MACHINE LEARNING BASED PREDICTION MODEL AND INTERPRETABLE ANALYSIS OF AIRFLOW TEMPERATURE IN WATER DROPPING SHAFT

Fei TANG [*]***, Yueping QIN, Shibin Wang, Peng Wang, Yanjie Yang, Mingyan Guo, Xiangyun Su, and Jiachang Wang***

School of Emergency Management and Safety Engineering, China University of Mining and Technology-Beijing, Beijing 100083, PR China

[*] Corresponding author; E-mail: 18325533132@163.com

Accurate prediction of the airflow temperature in the mine water dropping shaft is crucial for scientifically guiding the prevention and control of thermal hazards in mines. To improve the precision, stability, and interpretability of the prediction model for the airflow temperature in mine drainage shafts, combined with relevant literature research and the actual characteristics of mine airflow temperature, Pearson correlation analysis was introduced to screen the characteristic variables. The parameters of the LSBoost model were optimized using the WOS model, and a mine drainage shaft airflow temperature prediction model based on WOS-LSBoost was established. Under the same sample set conditions, compare the prediction precision of the five models with the established DT model, RF model, LSBoost model, and SVM model, and conduct interpretability analysis on the prediction models based on the SHAP analysis method. Research has demonstrated that the WOS-LSBoost model has the best predictive property, stronger generalization ability, and can effectively improve the precision of wind temperature prediction; the temperature of the wellhead airflow has the largest effect on the prediction model, while the surface pressure has the smallest effect on the prediction model; This research can offer scientific reference for the prevention and control of thermal hazards in mines.

Key words: *Water dropping shaft, airflow temperature prediction, Pearson correlation, WOS-LSBoost model, interpretability analysis*

1. Introduction

With the extension of mining depth and the increase in mining intensity, the issue of high-temperature thermal damage in mines is growing increasingly serious [1, 2]. The high temperature and high humidity underground environment can lead to problems such as reduced labor efficiency of miners, increased safety production accident rates, and increased mechanical equipment failure rates. Due to the extremely complex heat transfer process of mine airflow, the difficulty of predicting air temperature has increased [3]. Therefore, studying the changes in mine airflow temperature, proposing reliable airflow temperature prediction models, and continuously improving the accuracy of airflow temperature prediction are of great significance for scientifically evaluating the risk level of mine heat hazards, improving the high-temperature environment of mines, guiding and formulating mine heat hazard prevention and control plans, and other aspects.

The prediction of airflow temperature in the mine drainage shaft is the most important and complex part of the entire mine airflow temperature prediction process. The heat exchange process involves heat and moisture transfer, which is influenced by multiple factors such as surface wind temperature, wellbore depth, and geothermal gradient, bringing enormous hardship to predict the wind temperature of the water dropping shaft. Many scholars have conducted extensive research on predicting mine airflow temperature. Bascompta et al. [4] established a temperature prediction model for mine ventilation lines based on measured data such as temperature, airflow volume, and ventilation distance. Zhu et al. [5] established a temperature prediction model for underground airflow routes in ultra-deep mines based on the concept of thermal equilibrium. Akhmetova et al. [6] used Comsol to simulate the changes in shaft airflow temperature and heat loss. In late years, along with the ongoing growth and extension of machine learning, many scholars have applied intelligent algorithms such as deep learning and neural networks to predict mine temperature. Alqahtani et al. [7] used various machine learning methods to predict borehole temperature and evaluate geothermal potential. Liu et al. [8] used machine learning methods to predict the highest temperature zone in the goaf of longwall mining. Zhang et al. [9] adopted GA-BP neural network approaches to forecast the airflow temperature at the excavation face. However, the above models are prone to getting stuck in local optimal solutions during operation, which may greatly affect the stability and accuracy of wind temperature prediction, and lack interpretability analysis of wind temperature prediction models.

To overcome the mentioned issues, this article establishes a prediction model for the airflow temperature of mine water dropping shafts based on WOS-LSBoost. Using 60 sets of measured data from 20 mines as samples, the LSBoost model was used to capture the characteristic information of predicting the airflow temperature in the water dropping shaft. Combined with the WOS algorithm, the problem of the prediction model easily getting stuck in local optima was solved, thereby improving the accuracy of airflow temperature prediction. By comparing and analyzing with other common models, the superiority of the established model is demonstrated. The SHAP analysis method is introduced to conduct an interpretability analysis of the prediction model, the effects of different characteristic covariates on the predictive model. This model provides a new method and approach for predicting the airflow temperature in water dropping shafts.

