IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 2 January 2024, accepted 26 January 2024, date of publication 13 February 2024, date of current version 21 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3365496

RESEARCH ARTICLE

Forecasting the Friction Factor of a Mine Airway Using an Improvement Stacking Ensemble Learning Approach

ZHIPENG QI[1,2,3], KE GAO[1,3], DARIUSZ OBRACAJ[2], YUJIAO LIU[1,3], AND KEYI YUAN[1,3]

[1]College of Safety Science and Engineering, Liaoning Technical University, Huludao, Liaoning 125105, China
[2]Faculty of Civil Engineering and Resource Management, AGH University of Kraków, 30-059 Kraków, Poland
[3]Key Laboratory of Mine Thermodynamic Disasters and Control, Ministry of Education, Huludao, Liaoning 125105, China

Corresponding author: Ke Gao (gaokelntu@163.com)

This work was supported by the National Science Foundation of China under Grant 52074148 and Grant 52104194.

ABSTRACT The prediction of frictional air resistance using the inherent properties of roadways is of great significance for ventilation network computation and flow regulation in underground mines. This study proposes an improved stacked learning and error correction-based prediction model for the frictional air resistance of mine airways, called friction factor. A prediction set is established by selecting ten factors, including tunnel spatial features and support forms, with the ventilation resistance coefficient as the label. The improved stacked model consists of two layers. The first layer is the base learning module, which is composed of four components: Principal Components Analysis and Back Propagation (PCA-BP), GA-Projection Pursuit Regression (GA-PPR), Random Forest (RF), LightGBM (LGBM). The second layer is the meta-learning module, which is composed of the Ridge Regression (RR). Compared to traditional stacked models, the improved model first uses the Extreme Gradient Boosting (XG Boost) learner to evaluate the significance of input feature variables to eliminate redundancy and improve accuracy, thus enhancing prediction precision and computational efficiency. Then, the first-layer prediction results are weighted based on the errors of different prediction models in the training set using K-fold cross-validation. Box-Cox transformation is applied to the training set data from the first layer to the second layer to improve prediction normality and homogeneity. The error correction prediction model extracts the historical prediction errors from the meta-learning module and constructs an error prediction model using support vector machines (SVR), which are then combined with the meta-learning results to obtain the final prediction. The improved stacked model is compared with traditional ensemble learning models and single prediction models, and quantified using three metrics: root mean square error (RMSE), mean absolute error (MAE), and R-squared (R^2). The results demonstrate that the proposed improved model effectively enhances the prediction accuracy of the ensemble learning models, providing a new prediction method for the accurate acquisition of the friction factor of mine airways.

INDEX TERMS Mine airways, frictional air resistance coefficient, improved stacking model, cross-validation, prediction accuracy.

I. INTRODUCTION

With the rapid expansion of the underground mining roadway structure, the design requirements for ventilation systems in complex networks have undergone significant changes. During the construction process of roadways, various types of waste gases and dust are inevitably generated, posing serious safety risks to construction personnel [1], [2], [3], [4], [5], [6], [7]. Therefore, the design of ventilation systems plays a crucial role in roadway development.

The associate editor coordinating the review of this manuscript and approving it for publication was Akin Tascikaraoglu.

In mine ventilation terminology, individual roadways are called airways and form a mine ventilation network through interconnections. Airway resistance plays the principal role in air distribution in the network. It is a critical parameter in ventilation network design, as it directly influences the assessment of air demand in mining areas of the mine. The prediction of airflow splitting in a new mine ventilation network is made by various software using numerical methods in which the resistances of airways play a crucial role [8].

Air resistance of the mine airway defines the pressure drop along a section of the airway during the air flow rate. Airway resistance is a proportional constant that considers both the frictional air resistance coefficient of straight sections of the airway and shock losses at bends, variations in cross-sectional area, or junctions of mine airways when splitting or merging airflows or other obstacles. Considering the equivalent airway resistance, the shock losses can be related to the equivalent length of a straight section of the airway. Therefore, airway resistance R (Ns^2/m^8) can be related proportionally to fundamental coefficient of friction λ, or if standard air density ρ is taken into account to frictional air resistance coefficient α, known as the Atkinson friction factor or simply friction factor.

$$R = \frac{\lambda \rho}{8} \frac{\left(L + L_{eq}\right) P}{A^3} = \alpha \frac{\left(L + L_{eq}\right) P}{A^3} = \frac{\Delta p}{Q^2} \qquad (1)$$

where λ is a dimensionless coefficient of friction (known as the Darcy friction factor); α is a frictional air resistance coefficient (known as the friction factor), kg/m^3; ρ is air density, kg/m^3; L is a length of airway section, m; L_{eq} is an equivalent length of shock loses in section of airway, m; P is perimeter of airway, m; A is a cross-sectional area of airway, m^2; Δp is a pressure drop, Pa; Q is an airflow rate, m^3/s.

At present, there are three methods to obtain the frictional air resistance coefficient. The first is to use the existing theoretical formula to calculate this coefficient. In the 18th century, Antoine de Chézy and next Henry Darcy carried out research on canal resistance and put forward Darcy equation or the Darcy-Weisbach equation, which opened the precedent for the study of frictional air resistance coefficient and gradually applied to other field [9]. In 1994, Moody conducted experimental verification to investigate the relationship between roughness, Reynolds number, and flow resistance [10]. Based on Nikuradse's experimental curve, Prandtl has established a resistance coefficient formula applicable to smooth pipes [11]. On the other hand, Theodore von Karman has established a frictional air resistance coefficient formula applicable to the rough region of turbulent flow. In this region, the resistance coefficient is independent of the Reynolds number and only depends on the relative roughness of the pipe wall. In 1939, Colebrook and White combined the aforementioned formulas and proposed a new equation known as the Colebrook-White (CW) equation [12]. The most well-known formula for estimating the coefficient of friction is as follows [13]:

$$\frac{1}{\sqrt{\lambda}} = -2 \log \left(\frac{e}{3.7 D_h} + \frac{2.51}{\mathrm{Re}\sqrt{\lambda}} \right) \qquad (2)$$

where e is the height of the roughening or asperity, m; D_h is the hydraulic mean diameter of the airway, m, Re is a Reynolds number.

Due to the complexity and relatively large computational errors involved in using the CW equation for large-scale mine tunnels or roadways, there are two main directions in the research on coefficient of friction λ. The first direction involves developing explicit approximation formulas for calculating the coefficient of friction based on the CW equation using various mathematical methods. The second direction focuses on exploring new equations that offer improved accuracy for specific application scenarios [14]. However, the latter approach often requires significant human and material resources and faces challenges in modifying empirical formulas. Currently, researchers commonly adopt the first approach, which involves establishing explicit calculation formulas based on the CW equation [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25]. Based on the current research status of the friction factor, it can be observed that although there are numerous theoretical formulas for the frictional air resistance, their practical application in engineering is challenging due to the difficulty in accurately measuring the absolute roughness. Therefore, these formulas are difficult to be applied in practical engineering scenarios.

The second method consists of in situ measurement of pressure drop and airflow in rectilinear sections of existing airways and determination of friction factors with air density correction taken into account [26], [27]. The results are subsequently employed to predict the resistances of similar planned airways. Often, due to the lack of similarities between mine airways resulting from the differences in shape and roughness of the roads, but also the type of equipment, unrepresentative obstructions or blockages, and the lack of longer straight sections, this is not always possible and brings the expected benefits.

The third method for obtaining friction factor α is by consulting empirical handbooks [28], [29], [30], [31]. However, with the rapid development of mining technology, changes have occurred in roadway cross-section dimensions, support types, and roadway infrastructure. The limited number of airway types provided in the handbooks is insufficient to cover various real-world applications. It is difficult to find exact values corresponding to a specific mine airway in the handbook, making the tabulation method rarely used in actual mine ventilation. Therefore, other methods of determining friction factors or dimensionless friction coefficients are sought. With the development of artificial intelligence, researchers are to use machine learning to predict the friction factor according to the inherent properties of tunnels such as support and shape.

In recent decades, evolutionary computing (EC) in the form of artificial intelligence (AI) based methods has been profusely applied to solve technical problems in different fields of pipeline airflow and underground mine ventilation networks. Among others, Decision Trees (DT), Gene-Expression Programming (GEP), Support Vector Machine (SVM), Naive Bayes (NB), and Logistic Regression (LR) techniques were frequently used to characterize a broad genre of different problems in airway resistances [32], [33], [34], [35], [36], [37], [38], [39]. Under the aegis of the recent investigations, those who are connoisseurs of artificial intelligence countenance the idea that the stacking model approach has been prosperously applied to different fields of engineering sciences. For mine ventilation applications, the stacking model is called a new soft computing model, which can extract an explicit mathematical form along with input and output variables to gain better physical insight into the processes involved. For example, the stacking model can be used in algorithms for electronic devices for tunnel ventilation [40], [41], predicting energy consumption [42], lake surface water temperature [43], compressive strength of rice husk ash concrete [44], emergency cases of patients with heart disease [45] or during the detection of leakages in water distribution systems [46].

To the best of the authors' knowledge, this is the first time applying the staking model to present formulations for evaluating friction factor α. Moreover, the staking model is capable of presenting the simplest nonlinear equations between input and output parameters, releasing an acceptable physical insight into problems. Having many data sets provided from field measurements of mine tunnels or roadways in China's mining sector, the stacking model can be evaluated and then used for the prediction of friction factor α. Thoroughly, the most crucial aim of this study is an attempt to achieve the accuracy performances of the improved staking model for predicting friction factor α. The improved staking model result of performances will be compared with those obtained using the Principal Components Analysis and Back Propagation (PCA-BP), Random Forest (RF), GA-Projection Pursuit Regression (GA-PPR) and light gradient-boosting machine (LGBM) models because they can also create a reliable relationship employed as an input-output system. Eventually, traditional methods will be validated for used data series in further comparisons.

The framework developed integrates feature importance analysis, an enhanced stacked learning model, cross-validation, performance evaluation, model ablation study, and prediction integration to create a powerful model for determining mine airway resistance. Practical application includes feature simplification, the utilization of diverse algorithms, performance validation through cross-validation, and ongoing model improvement for effective deployment in mining engineering.

The purpose of this study is to address the challenges associated with the measurement of ventilation friction resistance in mines, such as the need for a significant number of personnel, prolonged measurement times, and the complexity of the tasks involved. The goal is to research and develop novel methods for predicting the friction factor, providing valuable insights for mine personnel. The research contribution of this paper lies in the analysis and prediction of the friction factor based on factors influencing it. Leveraging the advantages of stacking predictive models, the study aims to learn from the strengths and weaknesses of individual models to make more accurate predictions. Furthermore, it seeks to enhance the algorithm and methodology of stacking predictive models. Using real measured data from an actual mine in China as the model's sample, the research aims to compare and select optimized models for predicting the friction factor. This, in turn, offers a novel approach for obtaining accurate predictions of the friction factor. The practical application value lies in addressing the significant demand for the informationization and intelligent upgrading of mine ventilation. Mine ventilation network calculation provides rational and effective design and optimization solutions for mine ventilation systems. The accuracy of the calculation results directly impacts the safety and working environment of miners. The mine ventilation friction factor of a mine airway, as a key parameter in network calculations, is crucial. Research on machine learning models for the rapid and accurate prediction of frictional airflow resistance forms the basis for real-time calculation of the ventilation network and precise adjustment of air flow rates. This effectively overcomes the challenges of mine ventilation being in a manual or semi-manual stage and the difficulty in meeting the requirements of intelligent mine construction.

