International Journal of Coal Science & Technology (2022) 9:51
https://doi.org/10.1007/s40789-022-00519-8

RESEARCH

Cooperative prediction method of gas emission from mining face based on feature selection and machine learning

Jie Zhou[1] · Haifei Lin[1,2] · Hongwei Jin[1,2] · Shugang Li[1,2] · Zhenguo Yan[1,2] · Shiyin Huang[1]

Received: 22 July 2021 / Accepted: 16 June 2022

Abstract

Collaborative prediction model of gas emission quantity was built by feature selection and supervised machine learning algorithm to improve the scientific and accurate prediction of gas emission quantity in the mining face. The collaborative prediction model was screened by precision evaluation index. Samples were pretreated by data standardization, and 20 characteristic parameter combinations for gas emission quantity prediction were determined through 4 kinds of feature selection methods. A total of 160 collaborative prediction models of gas emission quantity were constructed by using 8 kinds of classical supervised machine learning algorithm and 20 characteristic parameter combinations. Determination coefficient, normalized mean square error, mean absolute percentage error range, Hill coefficient, mean absolute error, and the mean relative error indicators were used to verify and evaluate the performance of the collaborative forecasting model. As such, the high prediction accuracy of three kinds of machine learning algorithms and seven kinds of characteristic parameter combinations were screened out, and seven optimized collaborative forecasting models were finally determined. Results show that the judgement coefficients, normalized mean square error, mean absolute percentage error, and Hill inequality coefficient of the 7 optimized collaborative prediction models are 0.969–0.999, 0.001–0.050, 0.004–0.057, and 0.002–0.037, respectively. The determination coefficient of the final prediction sequence, the normalized mean square error, the mean absolute percentage error, the Hill inequality coefficient, the absolute error, and the mean relative error are 0.998%, 0.003%, 0.022%, 0.010%, 0.080%, and 2.200%, respectively. The multi-parameter, multi-algorithm, multi-combination, and multi-judgement index prediction model has high accuracy and certain universality that can provide a new idea for the accurate prediction of gas emission quantity.

Keywords Gas emission prediction · Machine learning · Feature selection · Cooperative prediction

1 Introduction

China is rich in coal resources and has a high dependence on coal in production and life for a long time. Coal mine gas is a main factor that affects coal mine safety production, and the prediction of gas emission in the working face is the main basis for determining gas emission grade in mine or horizontal mining. (Zhou et al. 2020a, b; Gao et al. 2020; Long et al. 2021). In 1964, Lindine (Shanjun 1998) established the first empirical model for predicting gas emission in coal mine. Since then, mine statistics method, separate-source prediction method, and gas geological map method were gradually applied in gas emission prediction (Zhang and Zhang 2005; Dai et al 2007). However, this type of prediction method does not consider the gas emission and its migration as a dynamic nonlinear system. For decades, the prediction technology of gas emissions from underground coal mining has been the subject of extensive research. The technology ranges from simple geometric models to modern finite element models (Wang et al. 2015; Guo et al. 2020; Liu et al. 2021). Researchers have adopted experiments and numerical simulations to study the occurrence of coal seam gas. In addition, to predict the gas emission rate of the longwall working face, a numerical gas emission model was established on the basis of ventilation pressure and the flow survey of the entire mine (p-Q survey) (Karacan 2008;

✉ Haifei Lin
lhaifei@163.com

[1] College of Safety Science and Engineering, Xi'an University of Science and Technology, Xi'an 710054, China

[2] Coal Industry Engineering Research Center for Western Mine Gas Intelligent Extraction, Xi'an 710054, China

Guo et al. 2012). The mathematical method of gas geology based on case analysis (Zhang and Yuan 1999; Zhang et al. 2009) has been developed rapidly. This method mainly uses machine learning algorithms and data mining techniques to establish a predictive model, these techniques can consider the dynamic changes of multiple factors by analyzing real-time data of gas emission. Scholars used methods based on statistics, principal component analysis (PCA), and artificial neural networks (ANN) to predict the ventilation gas emission rate of longwall mines in the United States (Karacan and Goodman 2012; Karacan and Olea 2014).

Recently, researchers have given increasing attention on the parameter selection and model establishment of gas emission prediction (typical reference was summarized in Table 1).