2. Prediction model construction

2.1. Pearson correlation analysis

Before starting the prediction model, it is necessary to conduct a correlation test on the feature variables. By conducting correlation analysis to select feature variables, this article uses the Pearson correlation coefficient method for feature variable analysis, which is calculated as follows :

$$P(x,y)=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{1}$$

Among them, $P\ (x, y)$ is the Pearson correlation coefficient; x_i and y_i respectively represent the values of two characteristic variables at each observation point; $\bar{x}$ and $\bar{y}$ are the means of two characteristic variables.

The Pearson correlation coefficient is in the interval [-1,1], and the greater the absolute value, the stronger the correlation of the characteristic variables. When the absolute value is less than 0.5, the two characteristic variables are lowly correlated.

2.2. WOS model

WSO is a metaheuristic optimization algorithm that is influenced by the attributes of great white sharks, such as their extraordinary hearing and smell when navigating and foraging. Fig. 1 shows the computational procedures of the model.

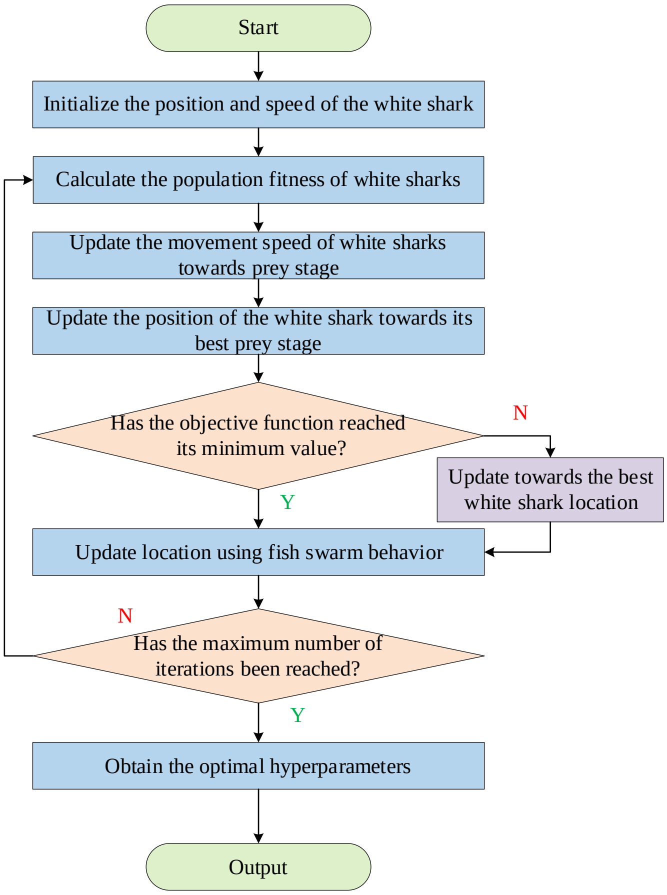

Figure 1. Algorithm Flow of WOS Model

From fig. 1, the model is calculated using the following procedure:

Step 1: Initialize the white shark population, set the initialization parameters and fitness function of the objective function, and randomly generate the positions and velocities of all white sharks.

Step 2: Based on the parameters of each white shark P, that is, evaluating the adaptation value of each white shark through the fitness function, as the selection criteria.

Step 3: Update the movement speed of each white shark towards the prey stage.

Step 4: Update the position of each white shark towards the best prey stage.

Step 5: If the fitness function reaches the threshold, update the position towards the best white shark and use fish swarm behavior to update all white shark positions.

Step 6: Output the global optimal solution.

2.3. LSBoost model

LSBoost is an algorithm that sequentially constructs weak learners to improve model performance. This algorithm minimizes the loss function, typically the squared error loss, by iteratively adding weak learners, with each model focusing on the residuals of the previous model. It can gradually improve the predictive performance of the model by iteratively correcting errors in previous models in data regression. Fig. 2 demonstrates the calculation steps of the model.