This investigation focuses on creating improvement traditional stacking predictive models. Utilizing measured data, a new predictive model for the friction factor of a mine airway is established. The paper analyzes the advantages of the new model in terms of predictive accuracy. The rest of this paper is organized as follows. Section II introduces the current research on predicting the friction factor through machine learning. Section III, improved the prediction model of the traditional stacking algorithm using XGBoost, BOX-COS, and SVR to enhance prediction accuracy. In Section IV, discuss the factors influencing the friction factor of mine airways. Analyze the relationships between these factors to provide learning data for constructing the new predictive model in the second section. In Section V, to showcase the enhanced accuracy of the improved stacking model, comparisons were made against individual base models, the application of boosting methods on individual base models, and other conventional predictive models. In Section VI, the corresponding conclusions are given.

II. RELATED WORKS

The ventilation friction factor prediction has always been a significant issue in mine ventilation safety management. One method to obtain the ventilation friction factor through machine learning is by solving the Colebrook–White equation. In traditional calculation methods, due to the implicit nature of the Colebrook–White equation, calculating the friction factor requires iterative solutions. This process becomes more time-consuming, especially when evaluating large mine networks multiple times. To address this issue, many scholars have employed explicit approximation methods for the Colebrook–White correlation, improving the prediction accuracy of the friction factor through various machine learning techniques. Past studies have utilized multiple machine learning models for predicting mine ventilation friction factor, such as Adaptive Neuro-Fuzzy Inference System (ANFIS), Evolutionary Polynomial Regression (EPR), and Model Tree (MT) [47], [48], [49], [50], [51]. Salmasi applied Artificial Neural Network (ANN) and Gene Expression Programming (GEP) to predict the friction factor, finding that the ANN model had lower accuracy than the GP model [50]. Similarly, Giustolisi and Savic developed the EPR model for the Chezy resistance coefficient and compared it with the rule-based genetic programming, finding the EPR model to be more accurate [48]. Recent studies have focused on developing explicit procedures for predicting friction factors in mine ventilation using AI techniques to overcome the time-consuming iterative solutions. Models based on ANN and Support Vector Regression (SVR) have been developed and validated against experimental data from various sources, including Nikuradse [52], [53], Princeton [54], [55], [56], and Oregon [57]. SVR, with its unique, optimal, and global solution for the quadratic programming problem, has advantages over traditional techniques like Multiple Linear Regression (MLR) and ANN [58], [59]. The Structural Risk Minimization (SRM) principle in SVR provides good generalization performance [60]. Fifteen well-established explicit correlations from the literature have been chosen to evaluate friction factors in turbulent flows in smooth and rough pipes. Predictions from these correlations were compared with experimental results and statistically with developed AI-based friction factor models [61]. Another approach to predicting the ventilation friction factor is using the support type of the roadway for prediction. A study employed an improved genetic algorithm (GA) to solve the inverse problem of ventilation resistance coefficients, enhancing the algorithm's global and local search capabilities [62]. In a different line of research, a hybrid prediction model for the local resistance coefficient of water transmission tunnel maintenance ventilation has been proposed based on machine learning [63]. The hybrid model introduced the hybrid kernel into a relevance vector machine to build the Hybrid Kernel Relevance Vector Machine model (HKRVM). An improved Artificial Jellyfish Search algorithm (IAJS) was applied for the kernel parameter optimization of the HKRVM model.

However, despite significant progress in predicting ventilation friction factors using the mentioned machine learning techniques, there are still some limitations. One major challenge is the potential limitation of these single models in handling complex nonlinear relationships, unable to fully capture the complex characteristics of friction factors. Additionally, models that perform well on specific datasets may have poor generalization abilities on different datasets, struggling to adapt to new data. To overcome these limitations, recent research has increasingly focused on the application of stacking algorithms in predicting ventilation friction factors. Stacking algorithms, as ensemble learning methods, combine predictions from multiple base models to improve overall performance. In predicting ventilation friction factors, stacking algorithms construct a meta-model that integrates outputs from various machine learning models, learning and predicting friction factors at a higher abstract level. Some studies have explored using stacking algorithms such as Stacked Autoencoder (SAE) or Stacked Long Short-Term Memory networks (Stacked LSTM) to address the prediction problem [64], [65], [66]. These methods, by introducing more layers and complexity, allow the model to better capture the nonlinear relationships and dynamic changes in ventilation systems. Empirical studies indicate that, compared to single models, these stacking algorithms exhibit higher prediction accuracy and robustness in predicting ventilation friction factors. Therefore, while traditional machine learning techniques have made some achievements in predicting ventilation friction factors, the introduction of stacking algorithms emphasizes improving the overall performance of models to better address the complexity and diversity of ventilation systems. This is expected to provide new perspectives and methods for research in predicting ventilation friction factors and contribute to developing intelligent mine ventilation networks.

III. EVOLUTION OF STACKING METHOD

Creating upon the traditional stacking algorithm, we have introduced enhancements. This section outlines the sub-prediction models within the stacking algorithm and the improvements made. The initial data is pre-processed using the XGBoost algorithm, and the intermediate predicted data is transformed using the Cox-Box method to achieve a bias towards a normal distribution. Prediction errors are forecasted through SVR, and the results of the error prediction are combined with the outcomes of the meta model to enhance the accuracy of predictions.

A. CHARACTERISTIC OF THE METHOD

Stacking is an ensemble learning technique that involves training multiple models and combining their outputs to improve prediction performance. It works by using the predictions of lower-level models as input to a higher-level model, enabling the ensemble to learn from the diverse range of models and improve its overall accuracy [67]. Stacking involves a two-layer learning structure, where the first layer consists of base models and the second layer consists of a meta-model that uses the outputs of the base models as input to make the final predictions. The core concept of the stacking model is to use a set of base learners to train on the original data and then generate a new data set from the output of each base learner. Subsequently, the new data set is used to train the second layer of meta-learners, which generates the final predictions. Figure 1 depicts the framework of the stacking ensemble strategy. Figure 2 illustrates the process flow of the stacking ensemble strategy algorithm [68]. In the stacking learning algorithm, the meta-learner's training set is generated by utilizing the outputs of the base learner. Using the output of base learners as the new training set for meta-learning can result in overfitting. To address this issue, the stacking framework employs k-fold cross-validation to generate new training samples for the meta-learner from the unused training samples of the base learner, thus preventing overfitting. Assume that the original training set is denoted as $A = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$. A is initially partitioned into k equal-sized subsets, denoted as $A_1, A_2, \ldots, A_k$. Ai represents the training set for the i-th fold, and $A_i{}'$ represents the test set for the i-th fold. The partitioning is done randomly to ensure that each subset has a similar distribution of data. Assuming there are G base learners, the base learners are obtained by training on subset A_j. The base learner's output for each sample x_i in $A_i{}'$ can be represented as z_{ig}. The new training set generated by the G base learners is $A' = \{(z_i, y_i)\}$, where $z_i = (z_{i1}; z_{i2}; \ldots; z_{iG})$ represents the combination of the predictions made by each of the G base learners for a particular input sample x_i.

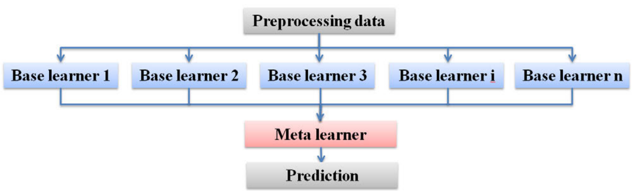


FIGURE 1. Stacking model framework.

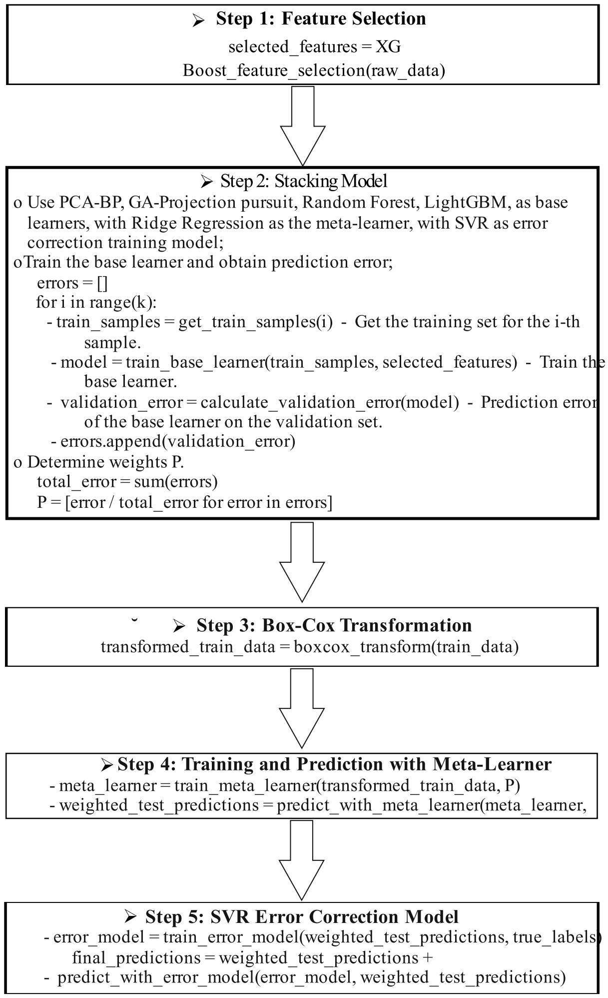


FIGURE 2. Calculation steps of improved stacking model [48].

B. LEARNERS

To build the stacking model, it is important to carefully select the base learners based on their accuracy and variability. The study exclusively used ensemble learning models as base learners. As a result, the top-performing ensemble learning models (PCA-BP; GA-PPR; RF; LGBM; RR; XG Boost algorithm; Box-Cox; SVR) were chosen for the study. To select the best base learner, one model with the highest performance was chosen among series algorithms, considering their differentiation. The reason for this is that using base learners with diverse characteristics can explore the relationship between input and output features from multiple perspectives. Below are brief descriptions of the four selected base learner models. The meta-learner typically selects a model that has fitting capabilities. Therefore, the stacking model including two layers was chosen as the meta-learner for this study.

1) BASCI MODEL

a: PCA-BP ALGORITHM

The PCA-BP algorithm is a hybrid machine learning algorithm that combines Principal Component Analysis (PCA) with Backpropagation (BP) neural networks. It involves using PCA to reduce the dimensionality of the input data and then feeding the reduced data into a BP neural network for training and prediction. This algorithm is capable of extracting the most relevant features from the data, thereby improving performance and accelerating training convergence speed.