Table 1 shows that common factors include coal seam thickness, buried depth, dip angle, gas content in coal seam, floor elevation of the coal seam, spacing between adjacent layers, thickness of adjacent layers of the coal seam, daily output, daily advancing distance, and pure amount of gas extraction. Most previous research on gas emission prediction only focuses on single parameter combination or single prediction algorithm. The accuracy, generalization, and reliability of the gas emission prediction method based on case analysis mainly depend on the influencing factors of gas emission and the selected algorithm. Consequently, the limitations of the prediction model must be extinguished, and various feature combinations should be effectively matched with different machine learning algorithms.

The foothold of this work was to propose a new gas emission prediction method. For a series of gas emission influencing factors, the feature selection method was used to form different gas emission factor combinations, and various machine learning algorithms were applied to traverse all the gas emission factors. The combination of factors and the machine learning algorithm were selected. This new prediction method avoids the limitations of using single combination of factors and single machine learning algorithm in previously published papers.

The new method contains multiple characteristic parameters, algorithms, combinations, and judgment indicators. Pearson correlation coefficient, full subset regression, recursive feature elimination (RFE), and random forest (RF) were applied to determine the optimal combination of gas emission feature parameters. Gaussian process regression (GPR), support vector machine (SVM), least squares SVM (LS-SVM), gradient boosted regression tree (GBRT), random forest (RF), multilayer perceptron (MLP), BP neural network (BPNN), and Elman neural network (ENN) were applied to construct dynamic prediction model with a multi-parameter combination. Normalized mean square error (NMSE), mean absolute percentage error (MAPE), Theil IC (TIC), and judgment coefficient (R^2) were applied to evaluate the accuracy of the model comprehensively. The new technique can provide a basis for the accurate prediction of gas emission.

2 Data processing

2.1 Data instance acquisition

The influencing factors of this paper can be divided into geological and mining factors, which are called first indicators. The secondary indicators that characterize geological factors include coal seam thickness (M), buried depth (H), dip angle (D), gas content in coal seam (GC), floor elevation of the coal seam (BLV), spacing between adjacent layers (SD), and thickness of adjacent layers (ML) of the coal seam. The factors that characterize mining include the daily output (DO), daily advancing distance (V), and pure amount of gas extraction (EP) of the working face. The predicted data were derived from Ma (2017) and Yan (2020).

2.2 Analysis of acquired data

A total of 60 groups of statistical parameters are shown in Fig. 1. To improve the generalization ability of the model and prevent the model from overfitting, the data set was shuffled randomly. The data set is divided into training set (40 sets of data) and verification set (20 sets of data), and the ratio was 2:1. The training set is used for model training, whereas the verification set is used to verify and evaluate the reliability and generalization performance of the trained model.

2.3 Data standardization

The 10 input parameters selected in the gas emission data set were all numerical data, and the value ranges of the different parameters varied and may even have diverse orders of magnitude. To obtain accurate prediction results and ensure that each parameter plays a role, Z-score standardization was performed on the parameters to reduce the influence of parameter scale on the model.

The sequence $x_1, x_2, \ldots, x_n$ is transformed:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{2}$$

