



OPEN Predicting water-conducting fracture zone height in three-soft coal seams using a BOA-MLP model

Zheng Tang¹, Haochen Tong¹, Dongsheng Zhang^{1✉}, Gangwei Fan¹, Chuangnan Ren¹, Yinglong Sun², Xinyao Gao¹, Yu Zhang² & Ziquan Wu¹

The height of water-conducting fracture zone (WCFZ) critically impacts coal mine safety and water hazard control. This study develops a Bayesian Optimization Algorithm-optimized Multilayer Perceptron (BOA-MLP) model to predict WCFZ height in three-soft coal seams in Henan Province. The Analytic Hierarchy Process-Entropy Weight Method integration identified mining height, face length, and burial depth as primary factors. Trained on 32 field-measured WCFZ samples, BOA intelligently optimized hyperparameters (hidden layer neurons, learning rate, batch size). The optimal configuration (64 neurons in the first hidden layer, 32 neurons in the second hidden layer, learning rate of 0.001, batch size of 32) yielded test-set RMSE = 1.98 m, MAE = 2.23 m, MAPE = 2.67%, $R^2 = 0.973$ —outperforming manual MLP with 15.0–28.5% error reductions and a 6.06% R^2 improvement. Statistical analysis using the Wilcoxon signed-rank test confirmed these improvements are significant ($p = 0.027$) with large effect size (Cohen's $d = 0.83$). Comprehensive uncertainty and sensitivity analyses demonstrated the model's robustness, with 84.4% of predictions falling within 95% confidence intervals and burial depth identified as the most influential parameter. Field validation at Panel 15,030 in Yaoling Mine confirmed the accuracy of BOA-MLP: predicted 28.1 m vs. measured 27.3 m (2.9% error), surpassing empirical formulas (18.3%) and manual MLP (7.3%). These results demonstrate that integrating Bayesian optimization substantially enhances the stability, generalizability, and predictive accuracy of the MLP model. The proposed BOA-MLP approach thus offers a robust and reliable methodology for forecasting WCFZ heights in three-soft coal seams, with significant implications for mine water management and safety planning.

Keywords Water-conducting fracture zone, Integrated subjective-objective weighting, Bayesian optimization algorithm, Neural network, Height prediction

Coal seam mining induces damage to the overlying strata, forming a WCFZ comprising caved and fractured zones. When groundwater or overlying aquifers infiltrate the mine workings through this zone, it can trigger surface water resource depletion and mine water inrush accidents^{1,2}. Under conditions characterized by weak roof, coal, and floor strata—termed “three-soft” coal seams—the low strength and high friability of the rock mass exacerbate the development scale of overburden fractures and the associated water-conducting risks. Predicting the height of the WCFZ in such seams is inherently challenging, compounded by the enhanced connectivity of the fracture network, which poses significant threats to aquifer protection. Consequently, accurate prediction of WCFZ height in three-soft coal seams remains a critical focus in mine water hazard prevention and a persistent challenge in water-preserved coal mining research^{3,4}.

Current methods for predicting WCFZ height primarily include empirical formulas, numerical simulation, field measurements, theoretical analysis, and physical similarity modeling. Empirical formulas typically rely on the “Three Under” code's simplified equations, which consider only mining height and thus fail to capture the complexity of modern mining conditions⁵. While field measurements provide precise height determination, they are operationally cumbersome and costly^{6,7}. The accuracy of physical similarity modeling is heavily contingent upon material proportions and demands substantial human and material resources^{8,9}. Numerical simulation accuracy depends critically on model parameter selection, introducing significant subjectivity^{10,11}. Theoretical

¹School of Mines, China University of Mining and Technology, Xuzhou 221116, China. ²Gongyi Yaoling Coal Industry Co., Ltd, Gongyi 450000, China. ✉email: ts24020083a31@cumt.edu.cn

analysis models often suffer from excessive idealization, failing to reflect actual complex hydrogeological conditions^{12,13}. Since the early 2010s, the rapid advancement of artificial intelligence has prompted researchers to apply machine learning (ML) algorithms to WCFZ height prediction, yielding numerous practically significant results^{14–19}. Wang et al. employed a Sparrow Search Algorithm (SSA) to optimize hyperparameters (e.g., number of trees, maximum depth) in a Random Forest (RF) model, establishing an SSA-RF predictive framework²⁰. However, this model requires constructing numerous decision trees, leading to computational complexity and limited generalization capability. Zhang et al. developed a Support Vector Machine (SVM) prediction model, enhancing its accuracy by optimizing penalty factors and kernel parameters using an improved Fruit Fly Optimization Algorithm (FOA)²¹. Xun et al. further improved model performance by introducing a Least Squares SVM (LSSVM) optimized with an Adaptive Particle Swarm Optimization (APSO) algorithm²². Li et al. established a Backpropagation Neural Network (BPNN) model based on field measurement data²³. Wu et al. compared a multivariate nonlinear regression model with a Genetic Algorithm-optimized BPNN (GA-BPNN) model, finding the latter offered superior accuracy²⁴. Liu et al. used the Synthetic Minority Over-sampling Technique for Regression with Gaussian Noise (SMO-GN) to expand small datasets and employed a Mutation Particle Swarm Optimization (MPSO) algorithm to refine a BPNN model²⁵.

Despite these advancements, significant limitations persist in current WCFZ height prediction research. First, previous studies lack models specifically tailored to the unique geological conditions of three-soft coal seams. The development height of the WCFZ in these seams, a particularly challenging stratum, is difficult to predict due to the complex coupling of factors such as mining height, burial depth, lithology, and goaf dimensions. Second, many studies use datasets with broad parameter ranges but limited sample sizes, resulting in models with poor generalization ability. Third, the selection of influencing factors for WCFZ height in most prior work relies primarily on empirical judgment. During model training, an excessive number of input factors can directly impair prediction accuracy. Consequently, existing methodologies provide inadequate guidance for predicting WCFZ height in three-soft coal seam conditions. Therefore, this study aims to address two core challenges: (i) establishing a robust predictive model for WCFZ height under three-soft coal seam conditions, and (ii) systematically determining the key influencing factors governing WCFZ height development.

Accordingly, we collected field-measured WCFZ height data from several mining areas in Henan Province, China. A combined weighting approach, integrating the Analytical Hierarchy Process (AHP) and the Entropy Weight Method (EWM), was employed to systematically identify and quantify the significance of influencing factors. We then constructed a Bayesian-optimized multilayer perceptron (BOA-MLP) model and conducted comprehensive evaluations including uncertainty analysis via Monte Carlo Dropout, sensitivity analysis using SHAP values, and statistical significance testing with Wilcoxon signed-rank test. This framework constitutes the proposed predictive model for WCFZ height in three-soft coal seams. Finally, we applied the model to forecast the WCFZ height for Working Face 15030 at the Yaoling Coal Mine in Gongyi and validated the results through field measurements. This research provides a scientific and accurate approach for predicting WCFZ height in three-soft coal seams.

Methodology

Engineering background

The Yaoling Coal Mine is situated in the eastern part of the Yanlong Coalfield, within the administrative boundaries of Xicun Town and Luzhuang Town, Gongyi City. Borehole data and post-mining geological investigations reveal that the coal-bearing strata belong to the Shanxi Formation of the Permian System. The roof and floor strata of the coal seam predominantly comprise mudstone, sandy mudstone, carbonaceous mudstone, fine-grained sandstone, and siltstone. The roof is characterized by its friability and susceptibility to caving, while floor heave is a notable phenomenon. The underlying L7 limestone, present throughout the mining area, exhibits high strength, a stable stratigraphic position, and a consistent thickness ranging from 10 to 15 m. The focus of this study, Panel 15030, has been fully extracted. It had an actual advance length of 810 m and a face width of 122 m. The mined coal seam 2-1 is a typical “three-soft” coal seam, characterized by low strength of the surrounding rock mass, significant fracture propagation, and pronounced rheological effects. The seam features a simple structure, an inclination of approximately 14°, and an average mining height of about 2.2 m. Extraction employed the longwall mining method along the strike with full-height retreat mining, with roof management using the caving method. A schematic overview of the Yaoling Coal Mine working section is presented in Fig. 1.

Based on the specific geological and mining conditions of the Yaoling mining area, five key influencing factors were preliminarily selected for subsequent analysis and modeling:

(1) Mining height

The mining height determines the vertical extent of the goaf. Under single-seam mining conditions, a greater mining height corresponds to a larger vertical mining disturbance zone, increased stress on the surrounding rock, and consequently, a larger development range of the overburden “three zones” (caved, fractured, and bent zones).

(2) Working face length

The face length governs the horizontal area of the goaf. For near-horizontal coal seams, a longer face length increases the cantilever span of the roof over the goaf, leading to higher pressure from the overlying strata on the unsupported roof. This results in greater deformation and failure of the overburden and an expanded range of the “three zones”.

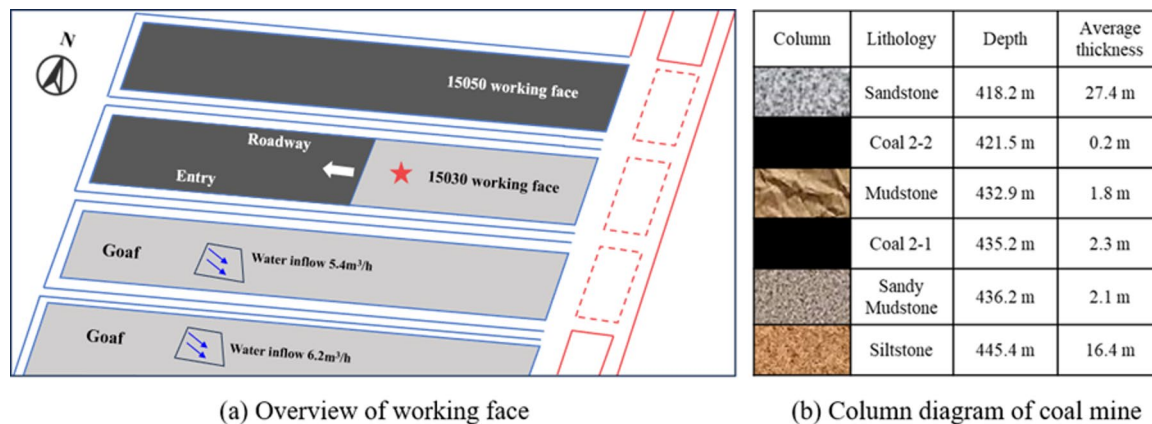


Fig. 1. Overview of the mining area of Yaoling coal mine.