PCA-BP prediction can be decomposed into two steps of sub-models. First, the basic principle of PCA involves the analysis of existing variable data to extract multiple sets of new data with significant information, making it an effective mathematical method for dimensionality reduction in multivariate statistics. PCA transforms an n-dimensional random vector composed of n initial variables into d new comprehensive variables through linear transformation, forming d (where $d < n$) final principal component factors. This achieves the dimensionality reduction of the initial features. In addition, a well-performing BP neural network requires input sample data, and the input of sampled data can sometimes affect the accuracy of the model. Combining the first step of principal component analysis can effectively address

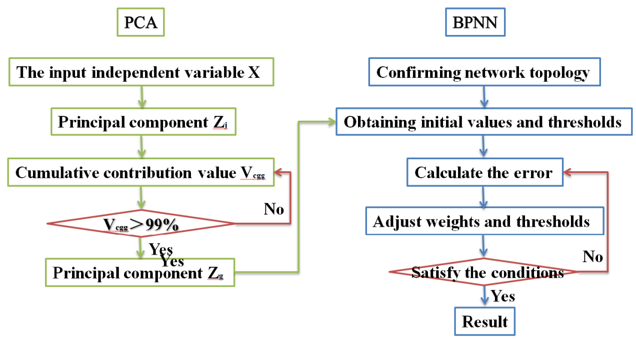


FIGURE 3. PCA-BP neural network modeling flowchart.

these issues, with the core content being feature extraction. Feature extraction involves finding the most relevant information to the problem from numerous data. What is truly needed at this point is a model that contains only the relevant inputs or features that lead to the fewest problems with free parameters. These features can generate the best model with sufficient generalization ability for a given dataset. The implementation method of feature extraction is data dimensionality reduction. Dimensionality reduction is a preprocessing method for high-dimensional feature data. It involves preserving some of the most important features of high-dimensional data, eliminating noise and unimportant features, and achieving the goal of improving data processing speed. In practical production and application, dimensionality reduction, within a certain range of information loss, can save time and cost. Dimensionality reduction can be applied in a very broad range of data preprocessing fields. The workflow diagram of the PCA-BP network is shown in Figure 3. More information on the PCA-BP model can be found in related articles [69], [70], [71].

b: GA-PPR ALGORITHM

GA-PPR is a novel artificial intelligence algorithm that combines genetic algorithms and projection pursuit techniques. PGT is designed to identify the most important features of complex data sets and trace their interactions by optimizing a projection index. The optimization process involves selecting a subset of features and projecting them onto a subspace using a projection index that maximizes the separability of the data points in the subspace. More information on the GA-PPR model can be found in related articles [72], [73], [74].

c: RF ALGORITHM

Random Forest (RF) is a widely used ensemble prediction algorithm. It is based on the ensemble method of DT to improve the overall predictive ability of the model. RF constructs multiple DT and combines them. It adopts the bootstrap sampling method to randomly sample the training data with replacement. Each DT is trained independently, and the final prediction result is obtained by voting or averaging. RF algorithm has good generalization ability, resistance to overfitting, and feature selection capability. It is widely applied in tasks such as classification, regression, and feature importance evaluation. More information on the RF model can be found in related articles [75], [76], [77].

d: LIGHT GRADIENT-BOOSTING MACHINE ALGORITHM

LGBM is an efficient gradient boosting tree algorithm known for its speed and accuracy. It uses histogram-based DT algorithms and discretization techniques to improve training speed and memory efficiency. LGBM performs exceptionally well in handling large-scale data sets and high-dimensional sparse data, demonstrating high accuracy and scalability. It is widely applied in various machine learning tasks such as classification, regression, ranking, and recommendation. More information on the LGBM model can be found in related article [78], [79], [80].

2) META MODEL

Ridge Regression (RR) was used to obtain the meta model. RR is a commonly used linear regression algorithm that solves the multicollinearity problem by introducing a parameter regularization term. In RR, we define the objective function as the sum of the original loss function and the regularization term, and we seek the optimal model parameters by minimizing the objective function. Compared to ordinary linear regression, RR can improve the model's generalization ability to some extent and is suitable for regression problems with high-dimensional data. More information on the RR model can be found in related articles [81], [82], [83].

3) IMPLEMENTATION USING THREE OPTIMIZATION MODELS

a: XG BOOST ALGORITHM

XGBoost is a machine learning algorithm based on Decision Trees (DT) ensembles, which uses gradient boosting technique to improve the performance and accuracy of the DT model. It is characterized by its efficiency, scalability, portability and accuracy, and has been widely used in areas such as data mining, machine learning and statistical modeling. XGBoost accelerates the training process of DT through parallel processing and pre-sorting techniques, and performs well on large-scale data sets. At the same time, it can handle various types of data, including numerical, categorical and textual data, and supports multiple loss functions and evaluation metrics. More information on the XG Boost model can be found in related articles [84], [85], [86].

b: BOX-COS ALGORITHM

Ridge Regression (RR) is a linear model, and the assumption for linear modeling is the presence of a linear relationship in the data. The Box-Cox transformation aids in making the data conform more closely to this linear assumption. Moreover, when there is heteroscedasticity in the original data, the Box-Cox transformation helps reduce heteroscedasticity, consequently improving the model's performance. Therefore,

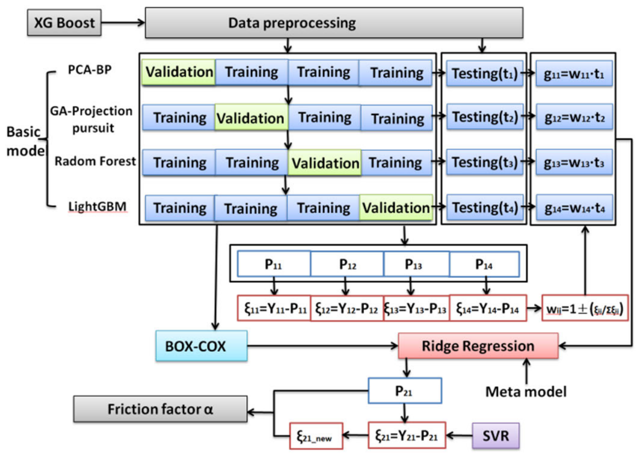


FIGURE 4. Improved stacking model for prediction the friction factor.

the results after Box-Cox processing contribute to better integration with Ridge Regression.

In stacking model, the base learning layer of the model, different learners were selected for stacking with the aim of comprehensively capturing the complex relationships of the ventilation friction factor by leveraging the strengths of multiple models. However, due to the different assumptions each model makes about the distribution of the data, it resulted in the data exhibiting skewed distributions. In the regression prediction problem of the meta-learner, assuming normality is crucial, and Box-Cox transformation is employed to make the data more conformant to a normal distribution, thereby enhancing the overall performance of the model. In other words, the transformation to a normal distribution is carried out to ensure that the assumptions of the meta-learner, especially the Ridge Regression model's assumption of normality, are met. More information on the Box-Cox model can be found in related articles [87], [88], [89].

c: SVR ALGORITHM

Support Vector Regression (SVR) is a powerful regression algorithm. Unlike traditional regression methods, SVR formulates the regression problem as an optimization problem to find the optimal hyperplane that best fits the data. It achieves this by using kernel functions to handle non-linear problems and map the data into a high-dimensional space, allowing for better capture of complex relationships among the data. More information on the XG Boost model can be found in related articles [90], [91], [92].

C. IMPROVED STACKING MODEL

Based on the XG Boost algorithm, weighted base models, and Box-Cox transformation, as well as SVR error correction, this study constructed a prediction model for friction factor α. Figure 4 shows the algorithm of using the improved stacking model.

The process is described as follows:

(1) Use the XGboost algorithm to perform feature selection on the original data of the friction factor, removing redundant features with low correlation. The remaining data is then input into an improved stacking model.

(2) In the stacking model, use PCA-BP, GA-PPR, RF, LGBM as the base learners, with RR serving as the meta-learner. For each base learner, train it with different samples and based on the prediction errors ($\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$) on the validation set, determine the weights (P_{11}, P_{12}, P_{13}, P_{14}) by calculating their proportional contribution to the total error. Then, weigh the output results of the first layer as the input for the second layer.

(3) Apply the Box-Cox transformation to the training set of the first layer to enhance its normality and predictability.

(4) Train the meta-learner using the transformed training set from the Box-Cox transformation. After training is complete, use this meta-learner to predict the weighted test set and obtain the final prediction results.

(5) After the initial predictions using the above models, train an SVR model using the error data from the initial predictions. This will provide an error distribution model. Combine the initial prediction results with the error predictions to obtain the final prediction results. The SVR error correction model effectively compensates for the errors inherent in the initial prediction model, further improving the accuracy of the predictions.

IV. PREDICTIVE MODELING OF FRICTION FACTORα

In this section, the influential factors of the friction factor of mine airways will be discussed. The interrelations among these factors will be analyzed, and three evaluation indicators will be established to prepare for subsequent prediction assessment.

A. DATASET

The data set used in this study comes from the measured data of mining areas in China, including 141 sets of data on factors affecting friction factor α. This dataset has been used in previous research study [93]. The data includes the proportion of different underground tunnel support types: wooden column support constitutes 24%, anchor spray support constitutes 27%, metal beam support constitutes 24%, shotcrete support constitutes 25%. Due to the distinct material requirements for different support forms, the wooden column support type data does not include information on metal beam thickness and column thickness. The anchor spray support type does not include information on diameter of wooden column and metal beam thickness. The metal beam support type does not include information on diameter of wooden column, longitudinal diameter, metal beam thickness and column thickness. The shotcrete support type does not include information on diameter of wooden column, longitudinal diameter, shotcrete part of the perimeter, metal beam thickness, column thickness and shed spacing. The adopted data has been used in previous studies for predicting the friction factor of a mine airway

TABLE 1. Statistical analysis using factors affecting friction factor.

Variable	Symbol	Unit	Mean	Minimum	Maximum	Standard deviation
diameter of wooden column	X_1	cm	19.75	15.12	26.12	3.68
longitudinal diameter	X_2	-	4.53	1.11	8.11	2.11
roadway cross-section area	X_3	m^2	6.42	3.01	19.58	3.14
roadway perimeter	X_4	m	10.07	7.21	16.99	1.92
shotcrete part of the perimeter	X_5	m	7.15	5.43	9.91	1.41
metal beam thickness	X_6	cm	13.00	10.00	18.00	2.49
column thickness	X_7	cm	44.00	40.00	50.00	4.47
shed spacing	X_8	m	0.61	0.27	1.53	0.35
roadway height	X_9	m	3.31	2.25	4.21	0.35
roadway width	X_{10}	m	4.31	2.74	5.62	0.64

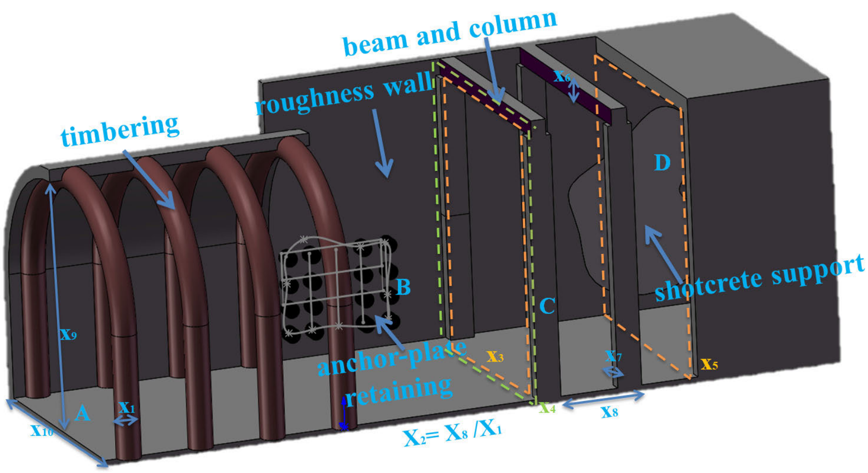


FIGURE 5. Factors affecting the friction factor of an airway.

using random forest prediction models, enhancing its authenticity in this research. In the new model development, these data are no longer classified by support type. Collectively, they constitute the training and testing sets for machine learning. Statistical analysis was conducted on the measured data of factors affecting the friction factor of a mine airway with different support types. The descriptive statistics, including mean, maximum, minimum, maximum, and variance, are presented in Table 1.

B. DATA PREPROCESSING

As described in Section I, the main factors affecting the value of coefficient α are the size and shape of the cross-section of mine airway. The collected data for this study concern variable roadway parameters with a rectangular shape and secured with mainly support frames consisting of props and metal beams, but also with additional prop-type support in the center of the frame. The input characteristic variables are diameter of wooden column, longitudinal diameter, roadway cross-section area, roadway perimeter, shotcrete part of the perimeter, metal beam thickness, column thickness, shed spacing, roadway height, roadway width. The output variable is the friction factor.

Figure 5 presents the visualization of these feature variables. As shown in Figure 5, the tested mine airway has four different supporting methods, represented by A for wooden column support, B for anchor spray support, C for metal beam support, and D for shotcrete support.