From fig. 2, initialization is first performed, usually by calculating the mean of the target variable to initialize the model, that is, finding the value that minimizes the loss function concerning the constant as the initial model. Continuing into the iterative process, in each iteration, the residual between the current model and the true value is first calculated. Then, the residual is used as a new "target value" to fit a regression tree, and the weights of this regression tree are calculated. Finally, the model is updated based on the weights. Repeat multiple iterations in this way until the preset number of iterations is achieved, and finally output the model after multiple iterations as the prediction model.

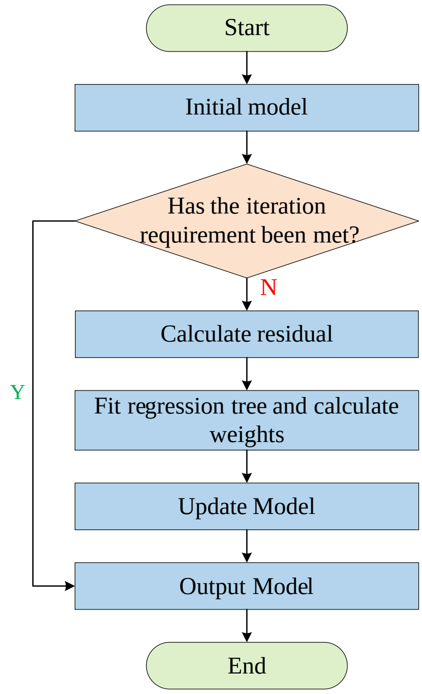

Figure 2. Calculation process of LSBoost model

2.4. WOS-LSBoost prediction model

Using the WOS model to optimize the LSBoost algorithm, establish a WOS-LSBoost prediction model for the airflow temperature of the water dropping shaft. Fig. 4 shows the calculation process of the WOS-LSBoost prediction model. The procedure of this model is as follows:

(1) Using Pearson correlation analysis to explore the relevance between feature variables and determine the feature variables of the prediction model.

(2) Screening and processing samples, importing measured data samples for wind temperature prediction; Dividing the wind and temperature data samples into training and testing sets, with the training and testing set accounting for 80% and 20% respectively.

(3) Establish a wind temperature prediction model and set parameters such as the number of LSBoost neurons. Search for relevant parameters through the WOS algorithm, optimize the

superparameters of the network model, and obtain the optimal network hyperparameters to prevent the model from getting stuck in local optimal solution problems.

(4) Import wind and temperature data training set, use LSBoost model to extract parameter features and establish WOS-LSBoost prediction model. Check if the set error of 0.001 has been reached, if so, proceed to the next step; Otherwise, retrain the training set until it meets the standard.

(5) Import wind temperature data test set and output predicted results of shaft wind temperature; Compare the absolute error, relative error, and other parameters with the measured wind temperature to assess the precision of the model.