Scale Value	Interpretation
1	A and B are equally important
3	A is slightly more important than B
5	A is strongly more important than B
7	A is very strongly more important than B
9	A is extremely more important than B
2, 4, 6, 8	Intermediate values between adjacent judgments
Reciprocal	If A/B = a, then B/A = 1/a

Table 1. 9-point scale table.

(3) Burial depth

Increased burial depth elevates both the in-situ stress and the original temperature of the coal mass. These factors influence the mechanical properties of individual rock layers and composite strata, thereby affecting their movement behavior in response to mining.

(4) Coal seam dip

The coal seam dip significantly influences the failure range of the overburden “three zones”. The extent of failure exhibits significant variations depending on the dip angle. Within a certain range (excluding steeply inclined seams), a steeper dip angle under identical mining conditions generally leads to a larger failure range in the overburden “three zones”

(5) Lithology

The extent of mining-induced disturbance in the overburden is closely related to its lithology. The mechanical properties of the rock mass determine its deformation characteristics under uniaxial stress. In engineering practice, these mechanical properties serve as crucial parameters for characterizing the quantitative changes in the overburden “three zones” induced by coal seam extraction.

Mathematical methods

Analytical hierarchy process (AHP)

The Analytical Hierarchy Process (AHP) is a pivotal method for transforming semi-qualitative and semi-quantitative problems into quantitative ones²⁶. It structures indicators into interrelated hierarchical levels. Based on expert experience and predefined rules, pairwise comparisons are scored to derive subjective weights for each indicator²⁷. The key computational steps are as follows:

(1) Constructing the judgment matrix

Within each hierarchical level, AHP constructs a judgment matrix by performing pairwise comparisons of the relative importance of all criteria. Each entry in the matrix quantifies the importance of the row element relative to the column element. These comparisons use a nine-level scale (Table 1) to define importance intensity (e.g., comparing Factor A vs. Factor B).

(2) Solving for the Eigenvalue

The maximum eigenvalue of the judgment matrix is calculated using the Root Mean Square method, as expressed in Eq. (1):

$$\lambda_{\max} = \sum_{i=1}^n \frac{(BW)_i}{nW_i} \quad (1)$$

where $(BW)_i$ represents the i -th component of matrix BW .

(3) Consistency check

To confirm the matrix's consistency, the consistency index (CI) and consistency ratio (CR) are calculated using Eq. (2):

$$CI = \frac{\lambda_{\max} - n}{n - 1}$$

$$CR = \frac{CI}{RI} \quad (2)$$

where RI is the random index value from Table 2.

A judgment matrix passes when $CR < 0.1$; otherwise, the pairwise comparisons are adjusted until consistency is achieved.

Entropy weight method (EWM)

The Entropy Weight Method (EWM) is an objective weighting approach that eliminates subjective bias. It quantifies the degree of variation (disorder) within each indicator using entropy. Indicators exhibiting greater variation (higher discrimination power) are assigned larger weights, signifying their stronger influence on the overall weighting calculation²⁸. The computational procedure involves:

- (1) Assemble the original decision matrix A for m schemes and n indicators:

$$A = (a_{ij}) \quad (i = 1, 2, \dots, m, j = 1, 2, \dots, n) \quad (3)$$

- (2) Apply dimensionless normalization to transform A into a standardized matrix $P=[p_{ij}]$:

$$p_{ij} = \frac{a_{ij}}{\sum_{i=1}^m a_{ij}} \quad (j = 1, 2, 3, \dots, n) \quad (4)$$

- (3) Compute the entropy H_j for each indicator j :

$$H_j = -\frac{1}{\ln m} \sum_{i=1}^m p_{ij} \ln p_{ij} \quad (5)$$

- (4) Derive each indicator's weight w_j from its entropy:

$$w_j = \frac{1 - H_j}{n - \sum_{j=1}^n H_j} \quad (6)$$

Multilayer perceptron (MLP)

The Multilayer Perceptron (MLP), also known as a multilayer feedforward neural network, comprises one or more fully connected layers. It is widely applied to regression, classification, and time-series forecasting tasks. The MLP achieves complex function approximation through successive nonlinear transformations across its layers²⁹. Its core principle involves abstracting input features hierarchically via hidden layers, culminating in the output layer performing the final regression or classification task³⁰. Mathematically, an MLP can be expressed as:

$$f(x) = \sigma_L(W^{(L)} \sigma_{L-1}(\dots \sigma_1(W^{(1)}x + b^{(1)}) \dots) + b^{(L)}) \quad (7)$$

Dimension	1	2	3	4	5	6	7
RI	0	0	0.52	0.89	1.12	1.26	1.36

Table 2. RI index stipulated values.

where $x \in R^n$ is the input vector, $W^{(l)}$ is the weight matrix for layer l , $b^{(l)}$ is the bias vector for layer l , $\sigma(\cdot)$ is the nonlinear activation function, and L is the number of layers.

In this study, the Rectified Linear Unit (ReLU) activation function was used for all hidden layers, defined as:

$$\sigma(z) = \max(0, z) \quad (8)$$

For the output layer in our regression task, a linear activation function was employed.

The MLP was selected as the core predictive model in this study primarily for the following reasons: The development of the WCFZ height in three-soft coal seams, with hydrogeological conditions similar to the study area, is a highly complex and nonlinear mechanical process. There exists a profound and non-intuitive mapping relationship between the influencing factors (e.g., mining height, burial depth) and the target value. The MLP, by virtue of its multi-layer structure and nonlinear activation functions, has been proven to be a powerful tool for capturing such complex nonlinear relationships. Compared to ensemble models based on decision trees, such as Random Forests, the MLP typically constructs smoother function approximations when handling continuous input and output variables, which is particularly crucial for regression tasks like fracture zone height prediction. Furthermore, the architecture of the MLP is highly compatible with hyperparameter search algorithms like Bayesian Optimization, facilitating the automated search for an optimal model configuration and thereby fully realizing its performance potential.

Bayesian optimization algorithm (BOA)

The Bayesian Optimization Algorithm (BOA) efficiently guides hyperparameter search by constructing a probabilistic surrogate model of the objective function, balancing exploration (searching new areas) and exploitation (refining known good areas). Its principle leverages Bayes' theorem to estimate the posterior distribution of the objective function. Based on this distribution, the next hyperparameter set to evaluate is selected. BOA leverages information from previous evaluations to progressively refine its understanding of the objective function's shape, ultimately locating the global optimum hyperparameters³¹. Formally, let $X = x_1, x_2, \dots, x_n$ represent a hyperparameter set and $f(x)$ the objective function; BOA seeks X that maximizes (or minimizes) $f(x)$.

BOA typically employs Gaussian Process Regression (GPR), assuming a joint Gaussian distribution over hyperparameters, to model the posterior distribution using the first n evaluated points. This provides an expected mean and variance for each potential x . Selecting points with high mean favors exploitation, while selecting points with high variance promotes exploration. To balance these, an acquisition function quantifies the potential improvement of sampling a point relative to the current best observed value. Common acquisition functions include Upper Confidence Bound (UCB), Probability of Improvement (PoI), and Expected Improvement (EI). This study uses the EI function. The function expression is as follows.

$$EI(x) = \begin{cases} (\mu(x) - f(x^*))\Phi(z) + \sigma(x)\phi(z) & \sigma > 0 \\ 0 & \sigma = 0 \end{cases} \quad (9)$$

$$z = \frac{\mu(x) - f(x^*) - c}{\sigma(x)} \quad (10)$$

where $f(x^*)$ is the current maximum value, μ represents the mean, σ denotes the standard deviation and c serves to balance the exploration-exploitation trade-off.

BOA effectively addresses the challenge of optimizing MLP hyperparameters by establishing this probabilistic surrogate model. The algorithm intelligently searches key MLP parameters (e.g., hidden layer structure, learning rate, batch size) via GPR, converging to the optimal hyperparameter set. BOA inherently avoids local optima, guiding the MLP to find the global optimum configuration with fewer evaluations, significantly enhancing model convergence speed and prediction accuracy. The training workflow of the BOA-MLP model is illustrated in Fig. 2.

To achieve efficient hyperparameter optimization, this study employed BOA with specified key parameters. The optimization process was terminated primarily based on a fixed number of iterations. In each iteration, the surrogate model proposed a new set of candidate hyperparameters for evaluation based on the existing evaluation history. The primary computational cost of the entire process was determined by the total number of iterations and the time required for model training and validation in each iteration.

Wilcoxon signed-rank test

To quantitatively evaluate whether the performance improvements of the BOA-MLP model over baseline approaches are statistically significant, we employed the Wilcoxon signed-rank test. This non-parametric statistical test is suitable for paired comparisons and does not assume a normal distribution of differences.

The Wilcoxon signed-rank test evaluates whether the median difference between paired observations is zero. The test procedure involves the following steps:

- (1) Calculate the differences between paired observations:

$$d_i = y_{1i} - y_{2i} \quad (11)$$

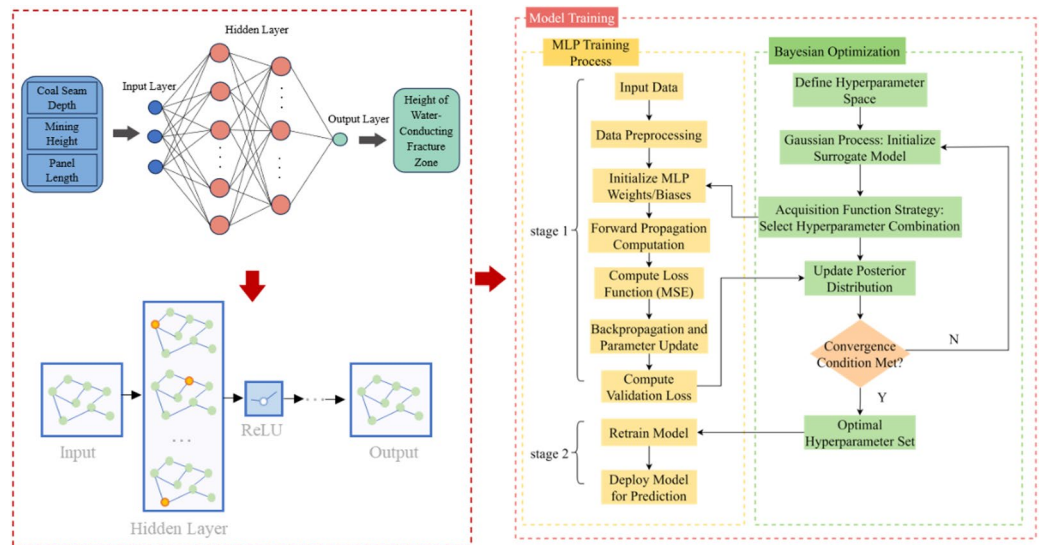


Fig. 2. BOA-MLP training workflow.

where y_{1i} and y_{2i} represent the absolute prediction errors of BOA-MLP and the baseline model for the i -th test sample, respectively.