The Figure 5 indicates the influencing factors X_1-X_{10}, where X_1 is the diameter of the wooden columns, X_2 is the longitudinal diameter (the ratio of the distance between adjacent support centrelines to the diameter or thickness of the supports), X_3 is the cross-section area, X_4 is the roadway perimeter, X_5 is the shotcrete part of the perimeter, X_6 is the metal beam thickness, X_7 is the column thickness, X_8 is the shed spacing, X_9 is the roadway height, and X_{10} is the roadway width.

Table 1 shows a brief statistical analysis of each characteristic variable. Correlations between variables were analyzed to assess the suitability of each characteristic variable. The Pearson correlation coefficient was used to examine the characteristic variables' association. Figure 6 displays the correlation matrix based on Pearson's coefficient. A coefficient closer to 1 indicates a stronger positive correlation between two variables. A coefficient closer to -1 indicates a stronger negative correlation. Figure 6 shows that there are correlations between the output variable and each input variable. All the features in the data set are numeric, and they vary in magnitude and scale. The numerical features were transformed into standardized values to eliminate differences in scales and magnitudes.

C. THE OPTIMAL HYPERPARAMETER

Grid search is an automated method for finding optimal hyperparameters. It systematically explores all possible combinations of hyperparameter values, training and evaluating models for each combination, and ultimately selecting the best-performing set. By eliminating the need for manual trial and error, grid search streamlines the process of hyperparameter tuning. The user defines the ranges and steps for each hyperparameter beforehand, and grid search automatically conducts a comprehensive search and evaluation.

In the context of ensemble algorithms, grid search is a powerful tool for systematically exploring multiple hyperparameter combinations to optimize model performance. This process involves creating a predefined grid of hyperparameter values, attempting each combination through training and cross-validation, and assessing the model's performance. For our stacked algorithm, consisting of PCA-BP, GA-PPR, RF and LGBM, grid search allowed us to finely tune the hyperparameters for each base learner.

The specific hyperparameter values optimized through grid search for PCA-BP include a principal component percentage of 90%, learning rate of 0.15, 17 hidden layer neurons, and a Sigmoid activation function.

Similarly, GA-PPR's optimized hyperparameters involve a population size of 35, a crossover probability of 0.65, a mutation probability of 0.03, 23 principal components, and a truncation value of 0.15.

For RF, grid search focused on parameters such as the number of trees 154, maximum depth 7, quality measure function gini, minimum samples per tree 3, minimum samples per leaf 2, feature consideration per split lg2, and bootstrap sample usage True.

LGBM's optimized hyperparameters include maximum leaf nodes per tree 31, learning rate 0.16, maximum tree depth 5, minimum samples per leaf 5, sample proportion for tree training 0.6, random feature sampling proportion 0.6, L_1 regularization weight 0.2, L_2 regularization weight 0.3, and defining the learning task and corresponding loss function as regression.

In summary, grid search efficiently optimized our model, ensuring it excelled in various tasks. This systematic tuning process guaranteed that our stacked algorithm leveraged the strengths of each base learner, resulting in a robust and powerful ensemble model.

D. MODELING BUILDING AND EVALUATION

The data set was divided into a training set and a test set prior to the development of the model. The model was trained using the training set, and the performance of the trained model was evaluated using the test set. This study used 75% of the data in the data set as the training set and the remaining 25% as the test set.

The hyperparameters have a critical impact on the model's performance. Optimization can be used to obtain the best-performing hyperparameters of a machine learning model on a given data set, leading to improved model performance. This study used a combination of grid search and 4-fold cross-validation to optimize the hyperparameters of the machine learning model. The grid search method involves trying out different combinations of hyperparameters and training the model for each combination, making it a thorough approach. The combination of hyperparameters that achieves the best performance among n training sessions is considered the optimal one. The 4-fold cross-validation method initially splits the data into 4 mutually exclusive subsets. The method selects three subsets of data to form the training set without overlap, and uses the remaining subset as the test set. This selection process was repeated 4 times to obtain 4 different combinations. The average of the 4 combinations is taken as the final result. In this research, the combination of grid search and 4-fold cross-validation methods was used to evaluate all possible hyperparameter combinations, which can significantly improve the model's generalization ability [74]. This research used the root mean square error (RMSE), correlation coefficient (R^2), and mean absolute error (MAE) as the evaluation metrics to assess the performance of the models. RMSE is highly responsive to the largest or smallest discrepancies between the predicted and observed values. The higher MAE implies greater predictive accuracy. R^2 coefficient quantifies the strength and direction of the linear relationship between variables. A higher R^2 indicates a stronger linear correlation between variables and a closer fit between the predicted and observed values. MAE is a commonly used metric to evaluate the goodness of fit of hydrological model simulations, with a range of values greater than or equal to 0. A lower MAE indicates better model performance. A value of MAE closer to 0 indicates that the model prediction is similar to the actual value of the observed data. If the MAE is significantly less than 0, it indicates that the model's performance is unreliable. The formulas used to calculate each performance indicator are given below:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i'\right)^2} \tag{3}$$

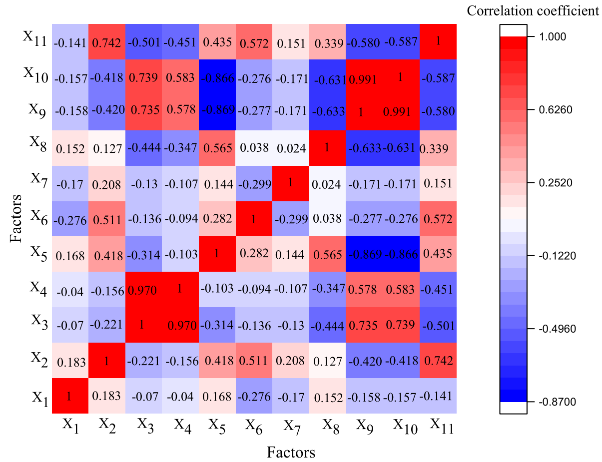


FIGURE 6. Pearson coefficient for the characteristic variables.

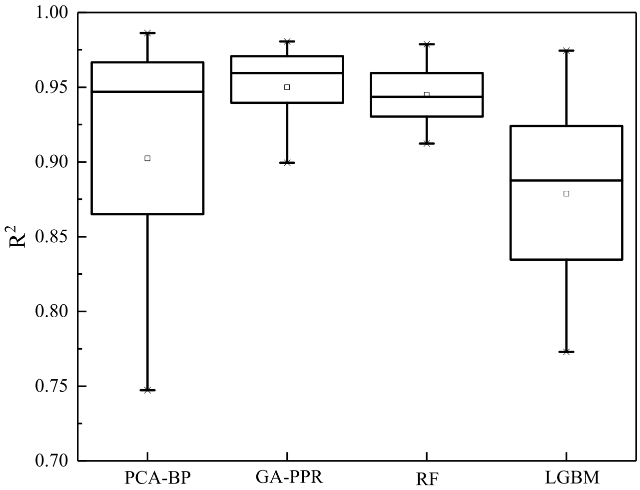


FIGURE 7. Box plot of basic learners' results.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - y_i'\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)^2} \tag{4}$$

$$MAE = \frac{\sum_{i=1}^{n}\left(y_i - y_i'\right)}{n} \tag{5}$$

where n is the number of samples in the data set, y_i and y_i' are the actual and predicted values of each sample.

V. RESULTS AND DISCUSSION

In the Results and Discussion Section, the improved stacking model is compared to individual base models, the application of the enhancement method to individual base models, and other traditional prediction models.

A. BASE-LEARNER INDEPENDENT PREDICTION RESULTS

Given that the characteristics of the base learners impact the predictive performance of the final stacking model, this study initially assessed the performance of the base learners to ascertain their suitability. Employing the base learners to make individual predictions on the test data set. The machine learning model underwent hyperparameter optimization using a combined approach of grid search and 4-fold cross-validation. R^2 was chosen as the primary evaluation metric for optimization to identify the best-performing hyperparameter combination. The average of the 4-fold cross-validation results serves as the final assessment metric. Below are the optimized hyperparameter combinations for the four base learners.

Four models as the base learners for the stacking model were selected, considering the diverse base of learners. Figure 7 displays the box plots representing the 4-fold cross-validation results of the four base learners, PCA-BP, GA-PPR, RF, and LGBM, on the training data set. Subsequently, the forecast accuracy of the hyperparameter-optimized base learners was assessed on the test data set. Table 2 presents the prediction outcomes of the four base learners on the test set. Based on the results presented in

TABLE 2. Base learner prediction results on the test set.

Model	RMSE	MAE	R^2
PCA-BP	3.121	0.288	0.984
GA-PPR	2.514	0.213	0.981
RF	2.337	0.189	0.979
LGBM	3.394	0.341	0.965

TABLE 3. Prediction results of improved stackingmodels with different combination of base learners.

Base Learner Combinations	RMSE	MAE	R^2
Improved stacking model	1.923	0.109	0.996
XGBoost+PCA-BP+SVR	2.996	0.249	0.984
XGBoost+ GA-PPR +SVR	2.519	0.268	0.982
XGBoost+RF+SVR	2.303	0.163	0.987
XGBoost+LGBM+SVR	2.212	0.157	0.983

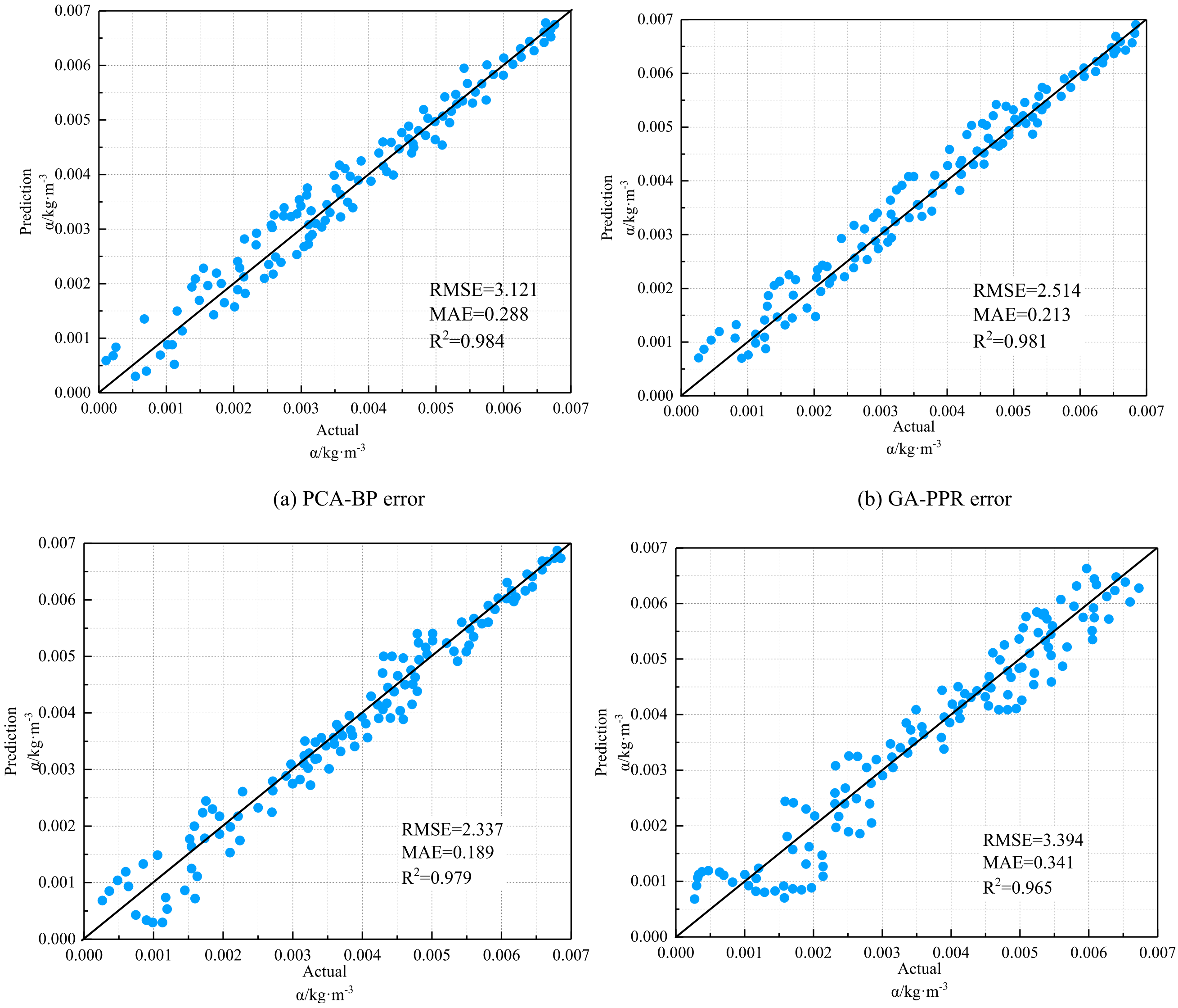
Table 2, the Randon Forest model demonstrates superior performance on the test set, while the LGBM model exhibits the lowest performance. Figure 8 visualizes the relationship between the predicted values and the actual values on the test data set. Considering that PCA-BP and GA-PPR are combination optimization algorithms, while LGBM and RF are ensemble-based algorithms.