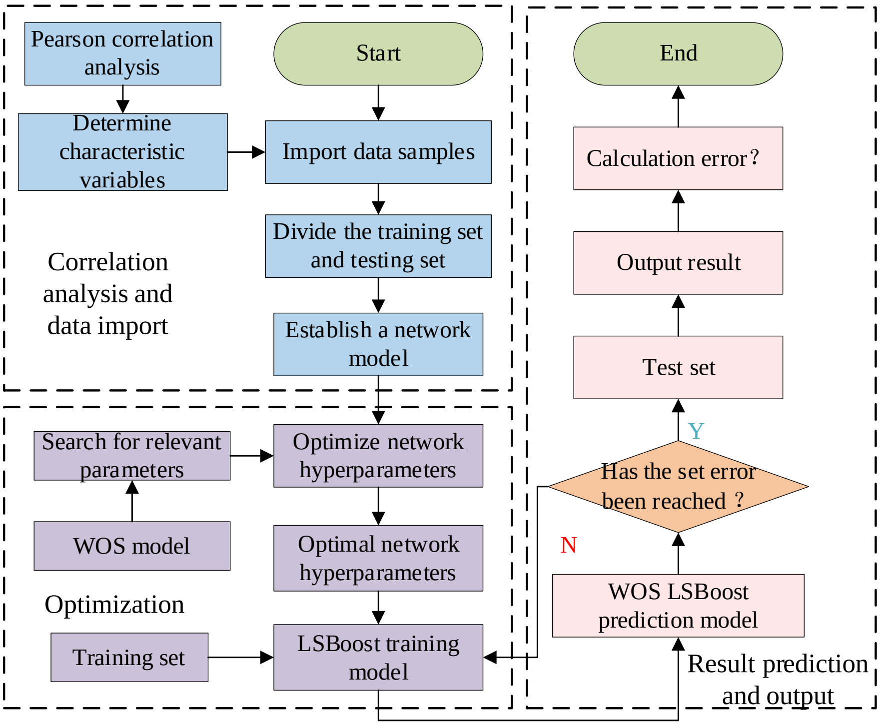

Figure 3. WOS-LSBoost Prediction Model Process

2.5. Prediction result evaluation

To evaluate the predictive performance and effectiveness of different prediction models, this research selects mean square error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), and coefficient of determination (R^2) as evaluation indicators for the prediction models. Among them, MSE, MAPE, RMSE, and MAE are all measures of the deviation between actual values and predicted values. The larger the value, the less accurate the prediction model is. R^2 reflects the distribution trend of measured values and predictions and can be used to assess the fitting effect of prediction models. The nearer the R^2 value is to 1, the greater the prediction effect and the higher the fitting accuracy of the model. The calculation formula is as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n}\left|P_j - Q_j\right| \tag{2}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{j=1}^{n}\left|\frac{P_j - Q_j}{Q_j}\right| \times 100\% \tag{3}$$

$$\mathrm{MSE}=\frac{1}{n}\sum_{j=1}^{n}\left(P_j-Q_j\right)^2 \tag{4}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(P_j-Q_j\right)^2} \tag{5}$$

$$R^2=1-\frac{\sum_{j=1}^{n}\left(Q_j-P_j\right)^2}{\sum_{j=1}^{n}\left(Q_j-\overline{Q}\right)^2} \tag{6}$$

Among them, n is the data number; P_j is the predicted value; Q_j is the measured value; $\overline{Q}$ is the mean of the measured values.

3. Predictive examples and result analysis

3.1. Factors affecting wind temperature and sample data

The heat transfer process in the water dropping shaft is complex, and the airflow is affected by factors such as surrounding rock heating, wind speed, rock thermal physical properties, and relative humidity of the airflow. There are heat transfer, moisture transfer, and mass transfer in this process, which makes it very difficult to theoretically predict the wind temperature in the mine drainage shaft. The overall thermal and physical properties of rocks in different mines are not significantly different, and the wind speed in each mine shaft is very close to the maximum safety standard of coal mines (15 m/s). Combining relevant literature on predicting the airflow temperature of mine dropping drainage shafts with actual on-site testing data. This article selects wellhead wind temperature, relative humidity of wellhead airflow, ground atmospheric pressure, shaft depth, and ground temperature gradient as characteristic variables for predicting actual wind temperature in the prediction model. Select 60 sets of on-site measured parameters such as wind temperature and humidity as sample data, with the first 48 sets as the training set and the last 12 sets as the testing set. Partial on-site data is shown in Table 1.

Table1 Sample data

Number	Airflow temperature /°C	Airflow relative humidity /%	Ground atmospheric pressure /Pa	Shaft depth /m	Bottom hole air temperature /°C	Geothermal gradient /（hm/°C）
1	26	91	100 380	476.35	25.6	2.8
2	26.6	81.7	100 240	476.35	25.8	2.4
3	26.4	85.4	100 339	476.35	25.5	3.1
4	29.8	63	100 520	518.78	28.6	2.8
5	28.8	62.6	100 880	518.78	27.6	3.2
⋮	⋮	⋮	⋮	⋮	⋮	⋮
58	26.8	79	99 965	673.20	26.8	3.5
59	28.1	80	99 868	673.20	27.9	3.3
60	15.7	67.3	100 350	482.02	16.6	2.7

3.2. Pearson correlation test

Calculate the Pearson correlation coefficients of each feature variable in Table 1 using equation (1), and draw a heatmap of the correlation coefficients of the feature variables, as shown in fig. 5. According to fig. 5, the correlation between geothermal gradient and shaft depth is high at 0.94, indicating a strong correlation. Due to the small difference in ground temperature gradient among different mines, while there are significant differences in shaft depth, we will exclude the ground temperature gradient. The Pearson correlation coefficients of the other four characteristic variables are all less than 0.5. The correlation between the four characteristic variables is relatively weak, meeting the requirement of correlation between them as prediction variables for shaft airflow temperature at the same time. Therefore, four variables such as wellhead airflow temperature and shaft depth are taken as the selected values of the prediction model.

