereby affecting the contours of over- and under-extraction.

The visualization of feature attribution can be realized through the concept of force, where each feature is represented as a force that enhances or diminishes the predicted outcome. The interactions of these forces reach a balance within the dataset. Figure 9 illustrates the visualization of each feature's force on overbreak prediction for the ninth group of data in the test set. From Figure 9a, the average predicted value for this instance is 4.55, where the red area represents forces that push the predicted value upwards, while the blue area indicates forces that lower the predicted value. Here, UCS and Ps are identified as the two most influential parameters, with UCS tending to decrease the predicted value, while Ps partially offsets this decrease and tends to increase the predicted value. The excessively high values of Ps and S contribute to increased overbreak.

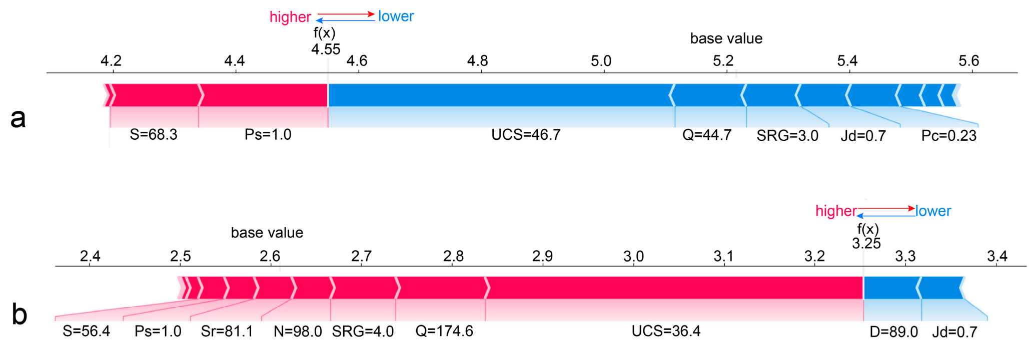

Figure 9. The overmining predictions using SHAP values: (**a**) from the apex to the shoulder; (**b**) from the shoulder to the waist.

In Figure 9b, the average predicted value is 3.25, and it is evident that all listed features tend to enhance the predicted value. When interpreting Figure 9, it is essential to consider the information from Figure 1. Figure 1 indicates that the UCS value is low, the SRG value is also low, the S value is relatively high within the dataset, the Sr value is located at the lower end of the dataset, and the Ps value is at the upper end, while the Q value is in the middle. According to the conclusions drawn from Figure 8, these feature configurations are detrimental to reducing overbreak and are the reasons for the positive influence on the predicted results.

5. Conclusions

This study utilizes 95 sets of data from different regions of the tunnel face to train four machine learning algorithmic models, utilizing the Whale Optimization Algorithm (WOA) for hyperparameter optimization, with the aim of enhancing the predictive performance of the models. The evaluation of model performance considered multiple metrics, including accuracy, Root Mean Square Error (RMSE), and Variance Accounted For (VAF). The results indicate that the random forest (RF) and Support Vector Machine (SVM) models demonstrate stable performance, making them suitable for subsequent research.

Given the inherent interpretability limitations and accuracy issues associated with machine learning algorithms, this study introduced SHapley Additive exPlanations (SHAP) as a method for post hoc interpretation of complex models. It was found that geological parameters play a crucial role among the factors influencing blast overbreak, particularly highlighting the significant impact of Unconfined Compressive Strength (UCS) on overbreak. Through SHAP analysis, this study explored the relationship between parameter values and overbreak outcomes, providing local explanations to further clarify the specific directional influence of parameter values (either increasing or decreasing overbreak). As UCS increases, i.e., with harder rock, the predicted overbreak area decreases. Conversely, with lower UCS, the prediction is the opposite. Other design parameters must be determined based on the rock conditions at the excavation tunnel face.

Author Contributions: Conceptualization, Q.W. and Y.F.; methodology, Q.W.; software, W.J.; validation, J.Y., J.W., A.L., Q.W. and Y.F.; formal analysis, W.J.; investigation, J.Y.; resources, Q.W.; data curation, Y.F.; writing—original draft preparation, J.Y.; writing—review and editing, J.W. and A.L.; visualization, Q.W.; supervision, Y.F.; project administration, W.J.; funding acquisition, Q.W. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the Research Funds of the Department of Transport of Shaanxi Province (No. 23-81X).