B. PREDICTION RESULTS OF THE IMPROVED SRACKING MODEL

As outlined in Section II, the base learners selected for the improved stacking model were XGBoost and SVR, while the meta-learner was LR. Figure 9 presents the 4-fold cross-validation results of the stacking model on the training data set. The stacking model achieved the following prediction results of the friction factor on the test set: RMSE=1.923, MAE=0.109, and $R^2 = 0.996$. Figure 10 illustrates the relationship between the predicted and actual results of the stacking model on the tested data set. When comparing the assessment metrics of the stacking model with those of the four individual learner models (PCA-BP, GA-PPR, RF and LGBM), it can be observed that the stacking model outstands each of the base learners. This suggests that the meta-learner in the stacking model has the ability to rectify the predictions of the base learner for samples that were initially predicted incorrectly, resulting in enhanced overall prediction accuracy of the model.

C. THE IMPACT OF BASE LEARNERS ON THE PERFORMANCE OF IMPROVED STACKING MODELS

This Section investigates the impact of different base learners on the overall performance of the improved stacking model, specifically focusing on the suitability of the XGBoost and SVR models. To assess this, the base learners in the improved stacking model were modified, and the effect on model performance was evaluated. The LR model was retained



(a) PCA-BP error (b) GA-PPR error

(c) RF error (d) LGBM error

FIGURE 8. Box plot of basic learners' results.

as the meta-learner, and the improved stacking model was retrained accordingly. In accordance with the base learners mentioned in Section II, the XGBoost model was substituted with other prediction models, such as GA-PPR and LGBM. Subsequently, the improved stacking model was trained using these different combinations of base learners. The prediction results obtained from the improved stacking model with various base learner combinations are presented in Table 3. The improved stacking model with XGBoost and SVR models as base learners demonstrates superior characteristics. This can be attributed to the superior performance of the XGBoost model and SVR model compared to the none of them in individual assessment. Utilizing the XGBoost algorithm, the original data of influencing factors on the friction factor in mine ventilation were subjected to feature selection to eliminate features with low correlation. Furthermore, employing the enhanced model for prediction, the predicted error data was input into an SVR model for training, thereby obtaining an error distribution model. Combining the preliminary prediction results with the predicted error values yields the final prediction outcome. The SVR error correction model effectively compensates for errors inherent in the initial prediction model, further enhancing the accuracy of the predictions. The superior performance of the base learner model contributes to the enhanced predictive capability of the final improved stacking model.

D. COMPARSION WITH OTHER ENSEMBLE STRATEGIES

This Section aims to compare improved stacking with other ensemble strategies, as different strategies can significantly

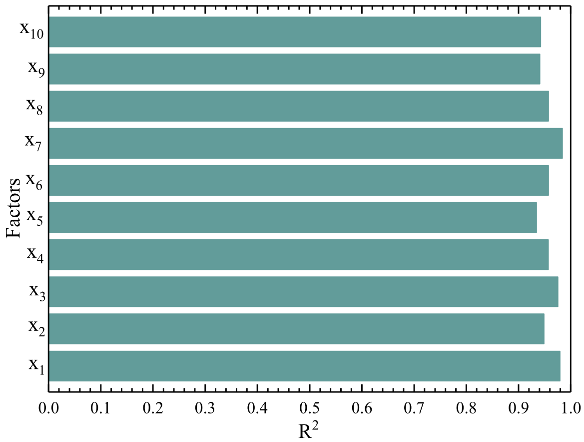


FIGURE 9. Improved stacking model on the training data.

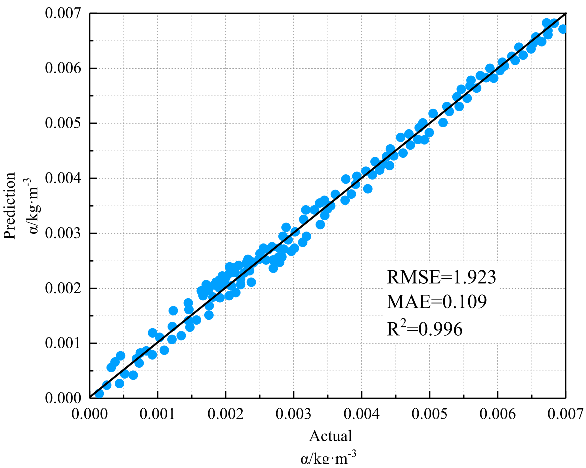


FIGURE 10. Improved stacking model on the test data.

impact the model's performance. The two most commonly used ensemble strategies in regression problems are simple averaging and weighted averaging. These methods involve averaging the predictions of the base learners to obtain the final projected outcomes. In the simple averaging method, equal weights are assigned to the predictions of all base learners. On the other hand, the weighted averaging method assigns different weights to each base learner based on their prediction performance. The RR and PCA-BP models as the base learners selected for both averaging methods. Table 4 presents the prediction results obtained from different ensemble strategies. It is evident that the stacking model exhibits superior properties compared to the averaging methods. This can be attributed to the introduction of a meta-learner in the stacking model. The meta-learner has the ability to correct mistakes made by the base learners, thereby improving the overall prediction accuracy of the model.

E. COMPARSION BETWEEN STATE-OF-THE-ART MODELS AND PREVIOUS WORKS

To establish the superiority of the stacking model, a comparison was conducted with other popular models. The evaluation

TABLE 4. Prediction results of friction factor according to different ensemble strategies.

Ensemble strategy	RMSE	MAE	R^2
Improved stacking model	1.923	0.109	0.996
Simple averaging	2.736	0.264	0.954
Weighted averaging	3.452	0.289	0.937

TABLE 5. Comparsion with mainstream machine learning models.

Model	RMSE	MAE	R^2
Improved staking model	1.923	0.109	0.996
Bagging	3.172	0.188	0.983
Boosting	2.672	0.183	0.976
Extra Tree	3.384	0.197	0.972
Veting	2.365	0.198	0.979

was based on three metrics: RMSE, MAE, and R^2. Inspired by [94], [95], [96], and [97], the data was normalized, and an 80% or 20% split was used for training and testing, respectively. Grid search combined with cross-validation was employed to optimize the hyperparameters for each model. The prediction results of each model on the test set are presented in Table 5. The stacking model exhibited impressive performance. A comparison with the other model revealed the most substantial decrease in RMSE and MAE by 76% and 81% respectively, along with the most 1.94% increase in R^2. Ensemble learning is a machine learning approach that combines the outputs of multiple learners to achieve stronger and more robust performance than individual learners. Bagging models, Boosting models, Extra Trees models, and Voting models are considered advanced ensemble learning models. They have demonstrated outstanding performance in various scenarios and achieved significant success in practical applications. The Bagging model constructs multiple base learners through random sampling and aggregates their predictions or averages them to obtain a robust and high-performance model. Boosting models, on the other hand, sequentially train a series of weak learners, with each focusing on correcting the errors of the previous learner, gradually enhancing the overall model performance. The Extra Trees model introduces additional randomness when constructing decision trees. As a method for integrating the opinions of different learners, the Voting model combines their predictions through voting or averaging to form the final prediction. In comparison with advanced Bagging models, our ensemble model demonstrated a reduction of 65% and 73% in RMSE and MAE, respectively, while increasing the R^2 score by 1.21%. Compared to advanced Boosting models, our ensemble model showed reductions of 39% and 68% in RMSE and MAE, respectively, along with an improvement in R^2 by 1.94%. In comparison with advanced Extra Trees models, the RMSE and MAE decreased by 76% and 81%, respectively, while the

TABLE 6. Comparsion with previous related works.

Model	RMSE	MAE	R^2
Improved staking model	2.014	0.124	0.995
GEP [62]	2.756	0.237	0.913
MT [98]	3.756	0.274	0.934

R^2 score increased by 2.35%. Furthermore, when compared to advanced Voting models, our ensemble model exhibited reductions of 23% and 82% in RMSE and MAE, respectively, and an improvement in R^2 by 1.67%. The enhanced performance of the ensemble model can be attributed to training multiple base learners and leveraging their collective output.

Additionally, the established improved stacking model is compared with relevant studies in the field. In [62], an improved genetic algorithm (GA) is proposed to solve the ventilation resistance coefficient inversion problem. By developing the model, the inversion problem is transformed into a nonlinear optimization problem based on the least squares principle. When defining the objective function, measured pressure, calculated pressure deviation, measured flow, and calculated flow deviation are considered, along with constraints on node pressure, air flow, and ventilation resistance coefficient range. Reference [98] introduces a model tree (MT) to provide an evaluation formula for pipeline friction factors. Effective parameters influencing ventilation friction factors include pipe diameter, average flow velocity, kinematic viscosity, and height of internal surface roughness. The model is enhanced using two non-dimensional parameters, Reynolds number and relative roughness. Machine translation methods are trained and tested using data sets obtained from approximate solutions. In Table 6, compared to GEP models, our ensemble model showed reductions of 37% and 48% in RMSE and MAE, respectively, along with an improvement in R^2 by 8.98%. In comparison with advanced Extra Trees models, the RMSE and MAE decreased by 46% and 55%, respectively, while the R^2 score increased by 6.53%. The results indicate that the improved stacking model outperforms the previous model.

F. INPUT FEATURE IMPORTANCE ANALYSIS

In this Section, the XGBoost model is considered as the base learner to analyze the importance of variables. The XGBoost model demonstrated the best importance among the base learners. The importance of each feature in the data set within the XGBoost model was computed to determine feature importance. The feature importance was evaluated based on the improvement in performance measure for each feature's split point within an individual boosted tree. The weight assigned to a feature depended on the magnitude of improvement observed in the split point performance measure.