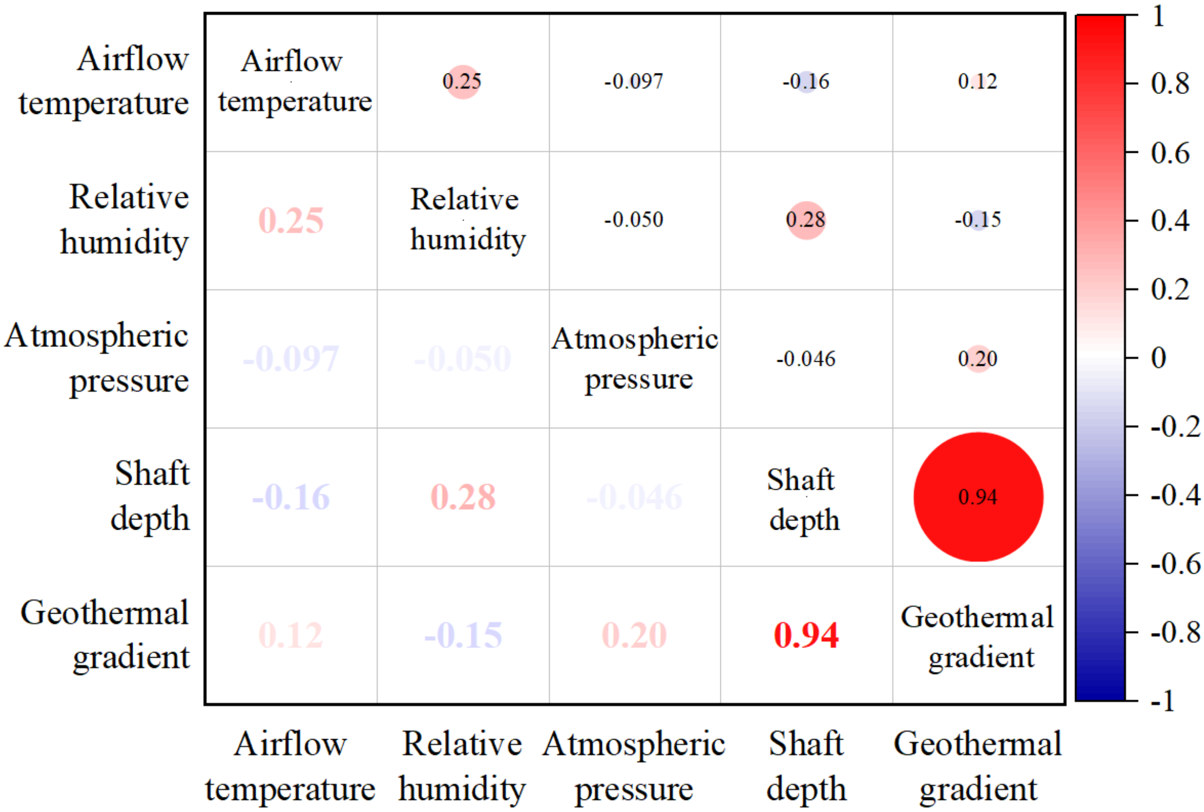

Figure 4. Heat map of correlation coefficients of characteristic variables

3.3. Additional instructions (Word Style TS Heading 2)

To assess the precision of the model more scientifically, MAE, MAPE, MSE, RMSE, and R^2 were utilized to assess the prediction performance of the five prediction models. Table 2 reveals the comparison of the evaluation results of the predicted values of five models. According to Table 2, compared with the DT model, RF model, MLP model, and LSBoost model, the WOS-LSBoost model has the lowest MAE, MAPE, MSE, and RMSE values, which are 0.42 ℃, 1.8%, 0.33, and 0.58, respectively. The R^2 values of the DT model, RF model, MLP model, LSBoost model, and WOS-LSBoost model are 0.956, 0.929, 0.932, 0.962, and 0.981, respectively. The R^2 values of all five models are very close to 1, but the WOS-LSBoost model has the highest R^2 value, indicating that this model has the optimal predictive properties. The prediction outcomes manifest that the WOS-LSBoost model has high accuracy and stability, which can improve the accuracy of predicting the air temperature in the water dropping shaft. In the current situation where the measured data variables in the mine are limited, this model only uses four sets of main control characteristic variables to achieve good prediction results in predicting the airflow temperature of the water dropping shaft and has great superiority.