- (2) Rank the absolute differences $|d_i|$ in ascending order, ignoring their signs.
- (3) Assign ranks R_i to the absolute differences, with the smallest absolute difference receiving rank 1.
- (4) Calculate the sum of ranks for positive differences (W^+) and negative differences (W^-):

$$W^+ = \sum_{i=1}^n [d_i > 0] \cdot R_i \tag{12}$$

$$W^- = \sum_{i=1}^n [d_i < 0] \cdot R_i \tag{13}$$

where $[\cdot]$ is the Iverson bracket that equals 1 if the condition is true and 0 otherwise.

- (5) The test statistic W is the smaller of W^+ and W^- :

$$W = \min(W^+, W^-) \tag{14}$$

- (6) For small sample sizes, the exact distribution of W is used to determine the p-value. The test was conducted at a significance level of $\alpha = 0.05$. Additionally, we calculated effect sizes using Cohen's d to quantify the magnitude of performance differences:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s_{pooled}} \tag{15}$$

where \bar{X}_1 and \bar{X}_2 are the sample means, and s_{pooled} is the pooled standard deviation.

Borehole imaging

To analyze WCFZ height in mined-out panels and predict it for unmined panels within the Yaoling mining area, a borehole inspection device was employed for field measurement. Combined with a mining borehole imaging trajectory detection system, this enables precise detection of WCFZ height. The survey procedure, borehole layout, and equipment are depicted in Fig. 3.

- (1) Borehole layout and parameter design

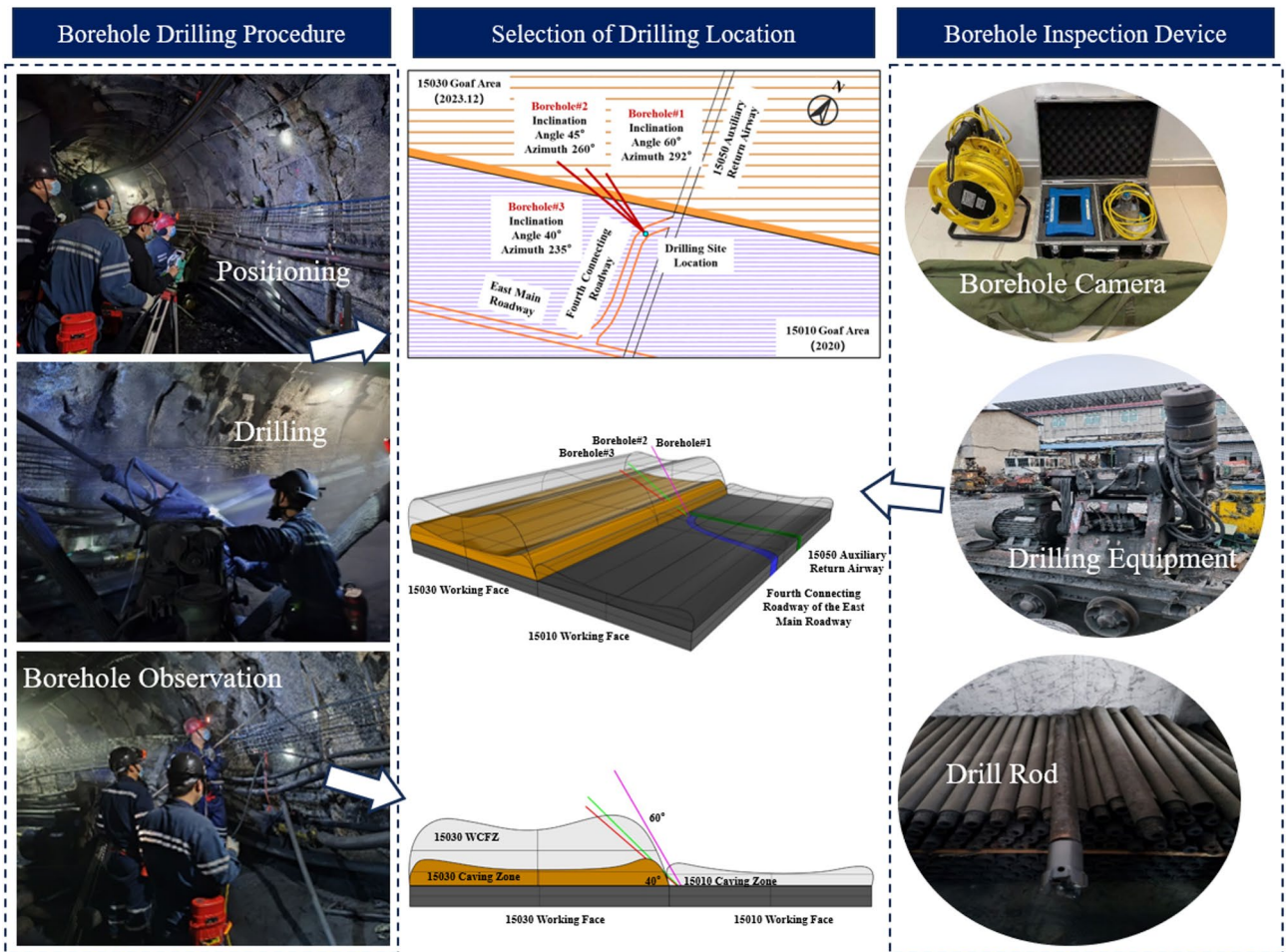


Fig. 3. Borehole imaging overall process.

Based on the mining area's specifics, three boreholes were strategically positioned at the corner of the East Main Road. Borehole #1, drilled at an inclination of 60° and an azimuth of 292° , served as the control borehole. Borehole #2 (inclination: 45° , azimuth: 260°) and Borehole #3 (inclination: 40° , azimuth: 235°) were designed to traverse the flanks of the roof saddle zone and intersect the boundary of the roof fracture zone.

(2) Drilling and imaging equipment

ZL 1700 hydraulic drill with 50 mm drill rods and bits for advancing the holes. ZKXG100 forward view borehole imager, whose probe houses LED illumination, an HD camera, and a 3D electronic compass, producing 360° unfolded borehole-wall images and spatial positioning data.

(3) Borehole drilling procedure

Using a total station to measure control point coordinates and azimuths within the roadway, followed by staking out the exact borehole collar locations at the drill site. Using the ZL-1700 rig and a 50 mm bit to drill to the specified inclination and azimuth. Boreholes were flushed to remove cuttings. Slowly lowering the ZKXG100 probe into the borehole while continuously acquiring and stitching 360° wall images. Depth, azimuth, and inclination data were recorded concurrently for subsequent WCFZ development analysis.

Results and discussion

Comprehensive weighting of influencing factors

The judgment matrix, calculated according to the steps in [Analytical Hierarchy Process \(AHP\)](#) section (1), is presented in Table 3.

For the consistency check, with five indicators selected, $RI = 1.12$. MATLAB calculations yielded a CR value of 0.037, satisfying the consistency criterion ($CR < 0.1$). The resulting subjective weights from AHP are listed in Table 4.

The entropy weights, calculated following the procedure in Sect. [Entropy Weight Method \(EWM\)](#), are shown in Table 5.

Criterion layer	Lithology	Coal seam dip	Working face length	Mining height	Burial depth
Lithology	1	1	1/3	1/5	1/7
Coal seam dip	1	1	1/3	1/3	1/7
Working face length	3	3	1	1/3	1/5
Mining height	5	3	3	1	1
Burial depth	7	7	5	1	1

Table 3. Judgment matrix.

Influencing factor	Lithology	Coal seam dip	Working face length	Mining height	Burial Depth
Weights	0.0559	0.0639	0.1322	0.3128	0.4532

Table 4. AHP-derived indicator weight results.

Influencing factor	Lithology	Coal seam dip	Working face length	Mining height	Burial depth
Weights	0.1356	0.1683	0.2644	0.2265	0.2032

Table 5. EWM-derived indicator weight results.

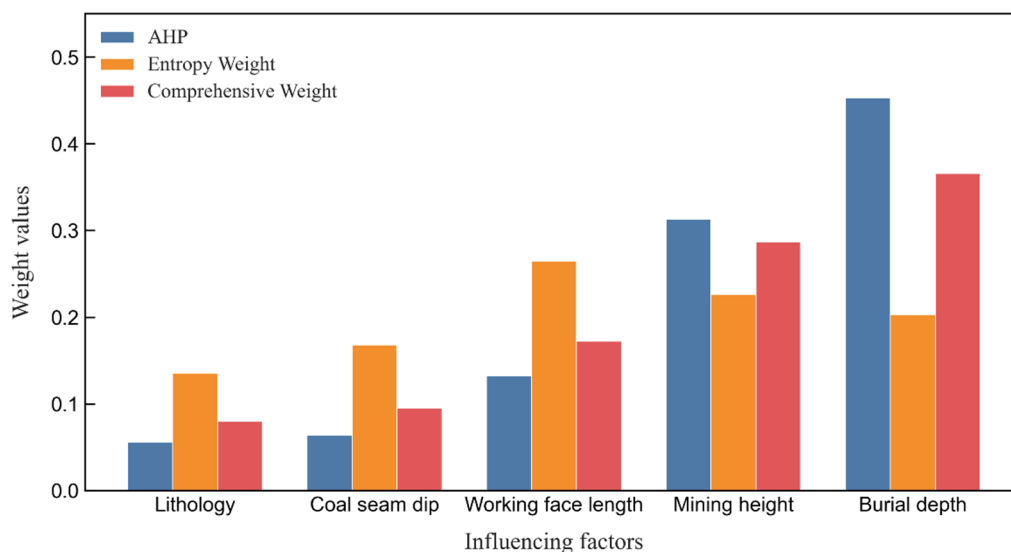


Fig. 4. Weight results derived from different methods.