Data Availability Statement: The derived data supporting the findings of this study are available from the corresponding author on request.

Conflicts of Interest: The authors J.Y., Q.W., Y.F., and W.J. are employed by the company CCCC Second Highway Engineering CO., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. Li, X.; Liu, K.; Sha, Y.; Yang, J.; Hong, Z. Experimental and numerical investigation on rock fracturing in tunnel contour blasting under initial stress. *Int. J. Impact Eng.* **2024**, *185*, 104844. [CrossRef]
2. Cui, J.; Yang, Z.; Azzam, R. Field measurement and numerical study on the effects of under-excavation and overbreak on ultra-deep foundation pit in coastal area. *J. Mar. Sci. Eng.* **2023**, *11*, 219. [CrossRef]
3. Liu, X.; Tao, T.; Tian, X.; Lou, Q.; Xie, C. Layout method and numerical simulation study of reduced-hole blasting in large-section tunnels. *Front. Earth Sci.* **2022**, *10*, 976419. [CrossRef]
4. Jang, H.; Kawamura, Y.; Shinji, U. An empirical approach of overbreak resistance factor for tunnel blasting. *Tunn. Undergr. Space Technol.* **2019**, *92*, 103060. [CrossRef]
5. He, B.; Jahed Armaghani, D.; Lai, S.H. Assessment of tunnel blasting-induced overbreak: A novel metaheuristic-based random forest approach. *Tunn. Undergr. Space Technol.* **2023**, *133*, 104979. [CrossRef]
6. Li, C.; Zhou, J. Prediction and optimization of adverse responses for a highway tunnel after blasting excavation using a novel hybrid multi-objective intelligent model. *Transp. Geotech.* **2024**, *45*, 101228. [CrossRef]
7. L'heureux, A.; Grolinger, K.; Elyamany, H.F.; Capretz, M.A. Machine learning with big data: Challenges and approaches. *IEEE Access.* **2017**, *5*, 7776–7797. [CrossRef]
8. Neto, M.P.; Paulovich, F.V. Explainable matrix-visualization for global and local interpretability of random forest classification ensembles. *IEEE Trans. Vis. Comput. Graph.* **2020**, *27*, 1427–1437. [CrossRef]