Table 1 Typical reference focusing on gas emission influencing factors

Reference	Factor		Method
	Geological factor	Mining factor	
Wang et al. (2011)	Buried depth; coal seam thickness; dip angle; gas content in coal seam; spacing between adjacent layers; gas content of adjacent layers; thickness of adjacent layers; interlayer lithology	Mining height; length of working face; mining speed; mining rate; daily output	Kalman Filtering; Artificial Neural Network (ANN)
Lv et al. (2012)	Buried depth; gas content in coal seam; buried depth; the coal seam thickness; dip angle; gas content of adjacent layers; thickness of adjacent layers; spacing between adjacent layers; interlayer lithology	Mining speed; mining rate; length of working face	Principal Component Analysis (PCA); Multistep Linear Regression (MLA)
Fu et al. (2014)	Gas content in coal seam; buried depth; coal seam thickness; dip angle; gas content of adjacent layers; thickness of adjacent layers; spacing between adjacent layers; interlayer lithology	Mining height; daily output; length of working face; mining rate; mining intensity	Ant Colony Clustering (ACC); Elman Neural Network (ENN)
Yuan et al. (2013)	Buried depth; coal seam thickness; gas content in coal seam; dip angle	Mining speed; daily output	Radial Basis Function neural network (RBFNN); BP neural network (BPNN)
Chen et al. (2015)	Buried depth; the coal seam thickness; gas content in coal seam; spacing between adjacent layers	Mining speed; daily output	Chaotic Immune Genetic Optimization Algorithm (CIGOA); Elman Neural Network (ENN)
Hu et al. (2017)	Buried depth; dip angle; gas content in coal seam; spacing between adjacent layers; coal seam thickness; gas content of adjacent layers; thickness of adjacent layers; interlayer lithology	Length of working face; mining speed; mining rate; daily output	Teaching Learning Based Optimization (TLBO)
Liu et al. (2019)	Buried depth; coal seam thickness; gas pressure; gas content in coal seam	Mining speed; daily output	Classification and regression tree (CART); Support vector machine (SVM)
Xu et al. (2019)	Gas content in coal seam; buried depth; coal seam thickness; spacing between adjacent layers	Mining speed; mining rate; daily output	Improved gravitational search algorithm (IGSA); BP neural network (BPNN)
Xiao et al. (2020)	Gas content in coal seam; gas content of adjacent layers; buried depth; coal seam thickness; dip angle; thickness of adjacent layers; spacing between adjacent layers	Mining height; length of working face; mining rate; mining intensity; mining speed	Contractive Mapping Genetic Algorithm (CMGA); BP neural network (BPNN)