The subjective weights w_j^1 reflect the expert-based ranking of indicator importance, while the entropy weights w_j^2 represent the objective statistical relationships revealed by the data. The integrated weight for the j -th indicator was calculated by coupling w_j^1 and w_j^2 using Eq. (11):

$$w_j^* = (w_j^1)^{1-\alpha} (w_j^2)^\alpha \tag{16}$$

where $\alpha=0.3$ denotes the relative importance proportion assigned to the subjective and objective weights.

The coupled weighting results for all indicators are visualized in Fig. 4. Based on the integrated weighting analysis, mining height, face length, and burial depth were ultimately selected as the primary influencing factors.

Development of the BOA-MLP neural network prediction model

Dataset establishment and processing

To accurately predict the WCFZ height for Panel 15030 at the Yaoling Coal Mine, field-measured WCFZ heights and mining parameters from mines in Henan Province with similar Carboniferous-Permian geological conditions were statistically compiled, as shown in Fig. 5.

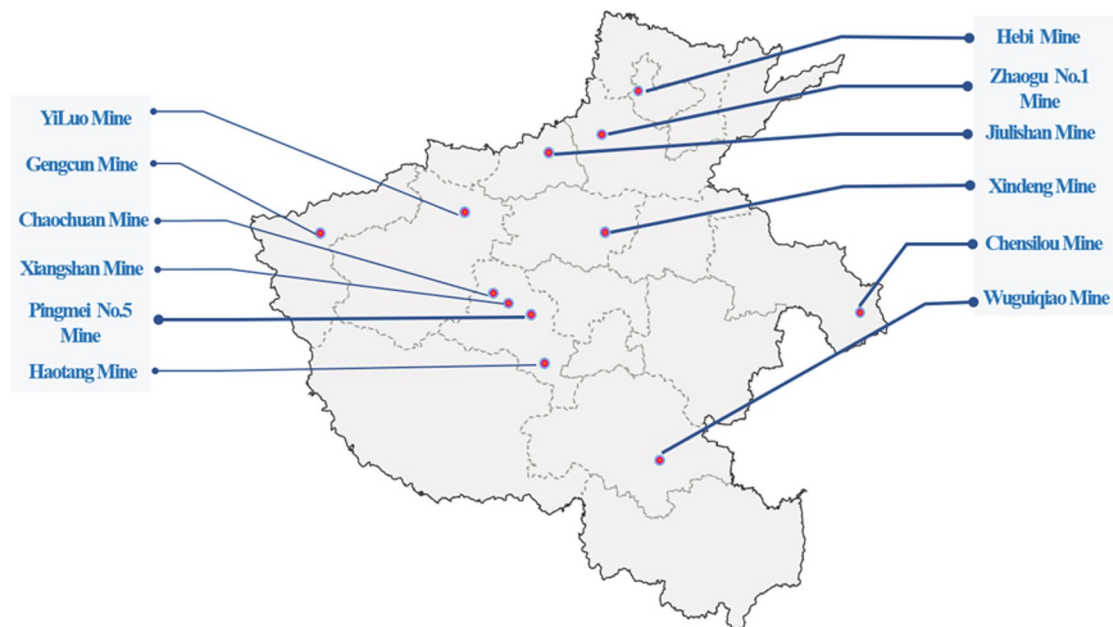


Fig. 5. Distribution of mines with similar engineering conditions in Henan Province.

Through literature review, 32 sets of mining parameters and measured WCFZ heights from working faces with engineering geological characteristics similar to Panel 15030 were collected. To eliminate dimensional heterogeneity while retaining distributional characteristics, all variables were standardized using Z-score normalization:

$$X_{\text{standardized}} = \frac{X - \mu}{\sigma} \quad (17)$$

where X is the original data value, μ is the mean of the data column, and σ is its standard deviation.

The collected and standardized dataset is presented in Table 6.

Development of the BOA-MLP neural network prediction model

The model training in this study was conducted in the Python 3.8 environment, using the PyTorch 1.13 framework for constructing the MLP network, the scikit-learn 1.2 library for data preprocessing, and the BayesianOptimization 1.4 library for implementing the Bayesian optimization algorithm. The dataset was partitioned using the `train_test_split` function from scikit-learn: 70% for training, 15% for validation, and 15% for testing, with a random seed of 42 to ensure reproducibility. This partitioning scheme, while working with the limited sample size ($n = 32$), ensures adequate training data while maintaining reliable validation and test sets for hyperparameter tuning and final evaluation, aligning with standard practices in small-sample studies.

An MLP neural network with two hidden layers was constructed. The input feature dimension was 3, and the output dimension was 1. The Rectified Linear Unit (ReLU) activation function was used for the hidden layers, the loss function was Mean Squared Error (MSE), and the Adam optimizer was employed. The hyperparameter search spaces for the Bayesian Optimization are detailed in Table 7.

The model was trained for 100 epochs during each BOA iteration. The Bayesian optimization used a Gaussian Process with a Matérn 5/2 kernel as the surrogate model. The process was initialized with 5 exploration points and run for 15 optimization iterations. This iteration count was set as the stopping criterion, providing a good balance between computational cost and performance gain.

Figure 6a–d clearly illustrate the evolution of each hyperparameter throughout the BOA iterations. It can be observed that the Bayesian optimization algorithm does not perform a random search but intelligently and directionally adjusts the hyperparameters. After the initial exploration, the values of each hyperparameter quickly aggregated into several regions with better performance. This indicates that BOA effectively learned the relationship between these parameters and model performance and identified the potential ranges for the optimal solution. Figure 6e shows the evolution of the best validation loss during the Bayesian optimization. It can be observed that the curve exhibits certain volatility throughout the iterations, rather than a smooth, monotonic descent. We attribute this primarily to the limited sample size ($n = 32$) in this study. On small datasets, the validation performance of neural network models is more sensitive to stochastic factors during training (e.g., weight initialization and mini-batch sampling), leading to inherent fluctuations in the validation loss. This ‘noise’ poses a challenge for Bayesian optimization in building a precise surrogate model. Nevertheless, BOA ultimately succeeded in identifying a hyperparameter configuration that outperformed manual tuning (see Table 8), demonstrating its practical utility even in challenging small-data scenarios. Figure 6f illustrates the relationship between hyperparameters and validation loss, where symbol size is proportional to batch size. The lowest

No.	Mine name	Measured values				Standardized values			
		Depth/m	Height/m	Length/m	WCFZ/m	Depth	Height	Length	WCFZ高
1	Xiangshan Mine	636	7.2	159	54.3	1.17	1.95	-0.83	0.73
2		578	3.7	271	18.5	0.62	-0.35	0.88	-0.86
3	Pingmei No.5 Mine	584	4.3	293	30.2	0.68	0.04	1.21	-0.34
4		752	6.3	288	77.1	2.27	1.36	1.14	1.74
5		587	2.9	242	18.7	-0.71	-0.88	0.44	-0.85
6	Xindeng Mine	494	3.0	148	21.5	-0.17	-0.81	-0.99	-0.73
7		315	6.0	190	52.6	-1.87	1.16	-0.35	0.65
8		301	6.2	127	57.8	-2	1.29	-1.31	0.88
9	Haotang Mine	647	4.1	207	32.6	1.28	-0.09	-0.09	-0.23
10		656	6.7	133	67.8	1.36	1.62	-1.22	1.33
11	Chensilou Mine	556	3.1	101	24.9	0.42	-0.75	-1.71	-0.58
12		490	2.4	278	18.1	-0.21	-1.21	0.99	-0.88
13	Yiluo Mine	434	3.8	238	32.2	-0.74	-0.29	0.38	-0.25
14	Zhaogu Mine	565	2.0	263	12.7	0.5	-1.47	0.76	-1.12
15		488	4.2	102	30.5	-0.23	-0.02	-1.69	-0.33
16		690	6.4	264	72.1	1.68	1.42	0.77	1.52
17		548	3.6	178	25.1	0.34	-0.42	-0.54	-0.57
18	Gengcun Mine	430	3.1	231	24.1	-0.78	-0.75	0.27	-0.61
19	Yangcun Mine	479	3.9	146	40.6	-0.31	-0.22	-1.02	0.12
20	Chaochuan Mine	489	3.5	154	24.7	-0.22	-0.48	-0.9	-0.59
21		477	5.5	240	81.5	-0.33	0.83	0.41	1.94
22	Hebi No.6 Mine	483	7.1	129	91.5	-0.28	1.88	-1.28	2.38
23	Chengjiao Mine	479	4.8	223	49.3	-0.31	0.37	0.15	0.51
24	Songshan Mine	463	2.3	289	14.4	-0.46	-1.27	1.15	-1.04
25	Quandian Mine	766	2.1	285	15.8	2.4	-1.4	1.09	-0.98
26		434	2.8	200	25.6	-0.74	-0.94	-0.2	-0.55
27	Zhaogu No.1 Mine	573	3.6	230	31.1	0.58	-0.42	0.26	-0.3
28	Yongsheng Mine	370	7.1	282	51.1	-1.35	1.88	1.05	0.59
29	Pingmei No.5 Mine	732	5.0	220	47.8	2.08	0.5	0.1	0.44
30	Pingmei No.6 Mine	769	2.1	270	18.0	2.43	-1.4	0.86	-0.88
31		569	6.4	146	65.8	0.54	1.42	-1.02	1.24
32		463	2.7	297	13.9	-0.46	-1.01	1.27	-1.06

Table 6. WCFZ height data and standardized values.