Table 2. Evaluation indicators for 5 prediction models

Model	MAE	MAPE	MSE	RMSE	R^2
DT	0.7173	0.03270	0.7894	0.8885	0.9561
RF	0.7622	0.03110	1.2670	1.1256	0.9294
MLP	0.8425	0.03493	1.2276	1.1080	0.9317
LSBoost	0.4995	0.02191	0.6857	0.8281	0.9618
WOS-LSBoost	0.4225	0.01848	0.3356	0.5793	0.9813

4. Explanation of SHAP based prediction model

The WOS-LSBoost model itself belongs to the "black box" model, and it is impossible to know the specific process of its training and prediction. To obtain a stronger appreciation of the sensitivity of various factors in the predictive model, this article uses the SHAP analysis approach to explain and analyze the established model.

The SHAP model for interpreting predictions from machine learning models is based on the SHAP value in game theory. SHAP value is a fair distribution game payment used to determine the contribution of each feature to the overall outcome of the game. Its calculation principle is as follows [10]:

$$\phi_i = \sum_{S \subseteq N\{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} \big(v(S \cup \{i\}) - v(S)\big) \tag{7}$$

In the formula: *N* is the collection of all features; *S* is any subset of features that do not contain feature *i*;| S | is the feature number in set S; *V* (*S*) is the contribution of the feature set *S* to the model's forecasted output; *v* (S U{*i*}) is the contribution of the feature set *S* {*i*} containing feature *i* to the model's forecasted output. Based on the additive feature attribute method, define a linear function *f*, that is:

$$f(x) = \Phi_0 + \sum_{m=1}^{M} \Phi_m x_m \tag{8}$$

In the formula, *f*(*x*) is the forecasted value of the post hoc explanatory model for sample *x*; The average value of the model's predicted values is represented by Φ_0; X_m is the mth feature sample.

According to equations (7) and (8), the importance ranking of feature variables and their impact on model prediction are shown in fig. 5 and 6. As shown in fig. 5, the importance of the model feature variables is ranked from large to small as wellhead airflow temperature, shaft depth, wellhead relative humidity, and surface atmospheric pressure. Among them, the wellhead airflow temperature has the greatest impact on the model prediction results and is the most important characteristic variable in the shaft airflow temperature prediction process.