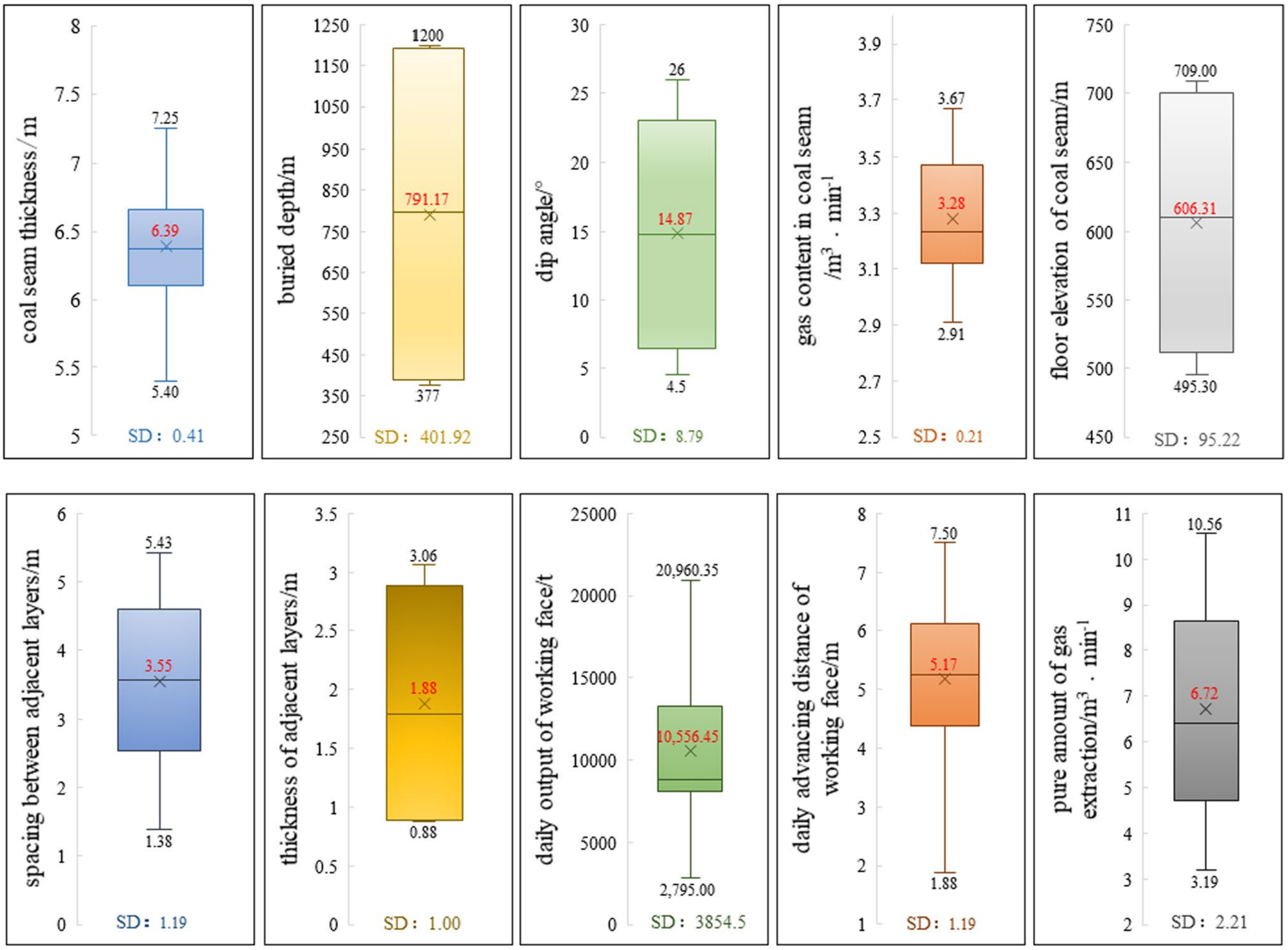


Fig. 1 Box plot of various gas emission parameters. *Notes*: The upper and lower data represent the maximum and minimum values of each parameter, respectively, and the red data represent the average values

$$h_i = \frac{x_i - \overline{x}}{s} \tag{3}$$

where x_i is the original sequence, $i \in [1, n]$; $\overline{x}$ is the average value of the sequence, s is the standard deviation, and h_i is the new sequence after transformation, $i \in [1, n]$.

3 Gas emission prediction model establishment process and primary selection

3.1 Establishment of the prediction model

(1) Sample data processing. The data set is standardized by Z-score method.

(2) Combination selection of feature parameters and determination of algorithm hyper-parameters.

The training set was divided into five parts, and then five-fold grid search cross-validation processes were performed, each time a different part is used as the validation set, and the four remaining parts were combined as the training set. Each sample was used as a validation sample in one experiment and a training sample in four experiments to obtain the optimal parameters for the algorithm with the highest accuracy.

In the grid search process, a series of priori candidate values of the algorithm-related parameters was given first, and all parameter value combinations were tested through loop traversal, and then the parameter value combinations that enable the algorithm to perform optimally were obtained.

(3) Establishment of the prediction model. Different supervised algorithms and characteristic parameters were used to establish the gas emission prediction model.

(4) Primary selection of the collaborative model. By analyzing the verification set data, the algorithm and the parameter combination with the average judgment coefficient R^2 greater than 0.80 is selected. Thus, the prediction cooperation model was selected preliminarily.

(5) Collaborative model optimization. In the above prediction models, the prediction model with the sum of MAPE and TIC less than 0.1 was selected (Ashis et al. 2013; Jin et al 2020, Seçkin et al. 2020), then the prediction model with the maximum relative error (RE_{max}) less than 15% and mean relative error (MRE) less than 10% was chosen as the optimization cooperation model.

(6) Collaborative model prediction. The predicted value ($\hat{y}_i$) was obtained after averaging the predicted data of each group of the optimized collaborative model.

The prediction flow of gas emission in the working face is shown in Fig. 2.

3.2 Primary selection of forecasting model

3.2.1 Feature combination selection

Feature selection refers to the selection of a feature subset according to the importance in a feature set. Few variables will lead to the low accuracy of the model, and excessive parameters cannot necessarily increase the accuracy of the model but lead to over-fitting problem. Furthermore, different feature combinations have diverse sensitivities to various machine learning algorithms. Therefore, the main function of the feature selection is to strengthen the generalization ability of the model, reduce over-fitting, and enhance the understanding between features and eigenvalues. Generally, the feature selection methods can be divided into three categories: direct method, univariate feature selection, and multivariate feature selection. In this paper, the Pearson correlation coefficient method, full subset regression, RFE, and RF were used to obtain the best input variable combination (Fig. 3).