Hyperparameters	Search space
Hidden layer 1 neurons	32–128
Hidden layer 2 neurons	16–64
Learning rate	e^{-4} – e^{-3}
Batch size	16–256

Table 7. Search space of hyperparameters.

validation loss points (darkest color) predominantly occur within medium-sized networks (Hidden1 \approx 80–100 neurons, Hidden2 \approx 25–50 neurons), small-to-medium batch sizes (smaller dots), and medium-to-low learning rates ($\approx 3 \times 10^{-4}$ – 6×10^{-4}). Larger networks (Hidden1 > 120 neurons or Hidden2 > 60 neurons) and very high learning rates ($> 8 \times 10^{-4}$) resulted in higher validation loss, indicating poorer generalization. Figure 7 provides an in-depth analysis of the distribution characteristics of the optimal hyperparameters through marginal histograms and pairwise scatter and density plots. The histograms on the diagonal indicate that the number of neurons in the first hidden layer is concentrated between 40 and 120, the second hidden layer between 10 and 80, the learning rate between 0.0001 and 0.0014, and the batch size is clustered around the four values of 32, 64, 128, and 256. The scatter plots in the upper triangle show that the configurations with the lowest validation loss are mostly concentrated in the ranges of Hidden1 \approx 60–100, Hidden2 \approx 30–60, LR \approx 0.0005–0.0010, and Batch Size \approx 32–128. The kernel density plots in the lower triangle further corroborate that high-performance hyperparameter combinations frequently occur within these aforementioned ranges. It can be inferred that employing a medium to large number of hidden units, small-to-medium learning rates, and medium batch sizes

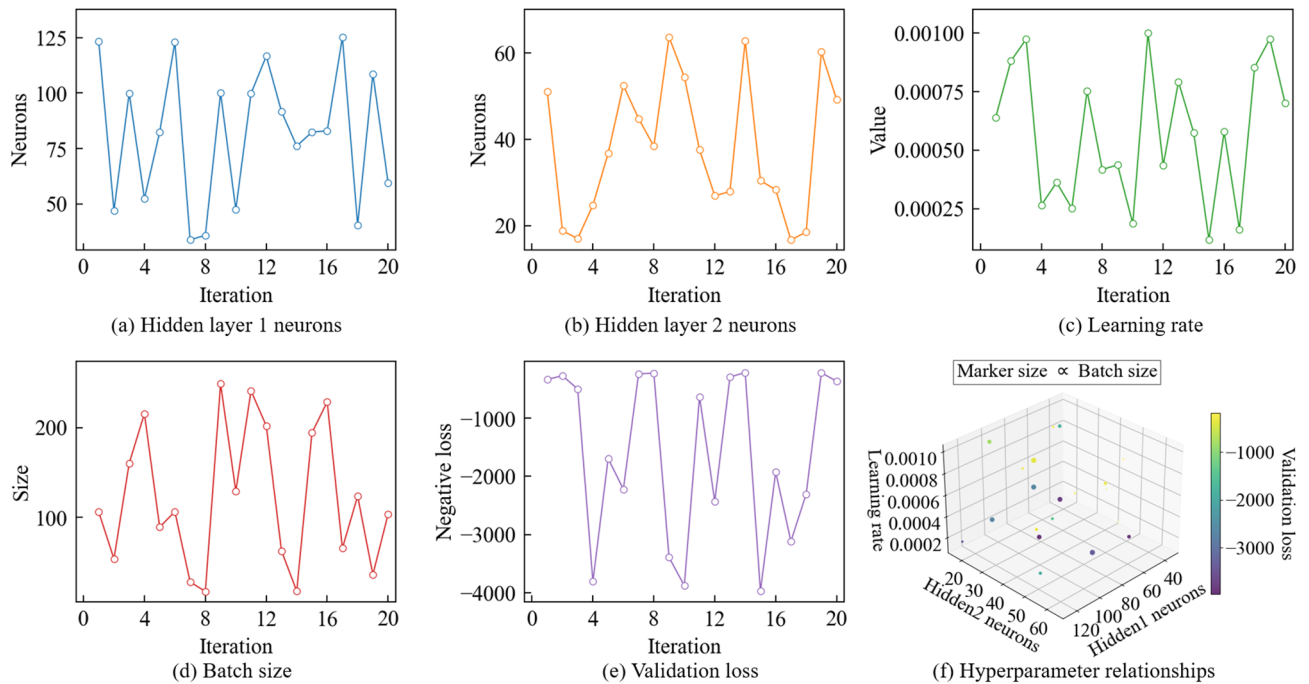


Fig. 6. Bayesian optimization progress analysis.

Hyperparameters	Value
Hidden layer 1 neurons	64
Hidden layer 2 neurons	32
Learning rate	0.001
Batch size	32

Table 8. Bayesian optimization-derived optimal parameters.

is a preferable choice for the current model. The best hyperparameter combination identified by BOA is listed in Table 8.

Using this optimal hyperparameter set, the final MLP model was trained for 400 epochs. To benchmark performance, a manually tuned MLP model (using expert-selected hyperparameters) was also trained. To comprehensively evaluate the model performance, we employed multiple statistical metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Percentage Peak Deviation (PPD), Coefficient of Determination (R^2), Nash–Sutcliffe Efficiency (NSE), and Root Mean Square Error to Standard Deviation Ratio (RSR). The formulas for these metrics are provided in Table 9.

These metrics comprehensively evaluate the model’s performance from multiple dimensions including error magnitude, percentage deviation, goodness-of-fit, and predictive efficiency. It is noteworthy that within the single-output regression framework of this study, the Coefficient of Determination (R^2) and the Nash-Sutcliffe Efficiency (NSE) are closely related in value, both serving to quantify the agreement between the model’s predictions and the measured values. Figure 8a illustrates the training and validation loss curves of both BOA-MLP and manual MLP models. The BOA-MLP converged faster and achieved a lower final loss. Figure 8b presents the learning curves, which depict the model performance on the training and validation sets as a function of the number of training samples. The BOA-MLP demonstrates superior generalization ability, as evidenced by the closer convergence of training and validation accuracy with increasing sample size. Figure 9a compares the performance metrics of both models on the test set. The BOA-MLP achieved superior results across all evaluated metrics: significantly lower error metrics (RMSE = 1.98 m, MAE = 2.23 m, MAPE = 2.67%) compared to the manual MLP (RMSE = 2.77 m, MAE = 2.98 m, MAPE = 3.14%); higher goodness-of-fit and efficiency coefficients ($R^2 = 0.973$, NSE = 0.973) versus the manual MLP ($R^2 = 0.914$, NSE = 0.914); and a lower peak deviation and error ratio (PPD = 0.982%, RSR = 0.180) than the manual MLP (PPD = 1.330%, RSR = 0.260). This comprehensive quantitative comparison unequivocally demonstrates the superior predictive accuracy, robustness, and generalization capability of the BOA-MLP model. Figure 9b visually compares the predicted values of the BOA-MLP and the manual MLP against the true values for the 10 samples in the test set. It can be clearly observed that the predicted values from the BOA-MLP align more closely with the true values, exhibiting smaller fluctuations and maintaining stable prediction accuracy across multiple consecutive samples. In contrast, the predictions from the manual MLP show significant deviations from the true values at several data points,

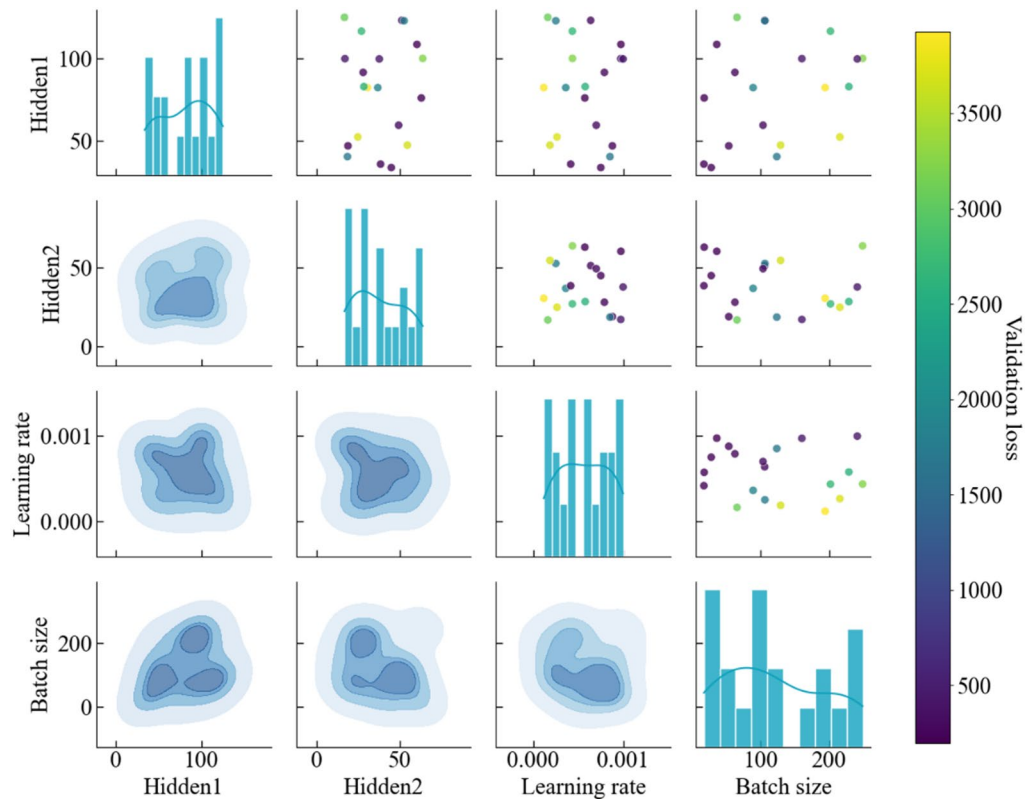


Fig. 7. Hyperparameter relationships matrix.

Metrics	Formulas
Root mean square error (RMSE)	$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$
Mean absolute error (MAE)	$MAE = \frac{\sum_{i=1}^n \hat{y}_i - y_i }{n}$
Mean absolute percentage error (MAPE)	$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left \frac{\hat{y}_i - y_i}{y_i} \right $
Percentage peak deviation (PPD)	$PPD = \frac{ \max \hat{y}_i - \max y_i }{\max y_i} \times 100\%$
Coefficient of determination (R ²)	$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$
Nash-Sutcliffe efficiency (NSE)	$NSE = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$
Root Mean Square Error to Standard Deviation Ratio (RSR)	$RSR = \frac{\sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$

Table 9. Model performance metrics and formulas.

indicating higher instability. This visually confirms the superior predictive accuracy and robustness of the BOA-MLP model.