Additionally, the more frequently a feature was selected by the boosting tree, the higher its importance. The results

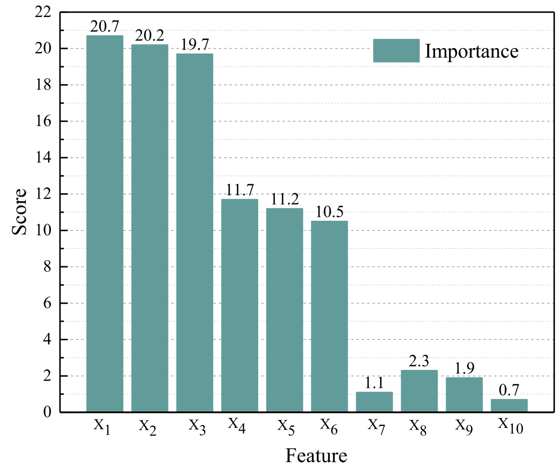


FIGURE 11. Importance of input feature.

from all boosting trees were aggregated and averaged, considering each feature's contribution to obtain the overall importance score. The XGBoost feature importance results are presented in Figure 11. Notably, the most crucial input feature variables were found to be the diameter of wooden columns, the longitudinal diameter, and the roadway cross-section area. In contrast, the roadway width and the roadway height of the tunnels were relatively less important. Among the top three important input features had the greatest influence on the friction factor. This is because the number of wooden columns is large, and rows of supporting obstacles generate additional air resistance in the airway, which has a huge obstacle effect on the airflow in the restricted space.

Using the XGBoost algorithm to rank the importance of input features is significant because it allows for the identification of the most crucial features for predicting the target variable. In this case, the features of wood column diameter, longitudinal diameter, and tunnel cross-sectional area are deemed more important. By selecting these three features, which exhibit higher correlation with the friction factor, through XGBoost, the model can be simplified and its interpretability improved. Furthermore, when dealing with large-scale data in subsequent predictions of new models, using fewer features can enhance computational efficiency. This is advantageous for both the training and inference stages of individual base learners in ensemble methods like stacking. Additionally, an excess of features may lead to overfitting of the model to the training data, and choosing features strongly correlated with the target can mitigate the risk of overfitting. In the context of practical data measurement, selecting these three features closely related to the actual application is beneficial. Notably, these features are easily measurable and obtainable in real-world scenarios. This facilitates the better application of the model's predictive results to practical engineering problems.

G. ANALYSIS OF ROBUSTNESS AND REPRODUCIBILITY

When selecting 4-fold cross-validation as the model evaluation method, various factors were thoroughly considered to ensure the chosen approach is both rational and effective. Firstly, 4-fold cross-validation is deemed an economically efficient choice in the context of our research background. Given the moderate size of our dataset, leave-one-out cross-validation might lead to prohibitively high computational costs, while k-fold cross-validation with a smaller k could overly influence performance evaluation due to the impact of a single split. Therefore, 4-fold cross-validation strikes a balance between computational efficiency and accurate assessment of model performance. Secondly, this study took into account that 4-fold cross-validation possesses characteristics that provide sufficiently reliable estimates of model performance. By dividing the dataset into four mutually exclusive subsets, we could evaluate the model's performance across multiple training/testing set combinations, thereby reducing the variance in evaluation results caused by specific splits. This contributes to enhancing our confidence in the model's generalization performance and ensures that the evaluations are relatively robust.

Although alternative methods exist, such as leave-one-out cross-validation, stratified k-fold cross-validation, and ShuffleSplit, our choice is based on a comprehensive consideration of data scale, computational resources, and evaluation precision. This study believes that 4-fold cross-validation can efficiently manage computational expenses while providing a reliable assessment of model performance, making it a suitable and feasible approach for our research.

4-fold cross-validation is a commonly used model evaluation method that divides the dataset into four equally sized, mutually exclusive subsets, with each subset aiming to maintain the distribution characteristics of the original dataset as closely as possible. Three subsets are then utilized as the training set to train the model, while the remaining one subset serves as the test set to evaluate the model's performance. This process is repeated four times, with each iteration selecting a different subset as the test set, and the final evaluation results are computed. 4-fold cross-validation is employed to assess the stability and reproducibility of the model. Through multiple repetitions of the training and testing process, the model's performance can be observed under different combinations of training and test sets, providing a better understanding of the model's generalization ability. The 4-fold cross-validation typically involves the following steps:

Step 1: Divide the dataset into four mutually exclusive subsets. The original dataset is partitioned into four equally sized, mutually exclusive subsets. Random partitioning is commonly used to ensure that each subset can represent the overall data distribution.

Step 2: Repeat the training and testing process. In each iteration, one of the four subsets is chosen as the test set, and the other three subsets are used as the training set. The model is trained with the training set and evaluated on the test set.

TABLE 7. 4-Fold cross validation method.

*i*th-fold	RMSE	MAE	R^2
1th-fold	1.965	0.118	0.987
2th-fold	2.006	0.120	0.991
3th-fold	1.968	0.114	0.983
4th-fold	2.018	0.123	0.981
Average	1.989	0.118	0.985

Step 3: Calculate performance metrics. In each iteration, the model's performance metrics on the test set need to be computed. The evaluation metrics in this model include RMSE, MAE, and R^2, which assess the predictive effectiveness of the model.

Step 4: Compute average performance metrics. After completing the four iterations, four performance metrics are obtained. Subsequently, the average of these metrics is calculated as the final evaluation result, offering a more accurate assessment of the model's performance on the entire dataset.

The 4-fold cross-validation of the stacked model on the training dataset is presented in Table 7. It depicts the RMSE, MAE, R^2, and other metrics under different folds in the 4-fold method. The mean values for RMSE, MAE, and R^2 are 1.989, 0.118, and 0.985, respectively. These evaluation metrics, when compared to those of a single base learner, demonstrate the superior performance of the results. For each fold, calculate the RMSE, MAE, and R^2 performance metrics of the model on that fold. These performance metrics show relatively consistent performance across all folds, indicating that the model is stable across different subsets of the data. Compute the average values of RMSE, MAE, and R^2 for the four-fold cross-validation. The model exhibits good performance on various data subsets, demonstrating its robustness across the entire dataset. This not only indicates the advantages brought about by ensemble learning relying on the complementarity of multiple models but also ensures the robustness and repeatability of the proposed improved ensemble model's prediction results.

H. THE ANALYSIS OF DIFFERENT COMBINATIONS OF BASE LEARNERS ON THE IMPROVED STACKING MODEL

Ablation study is a highly useful tool that aids in better understanding the internal mechanisms of integrated predictive models and provides valuable guidance for algorithms. In the field of machine learning, ablation study is commonly used to assess the impact of different features on model performance, determining which factors are essential for achieving good performance. During the ablation study, this paper systematically "removes" one or more predictive models from the ensemble of base learners and compares the performance of each altered configuration. This helps identify which predictive models contribute the most to improving the performance of the stacked prediction algorithm and which ones can be removed or improved to enhance overall performance.

TABLE 8. Ablating basic learner(s) from the improved stacking model.

left basic learner(s)	RMSE	MAE	R^2
PCA-BP +GA-PPR+ RF	2.221	0.158	0.993
PCA-BP+ GA-PPR+LGBM	2.101	0.151	0.986
PCA-BP+ RF+ LGBM	2.019	0.149	0.991
GA-PPR+ RF+ LGBM	2.196	0.145	0.989
RF+LGBM	2.198	0.157	0.988
GA-PPR+LGBM	2.199	0.153	0.985
GA-PPR+ RF	2.229	0.161	0.988
PCA-BP+LGBM	2.110	0.155	0.985
PCA-BP+ RF	2.298	0.159	0.988
PCA-BP+ GA-PPR	2.398	0.202	0.985

As shown in Table 8, the first column represents the remaining predictive models after removing one or more from the ensemble of base learners. The impact of the ablation results is characterized by RMSE, MAE, and R2. It can be observed that after removing one predictive model, the combination of PCA-BP + RF + LGBM exhibits the best performance, and after removing two predictive models, the combination of PCA-BP + RF + LGBM achieves the minimum root mean square error (RMSE), indicating superior performance. Analyzing the data reveals that the ablation of one learner's predictive indicators overall outperforms the ablation of two indicators' predictions. This suggests that retaining more learners in the model positively contributes to overall performance.

This situation indicates that different learners provide diverse information, and each learner has a unique contribution to the overall performance of the model. The effectiveness of the model lies in the complementarity between different learners; they can capture different patterns or features, thereby synergistically enhancing the model's generalization performance. It is evident that, for stacking algorithms, preserving more learners leads to better overall performance. This underscores the advantage of stacking algorithms in effectively combining multiple learners to achieve stronger generalization capabilities on complex problems.

VI. CONCLUSION

The predictive model of the friction factor (frictional air resistance coefficient α) in a mine airway was established based on the improved learning process. Important indicators with minimal impact were selected based on XGBoost scores. Weight allocation and BOX-COX transformation were employed to re-quantize the data of input meta-learners. Four heterogeneous algorithms were used as learners for predicting the friction factor in the improved stacked model. Experimental comparisons with traditional predictive models validate the superiority of the improved stacked model. The main results of the study are as follows:

(1) The XGBoost model assesses the significance of input feature variables and conducts a feature importance analysis based on the scores. By filtering out less important features, namely column thickness, shed spacing, roadway height and width, a prediction model for the friction factor is constructed with improved efficiency, accuracy, and reduced complexity.

(2) Through comparative research, it has been demonstrated that the improved stacked model proposed in this study outperforms traditional stacking models and other popular prediction models. The results indicate that the improved stacking model exhibits superior stability and higher prediction accuracy than individual base learners.

(3) By incorporating an error correction algorithm, the dynamic error of the prediction results from the fitted stacked model is calculated, leading to residual predictions. The final prediction value for the friction factor is obtained by combining the predicted values and residuals using an additive model fusion. The results indicate that incorporating residuals into the prediction results helps decrease prediction errors.

(4) The future research direction will focus on enhancing the performance of stacking algorithms in predicting the friction factor in mine ventilation network. There will be a particular emphasis on optimizing stacking model combinations, collecting datasets from different mines and ventilation periods and improving model interpretability. Specifically, we will focus on enhancing the predictive performance by incorporating the improved stacking model as a base learning model. Additionally, conducting on-site validation and comparing case studies will further propel the application of stacking algorithms in mining engineering. This will contribute to better adaptability to diverse mine conditions, ultimately enhancing the accuracy and operability of ventilation system design.

The achieved framework in this study can be used to determine mine airway resistances. It may also encourage researchers to use the improved stacked learning method to determine design factors that are difficult to verify experimentally.

REFERENCES

[1] X. Chang, J. Chai, Z. Liu, Y. Qin, and Z. Xu, "Comparison of ventilation methods used during tunnel construction," *Eng. Appl. Comput. Fluid Mech.*, vol. 14, no. 1, pp. 107–121, Jan. 2020.

[2] X. Chang, J. Chai, J. Luo, Y. Qin, Z. Xu, and J. Cao, "Tunnel ventilation during construction and diffusion of hazardous gases studied by numerical simulations," *Building Environ.*, vol. 177, Jun. 2020, Art. no. 106902.

[3] Y. Fang, J. Fan, B. Kenneally, and M. Mooney, "Air flow behavior and gas dispersion in the recirculation ventilation system of a twin-tunnel construction," *Tunnelling Underground Space Technol.*, vol. 58, pp. 30–39, Sep. 2016.

[4] Q. Liu, W. Nie, Y. Hua, S. Yin, L. Guo, H. Peng, H. Ma, and W. Zhou, "Investigation of efficient dust control strategy for construction tunnels: Ventilation System's implications for cleaner production," *Building Environ.*, vol. 180, Aug. 2020, Art. no. 107032.

[5] F. R. Mazarrón, C. Porras-Amores, and I. Cañas-Guerrero, "Annual evolution of the natural ventilation in an underground construction: Influence of the access tunnel and the ventilation chimney," *Tunnelling Underground Space Technol.*, vol. 49, pp. 188–198, Jun. 2015.

[6] Y. Xia, D. Yang, C. Hu, C. Wu, and J. Han, "Numerical simulation of ventilation and dust suppression system for open-type TBM tunneling work area," *Tunnelling Underground Space Technol.*, vol. 56, pp. 70–78, Jun. 2016.