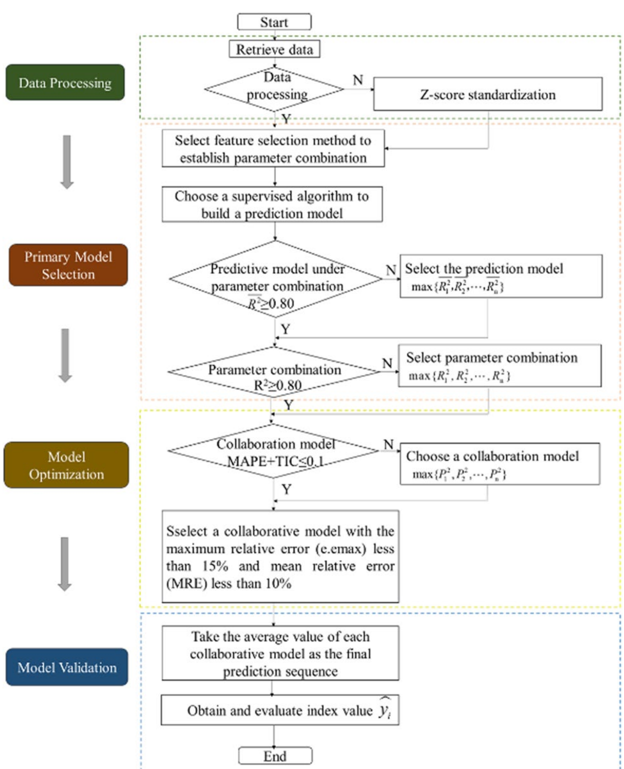


Fig. 2 Establish flow chart of prediction model

Figure 3a shows the correlation analysis by using the Pearson correlation coefficient method (Dominic et al. 2020). Pearson correlation coefficient was used for measuring the correlation between N and M. Its value is between − 1 and 1. Pearson correlation coefficient can be expressed as:

$$r = \frac{\sum_{i=1}^{n}(N_i - \overline{N})(M_i - \overline{M})}{\sqrt{\sum_{i=1}^{n}(N_i - \overline{N})^2}\sqrt{\sum_{i=1}^{n}(M_i - \overline{M})^2}} \tag{4}$$

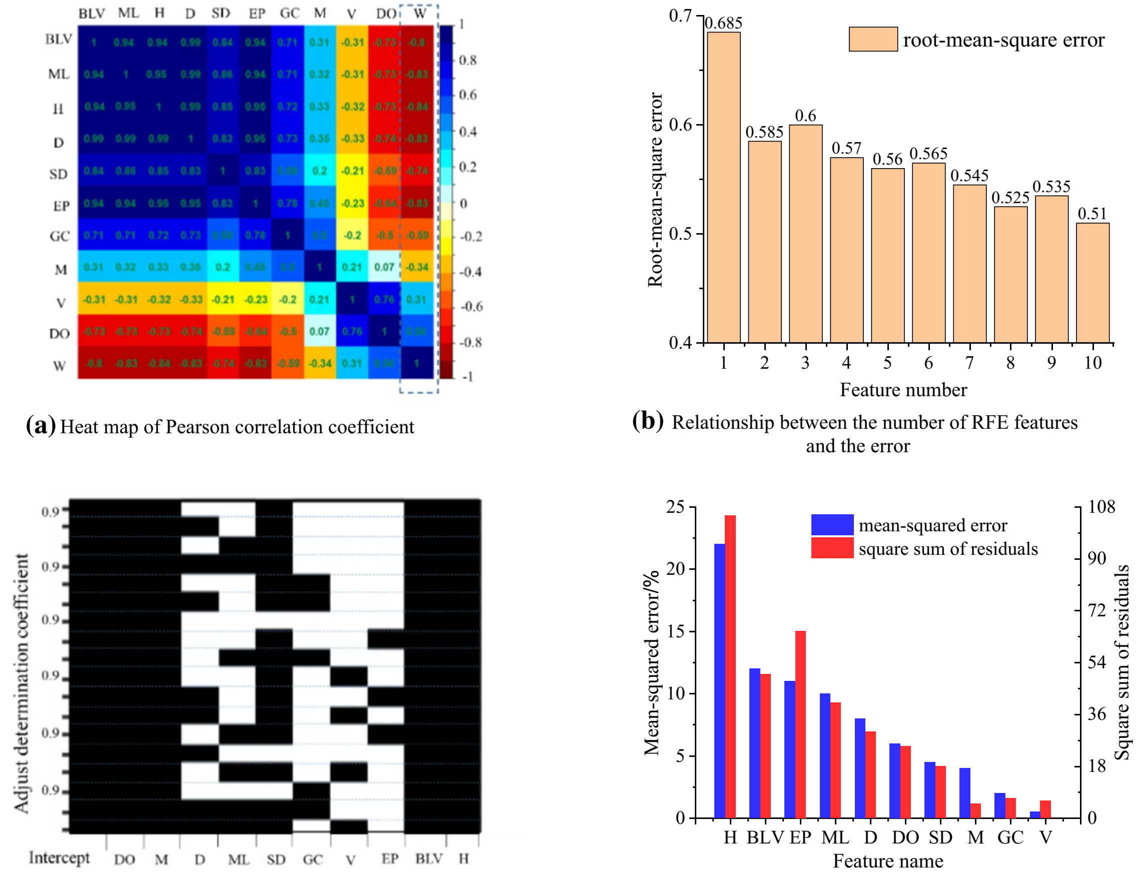
where N and M represent two pairs of continuous variables.