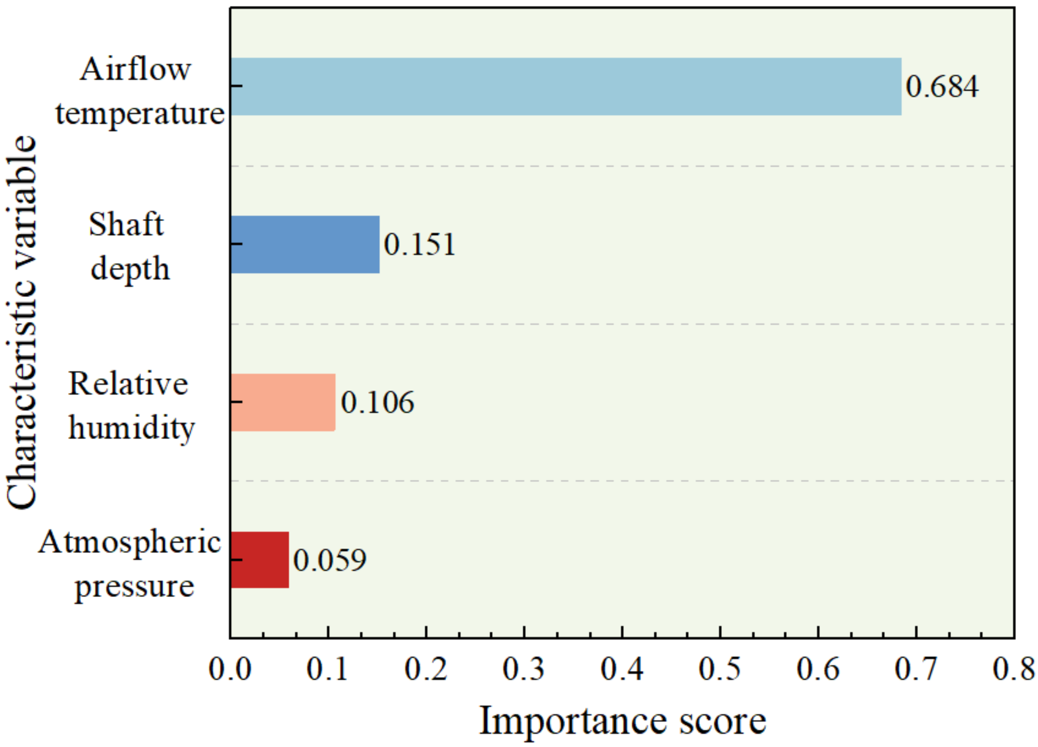

Figure 5. Importance ranking of characteristic variables

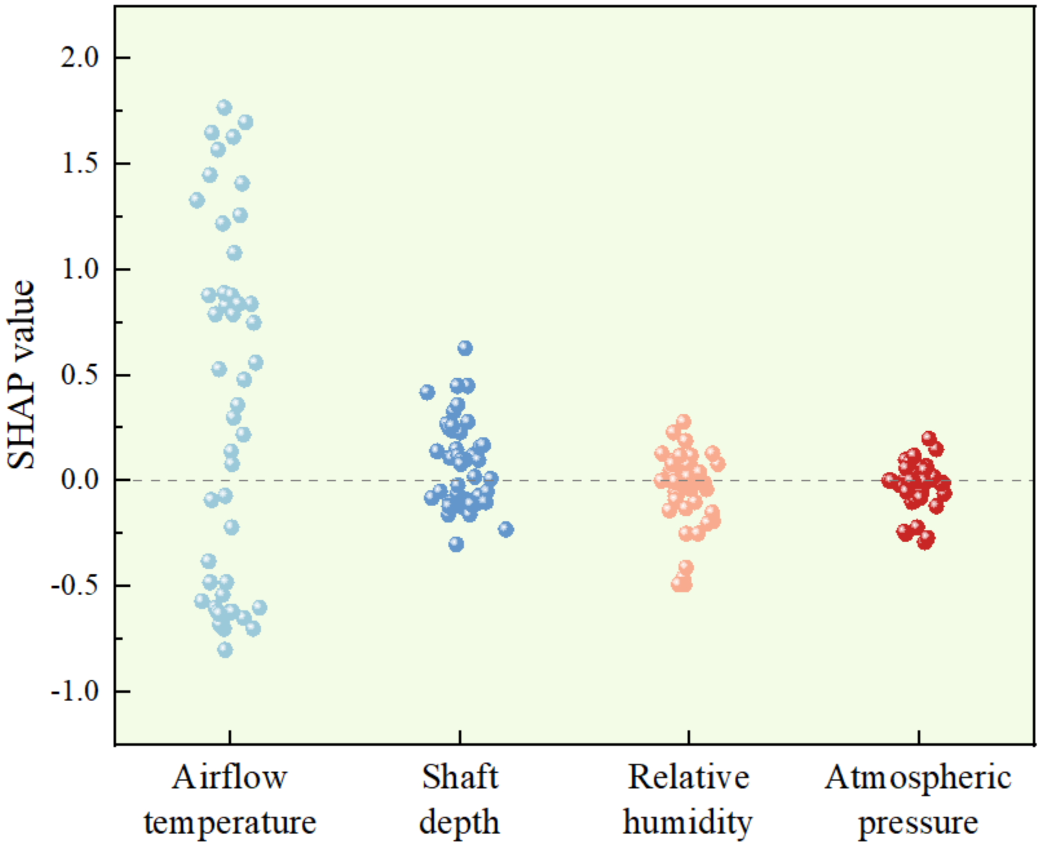

Figure 6. Impact of Characteristic Variables on Model Predictions

A positive SHAP value in fig. 6 indicates a positive correlation between feature factors, and vice versa. As shown in fig. 6, the SHAP value increases with the increase of wellhead airflow temperature and shaft depth. The larger the wellhead airflow temperature and shaft depth, the higher the bottom airflow temperature of the shaft. The SHAP value decreases with the increase of relative humidity at the wellhead and surface atmospheric pressure. The higher the relative humidity at the wellhead and the atmospheric pressure on the ground, the lower the airflow temperature at the shaft's bottom. In summary, the best way to suppress the augment in shaft airflow temperature is to control the temperature of the airflow at the wellhead and; Secondly, reduce the depth of the shaft.