To quantitatively validate that the performance enhancements of the BOA-MLP model over the manually tuned MLP are statistically significant, we conducted Wilcoxon signed-rank tests on the absolute prediction errors of both models across the test set. The null hypothesis stated that the median difference between the absolute errors of the two models is zero, while the alternative hypothesis posited that the BOA-MLP produces significantly lower errors.

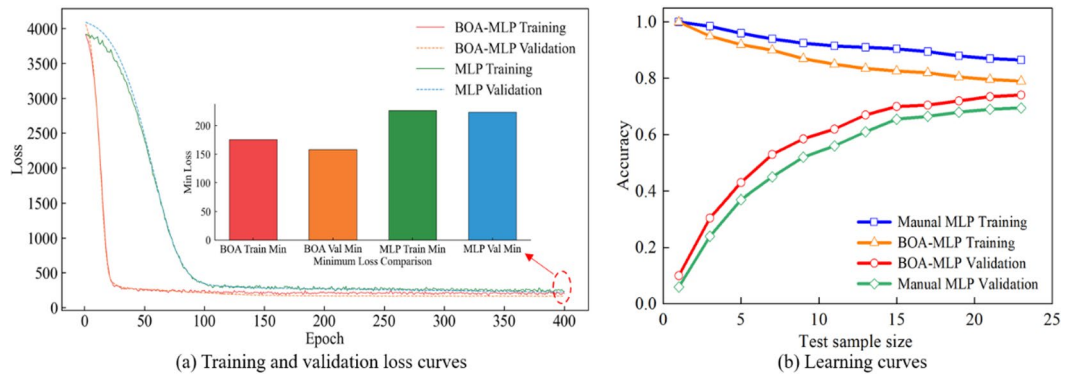


Fig. 8. Training dynamics and generalization performance of the BOA-MLP and manual MLP models.

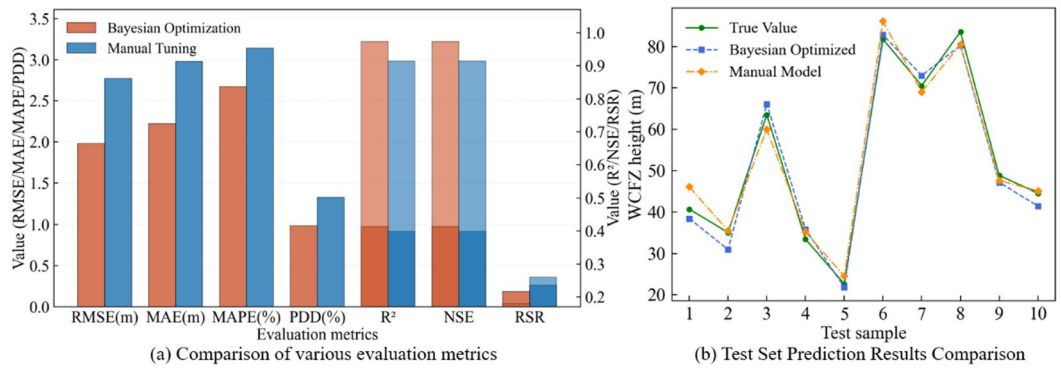


Fig. 9. Comprehensive evaluation and predictive effectiveness of the models on the test set.

Source (Reference)	Model used	R ²	RMSE / m
Zhao et al. ¹⁷	RFR	0.968	2.64
Wu et al. ¹⁸	XGBRegressor	0.925	1.90
Wang et al. ²⁰	SSA-RF	0.958	5.61
Xun et al. ²²	APSO-LSSVR	0.948	1.65
Xu et al. ³²	CNN	0.970	2.60
This study	BOA-MLP	0.973	1.98
This study	Manual MLP	0.914	2.77

Table 10. Performance comparison of different models for WCFZ height prediction.

Following the computational procedure outlined in [Wilcoxon Signed-Rank Test](#) section, the statistical analysis yielded a *p*-value of 0.027, which is substantially below the conventional significance threshold of 0.05. This result provides strong evidence to reject the null hypothesis, confirming that the BOA-MLP model achieves a statistically significant improvement in prediction accuracy compared to the manually tuned MLP.

Furthermore, we calculated Cohen's *d* effect size according to Eq. 15 to quantify the magnitude of this improvement. The obtained effect size of 0.83 represents a large effect according to conventional benchmarks (where *d*=0.2 indicates a small effect, *d*=0.5 a medium effect, and *d*=0.8 a large effect). This substantial effect size underscores the practical significance of the performance enhancement achieved through Bayesian optimization.

Comparative analysis of model performance

To provide a more comprehensive evaluation of the proposed BOA-MLP model, a quantitative comparison with other advanced prediction models reported in recent literature is summarized in Table 10.

As clearly demonstrated in Table 10, the proposed BOA-MLP model achieves the highest R² value (0.973) among all the compared models, indicating its superior ability to explain the variance in the WCFZ height data. While the APSO-LSSVR model reports a slightly lower RMSE (1.65 m), its R² value (0.948) is notably lower than that of our model. Conversely, the RFR and CNN models achieve comparable R² values (0.968 and 0.970) but

exhibit higher RMSE values (2.64 m and 2.60 m). This positions the BOA-MLP model as offering an excellent balance between high explanatory power and low prediction error. Furthermore, the significant performance gap between the BOA-MLP and the Manual MLP within this study (R^2 : 0.973 vs. 0.914; RMSE: 1.98 m vs. 2.77 m) quantitatively underscores the critical contribution of the Bayesian Optimization process in enhancing model performance. The results collectively confirm that our BOA-MLP framework is highly competitive and represents a state-of-the-art approach for predicting WCFZ height.

Uncertainty and sensitivity analysis

To comprehensively evaluate the robustness and interpretability of the proposed BOA-MLP model within the constraints of our limited dataset ($n=32$), we conducted integrated uncertainty and sensitivity analyses. These analyses were specifically designed to maximize data utilization while providing crucial insights into prediction reliability and feature importance.

Given the limited sample size, we employed a cross-validation strategy combined with Monte Carlo Dropout to quantify prediction uncertainty. This approach ensures that all 32 samples in our dataset are used for both training and evaluation, thereby providing a comprehensive assessment of model performance across the entire data distribution. The Monte Carlo Dropout technique performs multiple stochastic forward passes during inference, enabling the estimation of prediction intervals by assessing the variability in model outputs under different dropout configurations. Our cross-validation framework involved iterative training and testing cycles where each sample served as a test instance in at least one fold, ensuring complete coverage of the dataset. As illustrated in Fig. 10, the model demonstrates well-calibrated uncertainty quantification across all 32 samples. The analysis reveals that 27 samples (84.4%) fall within the 95% confidence intervals, while only 5 samples (15.6%) exhibit minor deviations beyond these bounds. This coverage closely approximates the nominal confidence level, indicating reliable uncertainty estimation despite the limited dataset size.

To elucidate the relative influence of input parameters on model predictions using our complete dataset, we employed SHapley Additive exPlanations (SHAP). This game-theoretic approach provides consistent and locally accurate feature importance measurements while being particularly suitable for small datasets due to its robust mathematical foundation. The utilization of all 32 samples for SHAP analysis ensures that the feature importance rankings are derived from the complete data distribution, minimizing sampling bias and providing more reliable insights into parameter sensitivities. This comprehensive approach is especially valuable given our limited sample size, as it maximizes the informational content extracted from each data point. The SHAP summary plot (Fig. 11) distinctly reveals the hierarchical importance of the three input features based on the complete dataset analysis. Burial depth emerges as the predominant factor, exhibiting the highest mean absolute SHAP value, followed by mining height and working face length. This feature importance ranking demonstrates remarkable consistency with the integrated AHP-EWM weighting results presented in [Comprehensive Weighting of Influencing Factors](#) section. The directional effects are clearly discernible from the feature value distributions across all 32 samples: elevated burial depth values (indicated by red markers) predominantly associate with positive SHAP values, signifying increased WCFZ height predictions. Conversely, reduced burial depth values (blue markers) generally correspond to diminished predictions. Analogous patterns, though with attenuated magnitudes, are observed for mining height and working face length parameters.

Discussion on generalization and overfitting risks

This study acknowledges that the dataset of 32 samples is relatively limited for neural network training, which carries a potential risk of overfitting. To proactively mitigate this issue and enhance the model's generalization capability, we implemented several key strategies, the effectiveness of which is now substantiated by rigorous quantitative analyses.

Firstly, the integrated AHP-EWM weighting method was employed to select only three primary influencing factors, thereby reducing the input space dimensionality and model complexity from the outset. Secondly,

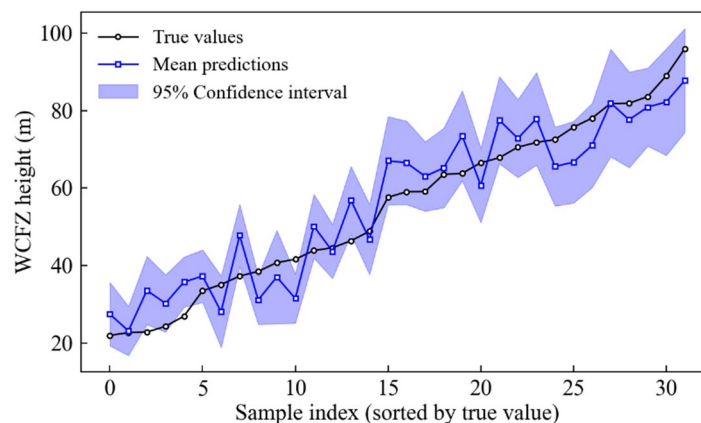


Fig. 10. Uncertainty quantification of BOA-MLP predictions using Monte Carlo dropout with cross-validation across all 32 samples.