[7] Y. Zhou, Y. Yang, R. Bu, F. Ma, and Y. Shen, "Effect of press-in ventilation technology on pollutant transport in a railway tunnel under construction," *J. Cleaner Prod.*, vol. 243, Jan. 2020, Art. no. 118590.

[8] M. Semin and L. Levin, "Mathematical modeling of air distribution in mines considering different ventilation modes," *Mathematics*, vol. 11, no. 4, p. 989, Feb. 2023.

[9] T. Long and S. Tong, *Fluid Mechanics*. Chongqing, China: Univ. Electronic Audiovisual Publishing House Co., 2018.

[10] L. F. Moody, "Friction factors for pipe flow," *Trans. Amer. Soc. Mech. Eng.*, vol. 66, no. 8, pp. 671–678, 1944.

[11] J. Nikuradse, "Laws of flow in rough pipe," *NACA Tech. Memorandum*, vol. 1292, pp. 1–63, Jan. 1933.

[12] C. F. Colebrook and C. M. White, "Experiments with fluid friction in roughened pipes," *Proc. Roy. Soc. London. Ser. A-Math. Phys. Sci.*, vol. 161, no. 906, pp. 367–381, 1937.

[13] C. F. Colebrook, "Turbulent flow in pipes, with particular reference to the transition region between the smooth and rough pipe laws," *J. Inst. Civil Eng.*, vol. 11, no. 4, pp. 133–156, Feb. 1939.

[14] J. Piggott, N. Revell, and T. Kurschat, "Taking the rough with the smooth—A new look at transmission factor formulae," in *Proc. PSIG Annu. Meeting*, Oct. 2002, Paper PSIG-0202.

[15] L. F. Moody, "An approximate formula for pipe friction factors," *Trans. ASME*, vol. 69, no. 12, pp. 1005–1011, 1947.

[16] S. W. Churchill, "Empirical expressions for the shear stress in turbulent flow in commercial pipe," *AIChE J.*, vol. 19, no. 2, pp. 375–376, Mar. 1973.

[17] B. Eck, *Technische Stromungslehre*. Berlin, Germany: Springer-Verlag, 2013.

[18] P. K. Swamee and A. K. Jain, "Explicit equations for pipe-flow problems," *J. Hydraul. Division*, vol. 102, no. 5, pp. 657–664, May 1976.

[19] G. F. Round, "An explicit approximation for the friction factor-Reynolds number relation for rough and smooth pipes," *Can. J. Chem. Eng.*, vol. 58, no. 1, pp. 122–123, Feb. 1980.

[20] D. Barr and C. White, "Solutions of the Colebrook-White function for resistance to uniform turbulent flow," *Proc. Inst. Civil Eng.*, vol. 71, no. 2, pp. 529–535, Jun. 1981.

[21] S. E. Haaland, "Simple and explicit formulas for the friction factor in turbulent pipe flow," *J. Fluids Eng.*, vol. 105, no. 1, pp. 89–90, Mar. 1983.

[22] G. Manadili, "Replace implicit equations with signomial functions," *Chem. Eng.*, vol. 104, no. 8, pp. 129–131, 1997.

[23] J. S. M. Li and W. K. Chow, "Numerical studies on performance evaluation of tunnel ventilation safety systems," *Tunnelling Underground Space Technol.*, vol. 18, no. 5, pp. 435–452, Nov. 2003.

[24] X. Tong, S.-W. Hong, and L. Zhao, "CFD modeling of airflow, thermal environment, and ammonia concentration distribution in a commercial manure-belt layer house with mixed ventilation systems," *Comput. Electron. Agricult.*, vol. 162, pp. 281–299, Jul. 2019.

[25] X. Fang, Z. Zhou, and D. Li, "Review of correlations of flow boiling heat transfer coefficients for carbon dioxide," *Int. J. Refrigeration*, vol. 36, no. 8, pp. 2017–2039, Dec. 2013.

[26] J. Kalenga and J. M. Akande, "Determination of airways resistance, volumetric efficiency and development of ventilation model of Rosh Pinah Zinc Mine Namibia," *J. Minerals Mater. Characterization Eng.*, vol. 9, no. 2, pp. 194–205, 2021.

[27] M. Bascompta, L. Sanmiquel, H. F. Anticoi, and J. Oliva, "Ventilation friction factor determination and comparison: Two case studies of potash mining," *J. Southern Afr. Inst. Mining Metall.*, vol. 119, no. 10, pp. 865–870, 2019.

[28] H. L. Hartman, J. M. Mutmansky, R. V. Ramani, and Y. J. Wang, *Mine Ventilation Air Conditioning*. New York, NY, USA: Wiley, 2012.

[29] M. J. McPherson, *Subsurface Ventilation and Environmental Engineering*. London, U.K.: Chapman & Hall, 1993.

[30] C. Montecinos and K. Wallace Jr., "Equivalent roughness for pressure drop calculations in mine ventilation," in *Proc. 13th United States/North Amer. Mine Ventilation Symp.*, Jun. 2010, pp. 225–230.

[31] H.-Y. Fu, Y.-Y. Zhao, H.-J. Ding, Y.-K. Rao, T. Yang, and M.-Z. Zhou, "A novel intelligent displacement prediction model of Karst tunnels," *Sci. Rep.*, vol. 12, no. 1, pp. 1–19, Oct. 2022.

[32] H. M. Azmathullah, M. C. Deo, and P. B. Deolalikar, "Neural networks for estimation of scour downstream of a ski-jump bucket," *J. Hydraulic Eng.*, vol. 131, no. 10, pp. 898–908, Oct. 2005.

[33] H. M. Azamathulla and A. A. Ghani, "ANFIS-based approach for predicting the scour depth at culvert outlets," *J. Pipeline Syst. Eng. Pract.*, vol. 2, no. 1, pp. 35–40, Feb. 2011.

[34] A. H. Alavi and A. H. Gandomi, "Prediction of principal ground-motion parameters using a hybrid method coupling artificial neural networks and simulated annealing," *Comput. Struct.*, vol. 89, nos. 23–24, pp. 2176–2194, Dec. 2011.

[35] H. M. Azamathulla and Z. Ahmad, "Estimation of critical velocity for slurry transport through pipeline using adaptive neuro-fuzzy interference system and gene-expression programming," *J. Pipeline Syst. Eng. Pract.*, vol. 4, no. 2, pp. 131–137, May 2013.

[36] L. Bonakdar and A. Etemad-Shahidi, "Predicting wave run-up on rubble-mound structures using M5 model tree," *Ocean Eng.*, vol. 38, no. 1, pp. 111–118, Jan. 2011.

[37] A. M. A. Sattar, "Gene expression models for prediction of dam breach parameters," *J. Hydroinformatics*, vol. 16, no. 3, pp. 550–571, May 2014.

[38] A. M. A. Sattar, "Gene expression models for the prediction of longitudinal dispersion coefficients in transitional and turbulent pipe flow," *J. Pipeline Syst. Eng. Pract.*, vol. 5, no. 1, pp. 1–11, Feb. 2014.

[39] M. Najafzadeh and H. Bonakdari, "Application of a neuro-fuzzy GMDH model for predicting the velocity at limit of deposition in storm sewers," *J. Pipeline Syst. Eng. Pract.*, vol. 8, no. 1, pp. 1–10, Feb. 2017.

[40] F. Wang, Y. Yang, T. Huang, and Y. Xu, "Lifetime prediction of electronic devices based on the P-stacking machine learning model," *Microelectron. Rel.*, vol. 146, Jul. 2023, Art. no. 115027.

[41] F. Akbari, S. Taghizadeh, D. Shvydka, N. N. Sperling, and E. I. Parsai, "Predicting electronic stopping powers using stacking ensemble machine learning method," *Nucl. Instrum. Methods Phys. Res. Sect. B: Beam Interact. with Mater. At.*, vol. 538, pp. 8–16, May 2023.

[42] J. García-Sanz-Calcedo, A. G. González, and D. R. Salgado, "Energy consumption in healthcare buildings in Spain," in *Proc. Int. Symp. Comput. Sci. Intell. Controls (ISCSIC)*, Oct. 2017, pp. 23–26.

[43] F. Di Nunno, S. Zhu, M. Ptak, M. Sojka, and F. Granata, "A stacked machine learning model for multi-step ahead prediction of lake surface water temperature," *Sci. Total Environ.*, vol. 890, Sep. 2023, Art. no. 164323.

[44] Q. Li and Z. Song, "Prediction of compressive strength of Rice husk ash concrete based on stacking ensemble learning model," *J. Cleaner Prod.*, vol. 382, Jan. 2023, Art. no. 135279.

[45] A. Ghasemieh, A. Lloyed, P. Bahrami, P. Vajar, and R. Kashef, "A novel machine learning model with stacking ensemble learner for predicting emergency readmission of heart-disease patients," *Decis. Anal. J.*, vol. 7, Jun. 2023, Art. no. 100242.

[46] S. Shu, S. Liu, and D. Zhang, "Water quality model calibration of water distribution network using parallel FmGA algorithm," in *Proc. Int. Conf. Elect. Control Eng. (ICECE)*, 2010, pp. 5801–5804.

[47] J. W. Davidson, D. Savic, and G. A. Walters, "Method for the identification of explicit polynomial formulae for the friction in turbulent pipe flow," *J. Hydroinformatics*, vol. 1, no. 2, pp. 115–126, Oct. 1999.

[48] O. Giustolisi and D. A. Savic, "A symbolic data-driven technique based on evolutionary polynomial regression," *J. Hydroinformatics*, vol. 8, no. 3, pp. 207–222, Jul. 2006.

[49] M. Özger and G. Yıldırım, "Determining turbulent flow friction coefficient using adaptive neuro-fuzzy computing technique," *Adv. Eng. Softw.*, vol. 40, no. 4, pp. 281–287, Apr. 2009.

[50] F. Salmasi, R. Khatibi, and M. A. Ghorbani, "A study of friction factor formulation in pipes using artificial intelligence techniques and explicit equations," *Turkish J. Chem.*, vol. 37, no. 4, pp. 121–138, 2012.

[51] G. Yıldırım, "Computer-based analysis of explicit approximations to the implicit Colebrook–White equation in turbulent flow friction factor calculation," *Adv. Eng. Softw.*, vol. 40, no. 11, pp. 1183–1190, Nov. 2009.

[52] M. Thakkar, A. Busse, and N. D. Sandham, "Direct numerical simulation of turbulent channel flow over a surrogate for nikuradse-type roughness," *J. Fluid Mech.*, vol. 837, pp. 1–11, Feb. 2018.

[53] P. Gualtieri, S. D. Felice, V. Pasquino, and G. P. Doria, "Use of conventional flow resistance equations and a model for the nikuradse roughness in vegetated flows at high submergence," *J. Hydrol. Hydromechanics*, vol. 66, no. 1, pp. 107–120, Mar. 2018.

[54] M. V. Zagarola and A. J. Smits, "Mean-flow scaling of turbulent pipe flow," *J. Fluid Mech.*, vol. 373, pp. 33–79, Oct. 1998.

[55] B. J. McKeon, J. Li, W. Jiang, J. F. Morrison, and A. J. Smits, "Further observations on the mean velocity distribution in fully developed pipe flow," *J. Fluid Mech.*, vol. 501, pp. 135–147, Feb. 2004.

[56] D. Taler, "Determining velocity and friction factor for turbulent flow in smooth tubes," *Int. J. Thermal Sci.*, vol. 105, pp. 109–122, Jul. 2016.

[57] C. J. Swanson, B. Julian, G. G. Ihas, and R. J. Donnelly, "Pipe flow measurements over a wide range of Reynolds numbers using liquid helium and various gases," *J. Fluid Mech.*, vol. 461, pp. 51–60, Jun. 2002.