9. Mohammadi, M.; Hossaini, M.F.; Mirzapour, B.; Hajiantilaki, N. Use of fuzzy set theory for minimizing overbreak in underground blasting operations—A case study of Alborz Tunnel, Iran. *Int. J. Mining. Sci. Technol.* **2015**, *25*, 439–445. [CrossRef]
10. Wu, B.; Sun, W.; Meng, G. Sensitivity Analysis of Influencing Factors of Karst Tunnel Construction Based on Orthogonal Tests and a Random Forest Algorithm. *Appl. Sci.* **2024**, *14*, 2079. [CrossRef]
11. Hong, Z.; Tao, M.; Liu, L.; Zhao, M.; Wu, C. An intelligent approach for predicting overbreak in underground blasting operation based on an optimized XGBoost model. *Eng. Appl. Artif. Intell.* **2023**, *126*, 107097. [CrossRef]
12. Kwak, Y.; Huh, J.H.; Jang, C. Development of a model predictive control framework through real-time building energy management system data. *Appl. Energy.* **2015**, *155*, 1–13. [CrossRef]
13. Koopialipoor, M.; Ghaleini, E.N.; Tootoonchi, H.; Jahed Armaghani, D.; Haghighi, M.; Hedayat, A. Developing a new intelligent technique to predict overbreak in tunnels using an artificial bee colony-based ANN. *Environ. Earth Sci.* **2019**, *78*, 165. [CrossRef]
14. Liu, Y.; Li, A.; Dai, F.; Jiang, R.; Liu, Y.; Chen, R. An AI-powered approach to improving tunnel blast performance considering geological conditions. *Tunn. Undergr. Space Technol.* **2024**, *144*, 105508. [CrossRef]
15. Antoniadis, A.; Lambert-Lacroix, S.; Poggi, J.M. Random forests for global sensitivity analysis: A selective review. *Reliab. Eng. Syst. Saf.* **2021**, *206*, 107312. [CrossRef]
16. Joshi, G.; Walambe, R.; Kotecha, K. A review on explainability in multimodal deep neural nets. *IEEE Access.* **2021**, *9*, 59800–59821. [CrossRef]
17. Wang, S.; Jin, S.; Bai, D.; Fan, Y.; Shi, H.; Fernandez, C. A critical review of improved deep learning methods for the remaining useful life prediction of lithium-ion batteries. *Energy Rep.* **2021**, *7*, 5562–5574. [CrossRef]
18. Zhang, W. Research on Blasthole Image Recognition Algorithms and Optimization of Smooth Blasting Parameters of Rock Tunnel. Ph.D. Dissertation, Shandong University, Jinan, China, 2019. [CrossRef]
19. Chun, P.J.; Hayashi, S. Development of a concrete floating and delamination detection system using infrared thermography. *IEEE/ASME Trans. Mechatron.* **2021**, *26*, 2835–2844. [CrossRef]
20. Lee, J.; Wang, W.; Harrou, F.; Sun, Y. Reliable solar irradiance prediction using ensemble learning-based models: A comparative study. *Energy Convers. Manag.* **2020**, *208*, 112582. [CrossRef]
21. Kotsiantis, S. Combining bagging, boosting, rotation forest and random subspace methods. *Artif. Intell. Rev.* **2011**, *35*, 223–240. [CrossRef]
22. Evangelisti, G.; Hirche, S. Data-driven momentum observers with physically consistent Gaussian processes. *IEEE Trans. Robot.* **2024**, *40*, 1938–1951. [CrossRef]
23. Dibike, Y.B.; Velickov, S.; Solomatine, D.; Abbott, M.B. Model induction with support vector machines: Introduction and applications. *J. Comput. Civ. Eng.* **2001**, *15*, 208–216. [CrossRef]
24. Chen, J.H.; Ong, C.F.; Zheng, L.; Hsu, S.C. Forecasting spatial dynamics of the housing market using support vector machine. *Int. J. Strateg. Prop. Manag.* **2017**, *21*, 273–283. [CrossRef]
25. Gao, X.; Huang, T.; Wang, Z.; Xiao, M. Exploiting a modified gray model in back propagation neural networks for enhanced forecasting. *Cogn. Comput.* **2014**, *6*, 331–337. [CrossRef]
26. Chakraborty, S.; Sharma, S.; Saha, A.K.; Saha, A. A novel improved whale optimization algorithm to solve numerical optimization and real-world applications. *Artif. Intell. Rev.* **2022**, *55*, 4605–4716. [CrossRef]
27. Abdel-Basset, M.; Mohamed, R.; Sallam, K.M.; Chakrabortty, R.K.; Ryan, M.J. An efficient-assembler whale optimization algorithm for DNA fragment assembly problem: Analysis and validations. *IEEE Access.* **2020**, *8*, 222144–222167. [CrossRef]
28. Gharehchopogh, F.S.; Gholizadeh, H. A comprehensive survey: Whale Optimization Algorithm and its applications. *Swarm Evol. Comput.* **2019**, *48*, 1–24. [CrossRef]
29. Hassan, A.A.; Abdullah, S.; Zamli, K.Z.; Razali, R. Combinatorial test suites generation strategy utilizing the whale optimization algorithm. *IEEE Access* **2020**, *8*, 192288–192303. [CrossRef]
30. Zhang, Z.H.; Guo, F.; Xu, Z.; Yang, X.Y.; Wu, K.Z. On retrieving the chromium and zinc concentrations in the arable soil by the hyperspectral reflectance based on the deep forest. *Ecol. Indic.* **2022**, *144*, 109440. [CrossRef]
31. Wu, X.; Manton, J.H.; Aickelin, U.; Zhu, J. On the generalization for transfer learning: An information-theoretic analysis. *IEEE Trans. Inf. Theory.* **2024**, *70*, 7089–7124. [CrossRef]
32. Zhang, Y.; Wang, S.; Ji, G. A comprehensive survey on particle swarm optimization algorithm and its applications. *Math. Probl. Eng.* **2015**, *2015*, 931256. [CrossRef]
33. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable AI: A review of machine learning interpretability methods. *Entropy* **2021**, *23*, 18. [CrossRef] [PubMed]
34. Kim, Y.; Kim, Y. Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models. *Sustain. Cities Soc.* **2022**, *79*, 103677. [CrossRef]
35. Bifarin, O.O. Interpretable machine learning with tree-based SHapley Additive exPlanations: Application to metabolomics datasets for binary classification. *PLoS ONE* **2023**, *18*, e0284315. [CrossRef]