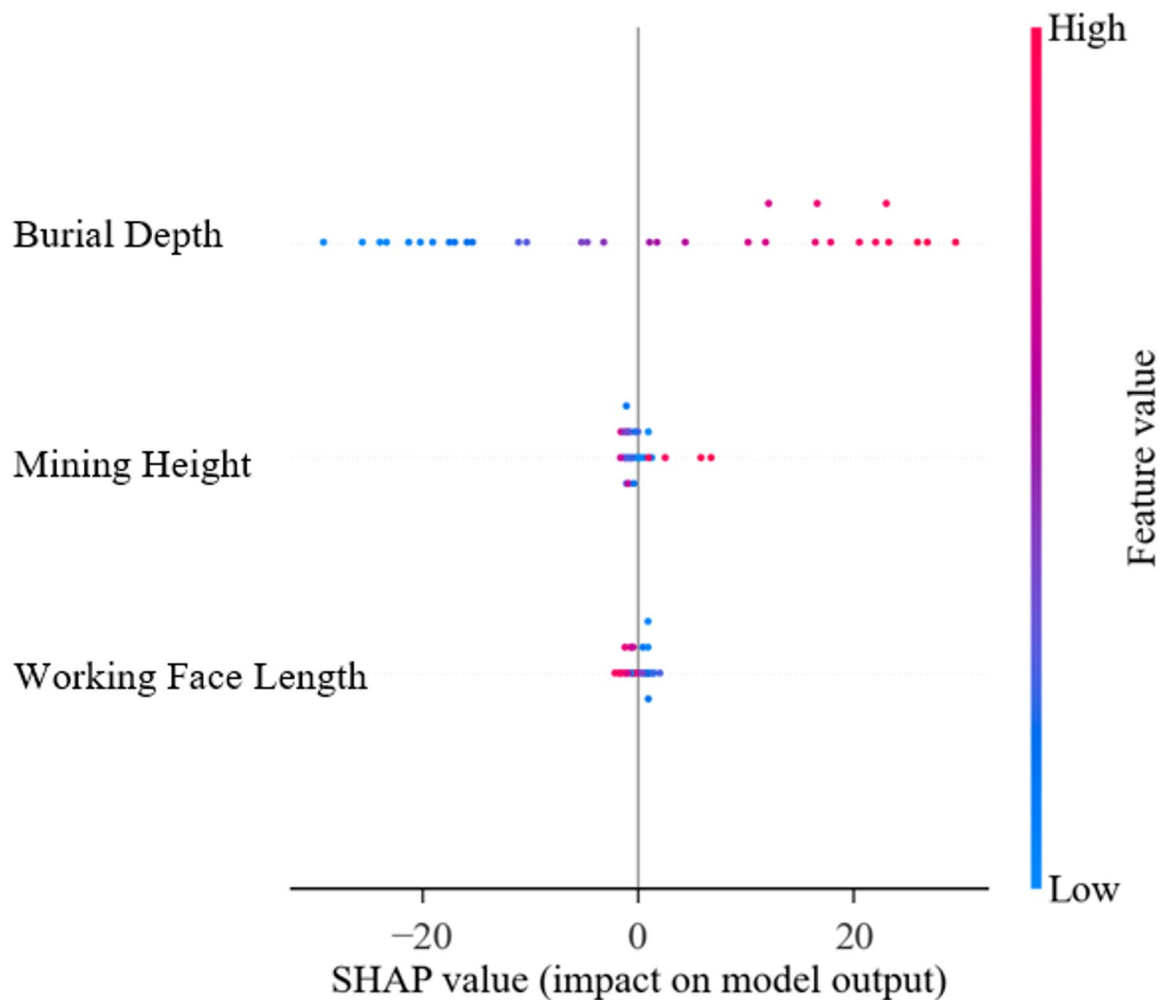


Fig. 11. SHAP-based sensitivity analysis of input features on BOA-MLP predictions using the complete dataset ($n = 32$).

the Bayesian Optimization process itself acted as a powerful regularizer by identifying hyperparameter sets that performed best on the validation set, inherently guiding the search towards configurations with stronger generalization power.

Critically, the results of our comprehensive post-hoc analyses provide direct evidence that overfitting was effectively controlled:

- (1) The statistical significance testing (Wilcoxon signed-rank test, $p = 0.027$) confirmed that the performance improvement of BOA-MLP over the baseline model is genuine and not attributable to overfitting to the training set or random chance.
- (2) The Monte Carlo Dropout uncertainty analysis ([Uncertainty and Sensitivity Analysis](#) section) demonstrated well-calibrated predictions, with 84.4% of test samples falling within the 95% confidence intervals. This indicates that the model recognizes the limits of its knowledge and does not exhibit overconfident predictions on unseen data, a key marker of good generalization.
- (3) The SHAP sensitivity analysis ([Uncertainty and Sensitivity Analysis](#) section) yielded a feature importance ranking (burial depth > mining height > working face length) that is consistent with the quantitative results from our integrated AHP-EWM weighting analysis ([Comprehensive Weighting of Influencing Factors](#) section). This cross-methodological validation strongly indicates that the model has learned physically plausible relationships rather than spurious correlations.

The combination of converging training and validation learning curves (Fig. 8b), excellent performance on the independent test set (Fig. 9), and the aforementioned quantitative evidence collectively demonstrates that the BOA-MLP model achieved robust generalization despite the limited sample size.

BOA-MLP model prediction

To further validate the model's accuracy, the WCFZ height for Panel 15030 at the Yaoling Coal Mine was predicted using three methods: the empirical formula from Ref⁵, the manually tuned MLP, and the BOA-MLP

Mining method	Mining height/m	Working face length/m	Burial depth/m
Top-coal caving	2.41	240	122

Table 11. Key geological parameters of working face 15,030.

Empirical formula prediction/m	MLP prediction / m	BOA-MLP prediction/m
22.3	29.3	28.1

Table 12. Prediction results of different models.

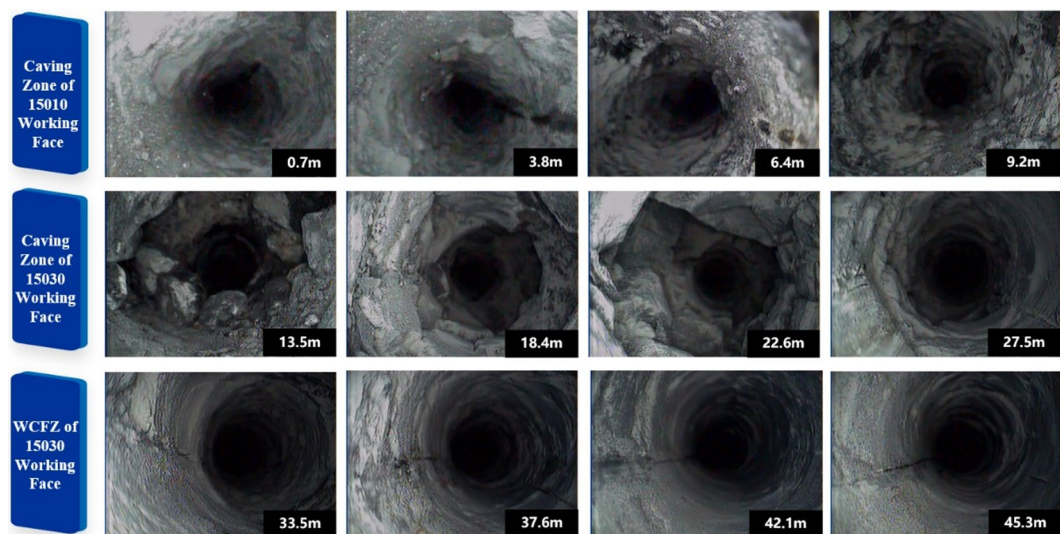


Fig. 12. Borehole #3 imaging zonation map.

model. The key geological parameters of the panel are listed in Table 11, and the final prediction results are presented in Table 12.

Development of the BOA-MLP neural network prediction model

Based on fracture development observed in Borehole #3, the profile was divided into three distinct zones: the caving zone of Working Face 15010, the caving zone of Working Face 15030, and the WCFZ immediately above Working Face 15030. Figure 12 illustrates these zones.

From the logged images, the fracture frequency was extracted and plotted versus depth in Fig. 13. Between 0.7 m and 9.2 m, fracture rates of 28–32% reflect compaction of the goaf. From 15.4 m to 22.6 m, the fracture rate peaks at 40%. It then slightly decreases to ~38% between 22.6 m and 27.5 m. The rock mass structure is severely damaged, exhibiting a complex network of interconnected fractures of varying sizes and orientations, characteristic of the caved zone. Beyond 27.5 m to 45.3 m, vertically dominant fractures with lateral extensions characterize the active water conducting fracture zone. Based on the borehole #3 imaging results, the developed height of the WCFZ above 15030 Working Face is determined to be approximately 27.3 m.

Comparison with field data confirms that BOA-MLP delivers the closest match, underscoring the efficacy of Bayesian optimization in enhancing MLP accuracy and stability.

Conclusions

- (1) Using an integrated subjective-objective weighting approach combining the AHP and EWM, the weights for the five primary influencing factors were calculated as follows: lithology (0.0798), coal seam dip (0.0952), working face length (0.1724), mining height (0.2869), and burial depth (0.3657). Based on this analysis, mining height, working face length, and burial depth were definitively selected as the final influencing factors for model development.
- (2) The BOA-MLP model for predicting WCFZ height achieved an RMSE of 1.98 m, MAE of 2.23 m, MAPE of 2.67%, PPD of 0.982%, R^2 of 0.973, NSE of 0.973 and RSR of 0.180. The model outperformed the manually tuned MLP model across all seven metrics, demonstrating its superior accuracy (lower error values) and enhanced generalization capability. This confirms its applicability for predicting WCFZ height within the study region.