[58] M. Najafzadeh, A. Etemad-Shahidi, and S. Y. Lim, "Scour prediction in long contractions using ANFIS and SVM," *Ocean Eng.*, vol. 111, pp. 128–135, Jan. 2016.

[59] M. Najafzadeh and S. Sarkamaryan, "Extraction of optimal equations for evaluation of pipeline scour depth due to currents," *Proc. Inst. Civil Eng. Maritime Eng.*, vol. 171, no. 1, pp. 1–10, Mar. 2018.

[60] R. Burbidge, M. Trotter, B. Buxton, and S. Holden, "Drug design by machine learning: Support vector machines for pharmaceutical data analysis," *Comput. Chem.*, vol. 26, no. 1, pp. 5–14, Dec. 2001.

[61] N. Parveen, S. Zaidi, and M. Danish, "Artificial intelligence (AI)-based friction factor models for large piping networks," *Chem. Eng. Commun.*, vol. 207, no. 2, pp. 213–230, Feb. 2020.

[62] K. Gao, L. Deng, J. Liu, L. Wen, D. Wong, and Z. Liu, "Study on mine ventilation resistance coefficient inversion based on genetic algorithm," *Arch. Mining Sci.*, vol. 63, no. 4, pp. 813–826, 2018.

[63] D. Tong, H. Wu, C. Liu, Z. Guo, and P. Li, "A hybrid prediction model for local resistance coefficient of water transmission tunnel maintenance ventilation based on machine learning," *Appl. Sci.*, vol. 13, no. 16, p. 9135, Aug. 2023.

[64] M. Yu, T. Quan, Q. Peng, X. Yu, and L. Liu, "A model-based collaborate filtering algorithm based on stacked AutoEncoder," *Neural Comput. Appl.*, vol. 34, no. 4, pp. 2503–2511, Feb. 2022.

[65] T. A. Farrag and E. E. Elattar, "Optimized deep stacked long short-term memory network for long-term load forecasting," *IEEE Access*, vol. 9, pp. 68511–68522, 2021.

[66] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," *PLoS ONE*, vol. 12, no. 7, Jul. 2017, Art. no. e0180944.

[67] M. Wu, Z. Lu, Q. Chen, T. Zhu, E. Lu, W. Lu, and M. Liu, "A two-stage algorithm of locational marginal price calculation subject to carbon emission allowance," *Energies*, vol. 13, no. 10, p. 2510, May 2020.

[68] Z. Zhou, *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL, USA: CRC Press, 2012.

[69] F. He and L. Y. Zhang, "Prediction model of end-point phosphorus content in BOF steelmaking process based on PCA and BP neural network," *J. Process Control*, vol. 66, pp. 51–58, Jun. 2018.

[70] X. Xie, B. Xia, and J. Yu, "A comprehensive evaluation method based on PCA and BP neural network," in *Proc. 5th Int. Conf. Inf. Comput. Sci.*, 2012, pp. 71–74.

[71] L. Zheng and H. He, "Share price prediction of aerospace relevant companies with recurrent neural networks based on PCA," *Expert Syst. Appl.*, vol. 183, Nov. 2021, Art. no. 115384.

[72] S. Espezua, E. Villanueva, and C. D. Maciel, "Towards an efficient genetic algorithm optimizer for sequential projection pursuit," *Neurocomputing*, vol. 123, pp. 40–48, Jan. 2014.

[73] Q. Boying and L. Xiankun, "Construction of response surface based on projection pursuit regression and genetic algorithm," *Phys. Proc.*, vol. 33, pp. 1732–1740, 2012.

[74] Y. Ren, H. Liu, S. Li, X. Yao, and M. Liu, "Prediction of binding affinities to $\beta 1$ isoform of human thyroid hormone receptor by genetic algorithm and projection pursuit regression," *Bioorganic Medicinal Chem. Lett.*, vol. 17, no. 9, pp. 2474–2482, May 2007.

[75] X. Ma, Z. Chen, P. Chen, H. Zheng, X. Gao, J. Xiang, L. Chen, and Y. Huang, "Predicting the utilization factor of blasthole in rock roadways by random forest," *Underground Space*, vol. 11, pp. 232–245, Aug. 2023.

[76] H. Li, J. Lin, X. Lei, and T. Wei, "Compressive strength prediction of basalt fiber reinforced concrete via random forest algorithm," *Mater. Today Commun.*, vol. 30, Mar. 2022, Art. no. 103117.

[77] C. Zhang, X. Ruan, Y. Zhao, and M.-H. Yang, "Contour detection via random forest," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 2772–2775.

[78] J. Cai, X. Li, Z. Tan, and S. Peng, "An assembly-level neutronic calculation method based on LightGBM algorithm," *Ann. Nucl. Energy*, vol. 150, Jan. 2021, Art. no. 107871.

[79] D.-N. Wang, L. Li, and D. Zhao, "Corporate finance risk prediction based on LightGBM," *Inf. Sci.*, vol. 602, pp. 259–268, Jul. 2022.

[80] J. Hancock and T. M. Khoshgoftaar, "Leveraging LightGBM for categorical big data," in *Proc. IEEE 7th Int. Conf. Big Data Comput. Service Appl. (BigDataService)*, Aug. 2021, pp. 149–154.

[81] S. Xiao, M. Shen, and L. Yu, "Energy saving analysis of refrigeration room group control based on kernel ridge regression algorithm," *Int. J. Refrig.*, vol. 153, pp. 345–355, Sep. 2023.

[82] X. Wang, X. Wang, B. Ma, Q. Li, C. Wang, and Y. Shi, "High-performance reversible data hiding based on ridge regression prediction algorithm," *Signal Process.*, vol. 204, pp. 1–12, Mar. 2023.

[83] D. Xiaohong, C. Huajiang, S. A. Bagherzadeh, M. Shayan, and M. Akbari, "Statistical estimation the thermal conductivity of MWCNTs-SiO_2/water-EG nanofluid using the ridge regression method," *Phys. A, Stat. Mech. Appl.*, vol. 537, Jan. 2020, Art. no. 122782.

[84] H. Yan, Z. He, C. Gao, M. Xie, H. Sheng, and H. Chen, "Investment estimation of prefabricated concrete buildings based on XGBoost machine learning algorithm," *Adv. Eng. Informat.*, vol. 54, Oct. 2022, Art. no. 101789.

[85] Y. Wang, F. Su, Y. Guo, H. Yang, Z. Ye, and L. Wang, "Predicting the microbiologically induced concrete corrosion in sewer based on XGBoost algorithm," *Case Stud. Construct. Mater.*, vol. 17, Dec. 2022, Art. no. e01649.

[86] L. Zhang, Y. Ji, T. Liu, and J. Li, "$PM_{2.5}$ prediction based on XGBoost," in *Proc. 7th Int. Conf. Inf. Sci. Control Eng. (ICISCE)*, 2020, pp. 1011–1014.

[87] B. Chen, J. Qin, and A. Yuan, "Variable selection in the Box–Cox power transformation model," *J. Stat. Planning Inference*, vol. 216, pp. 15–28, Jan. 2022.

[88] Y. Peng, T. Feng, and H. J. P. Timmermans, "Expanded comfort assessment in outdoor urban public spaces using box-cox transformation," *Landscape Urban Planning*, vol. 190, Oct. 2019, Art. no. 103594.

[89] H. Zhou and H. Zou, "The nonparametric Box–Cox model for high-dimensional regression analysis," *J. Econometrics*, pp. 1–19, Mar. 2023.

[90] Ö. Baydaroğlu and K. Koçak, "SVR-based prediction of evaporation combined with chaotic approach," *J. Hydrol.*, vol. 508, pp. 356–363, Jan. 2014.

[91] S. Zaidi, "Development of support vector regression (SVR)-based model for prediction of circulation rate in a vertical tube thermosiphon reboiler," *Chem. Eng. Sci.*, vol. 69, no. 1, pp. 514–521, Feb. 2012.

[92] Y. Zhong and J. Xiang, "Pipe damage detection method based on BSWI and SVR," in *Proc. Int. Conf. Measuring Technol. Mechatronics Autom.*, vol. 2, Mar. 2010, pp. 899–902.

[93] S. Li, K. Gao, Y. Liu, H. Zhou, and Z. Liu, "Random forest inversion method for mine ventilation resistance coefficient," *Modern Mining*, vol. 36, no. 5, pp. 205–207, 2020.

[94] F. Pedregosa, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.

[95] X.-W. Ye, X.-L. Zhang, H.-Q. Zhang, Y. Ding, and Y.-M. Chen, "Prediction of lining upward movement during shield tunneling using machine learning algorithms and field monitoring data," *Transp. Geotechnics*, vol. 41, Jul. 2023, Art. no. 101002.

[96] K. Kahar, R. Dhekekar, M. Bhaiyya, S. K. Srivastava, P. Rewatkar, S. Balpande, and S. Goel, "Optimization of MEMS-based energy scavengers and output prediction with machine learning and synthetic data approach," *Sens. Actuators A, Phys.*, vol. 358, Aug. 2023, Art. no. 114429.

[97] Q. Liu, Y. Xue, G. Li, D. Qiu, W. Zhang, Z. Guo, and Z. Li, "Application of KM-SMOTE for rockburst intelligent prediction," *Tunnelling Underground Space Technol.*, vol. 138, Aug. 2023, Art. no. 105180.

[98] M. Najafzadeh, J. Shiri, G. Sadeghi, and A. Ghaemi, "Prediction of the friction factor in pipes using model tree," *ISH J. Hydraulic Eng.*, vol. 24, no. 1, pp. 9–15, Jan. 2018.

ZHIPENG QI was born in Dalian, Liaoning, China, in 1994. He received the bachelor's degree from Liaoning Technical University in 2018, where he is currently pursuing the Ph.D. He served as a Joint Supervisor and Researcher at AGH University of Krakow, Poland. He has already published three high-quality international research articles and holds two granted invention patents. His primary research interest includes ventilation-on-demand in complex mining ventilation networks.

KE GAO was born in Laiwu, Shandong, China, in 1983. He received the master's degree from Central South University, in 2010, and the Ph.D. degree from Liaoning Technical University, in 2016. He was a Visiting Scholar with the University of Technology Sydney, Australia, from 2020 to 2021. Currently, he is the Vice Dean and a Professor with the College of Safety Science and Engineering, Liaoning Technical University. He has published more than 30 high-quality international research articles and holds three granted invention patents. His primary research interests include disaster computation and intelligent ventilation.

DARIUSZ OBRACAJ was born in Bielsko-Biala, Poland, in 1969. He received the master's degree from the Mining Faculty, AGH University of Science and Technology, Kraków, Poland, and the Ph.D. and D.Sc. degrees from the AGH University of Science and Technology. Currently, he is the Deputy Head of the Department of Environmental Engineering, AGH UST. He has published more than 20 articles in high-quality international journals, 12 monographs and books, 30 book chapters, and more than 90 technical papers in conference proceedings as the author or coauthor. He holds five granted invention patents. His primary research interests include mine ventilation and mining safety.

YUJIAO LIU was born in Chenzhou, Hunan, China, in 1984. She received the Ph.D. degree from Liaoning Technical University, in 2016. She was a Visiting Scholar with the University of Technology Sydney, Australia, from 2020 to 2021. She is currently an Associate Professor with the College of Safety Science and Engineering, Liaoning Technical University. Her research focuses on intelligent ventilation in mining. She has authored more than 20 high-quality research articles. Her research interest includes optimizing mine ventilation systems.

KEYI YUAN was born in Guizhou, China, in 2000. He is currently pursuing the master's degree in safety engineering with Liaoning Technical University, Huludao, China. His primary research interest includes intelligent mine ventilation.