5. Conclusion

This article establishes a water dropping shaft airflow temperature prediction model based on WOS-LSBoost. By utilizing the LSBoost model to more effectively capture sample features and combining it with the WOS algorithm to solve the problem of the model easily getting stuck in local optima, the accuracy and stability of wind temperature prediction have been improved. The SHAP analysis method was used to conduct an interpretable analysis of the prediction model, and the main conclusions are:

1) The average relative errors of the DT model, RF model, MLP model, LSBoost model, and WOS-LSBoost model are 3.27%, 3.11%, 3.49%, 2.19%, and 1.85%, respectively. The WOS-LSBoost model has the highest prediction accuracy. This model can be used in situations where there are few data samples and few feature variables.

2) In contrast to the DT model, RF model, MLP model, and LSBoost model, the WOS-LSBoost model has lower prediction errors, with MAE, MSE, and RMSE of 0.42 ℃, 0.33, and 0.58, respectively, and an R^2 value of 0.981, demonstrating that the model has stronger generalization ability and better prediction performance.

3) Based on the SHAP interpretation model, it can better explain the logic behind the prediction model, analytics the influence of various feature variables on the forecast model, and prioritize the control measures for the wellhead airflow temperature, which has the greatest impact on the forecast results of the model, to suppress the rise of bottom hole airflow temperature.

Acknowledgment

In this study, the authors gratefully acknowledge support from the National Key Technology Research and Development Program of China [Grant No. 2022YFC2904100].

References

[1] Guo, M.Y., *et al.*, Prediction of airflow temperature in coal mining face under the influence of multiple heat sources coupling: Numerical simulation and verification, *Case Studies in Thermal Engineering*, *61*. (2024), DOI No. 10.1016/j.csite.2024.105128

[2] Tang, F., *et al.*, Thermal risk assessment in the mine of TOPSIS based on game theory subjective and objective comprehensive weighting, *Thermal Science and Engineering Progress*, *59*. (2025), p. 103367, DOI No. https://doi.org/10.1016/j.tsep.2025.103367

[3] Tang, F., *et al.*, Multi-angle research on the performance of self-foaming filling materials and analysis of engineering applications, *Construction and Building Materials*, *449*. (2024), DOI No. 10.1016/j.conbuildmat.2024.138447

[4] Bascompta, M., *et al.*, Temperature Prediction Model in the Main Ventilation System of an Underground Mine, *Applied Sciences-Basel*, *10*. (2020), 20, DOI No. 10.3390/app10207238

[5] Zhu, S., *et al.*, An Underground Air-Route Temperature Prediction Model for Ultra-Deep Coal Mines, *Minerals*, *5*. (2015), 3, pp. 527-545, DOI No. 10.3390/min5030508

[6] Akhmetova, G., *et al.*, Simulation of wellbore temperature and heat loss in production well in geothermal site, *Energy Sources Part a-Recovery Utilization and Environmental Effects*, *45*. (2023), 2, pp. 6262-6273, DOI No. 10.1080/15567036.2023.2213673

[7] Alqahtani, F., *et al.*, Machine Learning Techniques in Predicting Bottom Hole Temperature and Remote Sensing for Assessment of Geothermal Potential in the Kingdom of Saudi Arabia, *Sustainability*, *15*. (2023), 17, DOI No. 10.3390/su151712718

[8] Liu, W., *et al.*, Dynamic prediction of high-temperature points in longwall gobs under a multi-field coupling framework, *Process Safety and Environmental Protection*, *187*. (2024), pp. 1062-1075, DOI No. 10.1016/j.psep.2024.04.097

[9] Zhang, X., *et al. Research on Forecasting of Airflow Temperature in Heading Face Based on The GA-BP Neural Network*,1st International Conference on Energy and Environmental Protection (ICEEP 2012),Hohhot, PEOPLES R CHINA,2012,524-527, pp. 668-672

[10] Palar, P.S., *et al.*, Enhancing the explainability of regression-based polynomial chaos expansion by Shapley additive explanations, *Reliability Engineering & System Safety*, *232*. (2023), DOI No. 10.1016/j.ress.2022.109045

Submitted: 29.03.2025.
Revised: 19.09.2025.
Accepted: 28.09.2025.