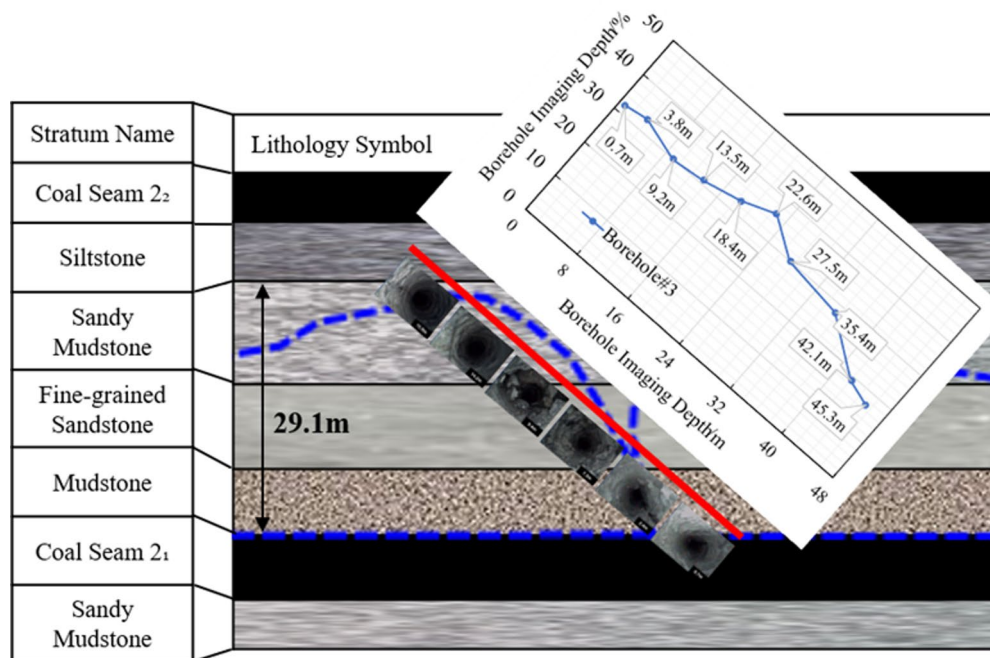


Fig. 13. Borehole #3 imaging rock strata fracture rate evolution schematic.

- (3) The integrated uncertainty and sensitivity analyses provided comprehensive validation of the BOA-MLP model. Monte Carlo Dropout uncertainty quantification demonstrated well-calibrated prediction intervals with 84.4% coverage of the 95% confidence bounds across test samples. SHAP-based sensitivity analysis established burial depth as the most influential parameter, followed by mining height and working face length, exhibiting perfect concordance with the prior integrated weighting methodology.
- (4) In situ borehole imaging at Yaoling Coal Mine determined a WCFZ height of approximately 27.3 m for Working Face 15,030. The BOA-MLP model predicted a height of 28.1 m, showing closer agreement with field measurements than predictions derived from empirical formulas or the manually tuned MLP model. Consequently, the BOA-MLP model provides practical guidance for WCFZ height estimation and mine water hazard prevention in geologically similar areas.

Data availability

The dataset supporting the findings of this study is available from the corresponding author upon reasonable request.

Received: 15 September 2025; Accepted: 10 December 2025

Published online: 03 January 2026

References

- Guo, C. F., Yang, Z., Li, S. & Lou, J. F. Predicting the Water-Conducting fracture zone (WCFZ) height using an MPGA-SVR approach. *Sustainability* **12** (5), 1809 (2020).
- Zhang, L. F., Zhang, Z. Z., Wang, K. K. & Tan, X. D. Characteristic developments of the water-conducting fracture zones in weakly cemented overlying strata of jurassic coal mines in Western China. *Water* **15** (6), 1097 (2023).
- Lu, C. J., Xu, J. P., Li, Q., Zhao, H. & He, Y. Research on the development law of water-conducting fracture zone in the combined mining of jurassic and carboniferous coal seams. *Appl. Sci.* **12** (21), 11178 (2022).
- Huang, Y. L. et al. Ecological and environmental damage assessment of water resources protection mining in the mining area of Western China. *Ecol. Ind.* **139**, 108938 (2022).
- State Administration of Work Safety, National Coal Mine Safety Administration, National Energy Administration, National Railway Administration. *Regulations for coal pillar retention and pressure mining under buildings, water bodies, railways and major shafts* (China Coal Industry Publishing House, 2017).
- Brown, S. & Walsh, R. The height of connected fracturing above a longwall coal mine: a field and isotopic investigation. In Proceedings of the 11th conference on mine subsidence. Pokolbin, NSW 135–140 (2022).
- Luo, B., Zhang, C. H., Zhang, P., Huo, J. Y. & Liu, S. D. A combined method utilizing microseismic and parallel electrical monitoring to determine the height of water-conducting fracture zones in Shengfu coal mine. *Water* **16** (21), 3047 (2024).
- Zhang, L. F. Development and height prediction of fractured water-conducting zone in weakly cemented overburden: a case study of Tashidian Erjingtian mine. *Sustainability* **15** (18), 13899 (2023).
- Miao, X. X., Cui, X. M., Wang, J. A. & Xu, J. L. The height of fractured water-conducting zone in undermined rock strata. *Eng. Geol.* **120**, 32–39 (2011).
- Zheng, L. L. et al. Study of the development patterns of water-conducting fracture zones under karst aquifers and the mechanism of water inrush. *Sci. Rep.* **14**, 20790 (2024).

11. Wang, X. H., Zhu, S. Y., Yu, H. T. & Liu, Y. X. Comprehensive analysis control effect of faults on the height of fractured water-conducting zone in Longwall mining. *Nat. Hazards*. **108** (2), 2143–2165 (2021).
12. Liu, Y., Yuan, S. C., Yang, B. B., Liu, J. W. & Ye, Z. Y. Predicting the height of the water-conducting fractured zone using multiple regression analysis and GIS. *Environ. Earth Sci.* **78** (14), 422 (2019).
13. Li, X. B., Li, Q. S., Xu, X. H., Zhao, Y. Q. & Li, P. Multiple influence factor sensitivity analysis and height prediction of water-conducting fracture zone. *Geofluids* **2021**, 8825906 (2021).
14. Xu, C., Zhou, K., Xiong, X., Gao, F. & Zhou, J. Research on height prediction of water-conducting fracture zone in coal mining based on intelligent algorithm combined with extreme boosting machine. *Expert Syst. Appl.* **249**, 123669 (2024).
15. Wang, H. Z., Zhu, J. Z. & Li, W. P. An improved back propagation neural network based on differential evolution and grey Wolf optimizer and its application in the height prediction of water-conducting fracture zone. *Appl. Sci.* **14** (11), 4509 (2024).
16. Zhao, D. K. et al. Using swarm intelligence optimization algorithms to predict the height of fractured water-conducting zone. *Energy Explor. Exploit.* **41** (5), 1603–1627 (2023).
17. Zhao, D. K. & Wu, Q. An approach to predict the height of fractured water-conducting zone of coal roof strata using random forest regression. *Sci. Rep.* **8** (1), 10986 (2018).
18. Wu, Z. Y., Chen, Y. & Luo, D. Y. Comparative study of five machine learning algorithms on prediction of the height of the water-conducting fractured zone in undersea mining. *Sci. Rep.* **14**, 21047 (2024).
19. Wang, M. et al. Prediction of coal mine water conduction fracture zone height based on integrated learning model. *Sci. Rep.* **15**, 28101 (2025).
20. Wang, Z. C. et al. Intelligent prediction of the development height of the water-conducting fracture zone based on the SSA-RF model. *Coal Sci. Technol.* <https://doi.org/10.11731/j.issn.1673-193x.2020.10.012> (2020).
21. Zhang, H. W., Zhu, Z. J., Huo, B. J. & Song, W. H. Water flowing fractured zone height prediction based on improved FOA-SUM. *China Saf. Sci.* **23** (10), 9–14 (2013).
22. Xun, B. H., Lyu, Y. Q. & Yao, X. Comparison of prediction models for the development height of water-conducting fractured zone. *Coal Sci. Technol.* **51** (3), 190–200 (2023).
23. Li, Z. H., Xu, Y. C., Li, L. F. & Zhai, C. Z. Forecast of the height of water flowing fractured zone based on BP neural networks. *J. Min. Saf. Eng.* **32** (6), 905–910 (2015).
24. Wu, J. H., Pan, J. F., Gao, J. M., Yan, Y. D. & Ma, H. Y. Research on prediction of the height of water-conducting fracture zone in Huanglong jurassic coalfield. *Coal Sci. Technol.* **51** (Sup.1), 231–241 (2023).
25. Liu, Q., Liang, Z. H. & Zi, J. X. A SMOGN-based MPSO-BP model to predict the height of a hydraulically conductive fracture zone. *Coal Geol. Explor.* **52** (11), 72–85 (2024).
26. Xiao, L. L. et al. Evaluation of water inrush hazard in coal seam roof based on the AHP-CRITIC composite weighted method. *Energies* **16** (1), 114 (2023).
27. Jhariya, D. C. et al. Assessment of groundwater potential zone using GIS-based multi-influencing factor (MIF), multi-criteria decision analysis (MCDA) and electrical resistivity survey techniques in Raipur city, Chhattisgarh, India. *Aqua-water Infrastructure Ecosyst. Soc.* **70** (3), 375–400 (2021).
28. Feng, D., Hou, E. K., Xie, X. S. & Hou, P. F. Research on water-conducting fractured zone height under the condition of large mining height in Yushen mining area. *China Lithosphere*. **2023** (1), 8918348 (2023).
29. Dolatshahi, A. & Molladavoodi, H. Prediction of rock tensile fracture toughness: hybrid ANN-WOA model approach. *Rudarsko-geolosko-nafini Zbornik*. **39** (3), 1–12 (2024).
30. Fadaee, M., Mahdavi-Meymand, A. & Zounemat-Kermani, M. Suspended sediment prediction using integrative soft computing models: on the analogy between the butterfly optimization and genetic algorithms. *Geocarto Int.* **37** (4), 961–977 (2020).
31. Rouhani, M. M. & Farrokh, E. Chisel Bits cutting force Estimation using XGBoost and different optimization algorithms. *Comput. Geotech.* **172**, 106465 (2024).
32. Xu, D. J. et al. Research on height zoning prediction of water-conducting fracture zone under the control of different overburden rock types in North China coal-rich region. *Coal Geol. Explor.* **53** (3), 177–189 (2025).

Author contributions

Conceptualization, Z.T., H.T. and D.Z.; methodology, Z.T., H.T.; software, H.T., C.R., Y.S.; validation, C.R., X.G.; data curation, H.T., C.R., Y.Z.; writing—original draft preparation, Z.T., H.T., D.Z.; writing—review and editing, D.Z., G.F., C.R., Z.W.; supervision, G.F., C.R. All authors have read and agreed to the published version of the manuscript.

Funding

Supported by the National Key Research and Development Program of China (Grant No. 2024YFC3909301), the National Natural Science Foundation of China (Grant No. 52204161, 52574177) and the Graduate Research and Innovation Projects of Jiangsu Province (Grant No. KYCX24_2859).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to D.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025