According to Eq. (4) and Pearson correlation coefficient classification rules, the absolute value of Pearson correlation coefficient that is greater than 0.4 is regarded as moderate correlation. In this example, the variable above moderately correlated is considered the input variable, and the dashed box represents the correlation degree between gas emission and each parameter.

RFE is a wrapper feature selection method, in which the search starting point is all features, and the evaluation criterion is the mean square error of each grouping. After cyclic iteration, each iteration eliminated the least relevant feature. The combination with the smallest mean square error is the optimal feature subset (You et al. 2014; Ke et al. 2015) (Fig. 3b). In Fig. 3b, the abscissa represents the number of features, whereas the ordinate represents the mean square error of a specific group. When the number of features were 10 (all features), the mean square error was the smallest.

Full subset screening was based on all possible combinations of different independent variables. The reduced variable combinations were fitted by the least square method, and a model with a corrected coefficient of determination greater than 0.9 was selected among all the possible models (Zhang et al. 2019a, b). The selection result was shown in Fig. 3c. In this example, 17 optimal combinations were obtained through full subset screening, and the determination coefficients of these 17 combinations were all greater than 0.9.

A large number of decision trees was used for the feature selection in RF (Speiser et al. 2019), and the variables obtained from each decision tree were synthesized to obtain the final variable importance ranking (Fig. 3d).



(a) Heat map of Pearson correlation coefficient

(b) Relationship between the number of RFE features and the error

(c) Relationship between the feature number of full subset and adjusted judgement coefficient

(d) Feature selection by random forest method

Fig. 3 Results of feature selection methods. *Notes*: The black box in **c** represents the selected parameter, and the blue dotted line represents the division of different parameter combinations, a total of 17 groups

In this example, according to the RMSE and square sum of residuals, nine factors, except for the buried depth of coal seam, are selected.

In summary, the Pearson correlation coefficient, RFE, full subset regression, and RF were used to select 10 influencing factors according to different laws. A total of 20 sets of feature combinations were obtained (Fig. 4).

3.2.2 Selection of prediction algorithm

(1) Regression algorithm

GPR (Mahmoodzadeh et al. 2021; Noori et al. 2019) has good adaptability and strong generalization ability to address high-dimensional, small-sample, nonlinear, and complex problems. Compared with neural network and SVM, this method has the advantages of easy implementation and adaptive acquisition of superparameters. SVM (Qian et al. 2014; Zhou et al. 2012) has shown many unique advantages in solving small sample, nonlinear, and high-dimensional pattern recognition problems. The ultimate goal of the LS-SVM (Xue and Xiao 2017) optimization problem is to obtain the optimized model parameters. The linear decision function constructed by LS-SVM not only has good fitting performance but also has strong generalization ability.

(2) Neural network

Multilayer is an essential feature of MLP (Teresa and Wilson 2013) that includes an input layer, an output layer, and a hidden layer. No specific number of hidden layers is provided. Thus, the appropriate number of hidden layers can be selected according to the requirements. The number of neurons in the output layer are unlimited. BPNN (Zhang et al. 2019a, b; Zhao et al. 2021) is a multi-layer perceptron network trained according to error back propagation and consists of an input layer, at least one hidden layer, and an output layer. ENN (Xie et al. 2019) is a kind of dynamic feedback network that not only has an input layer, a hidden layer, and an output layer unit but also has a special connection unit. The special connection unit

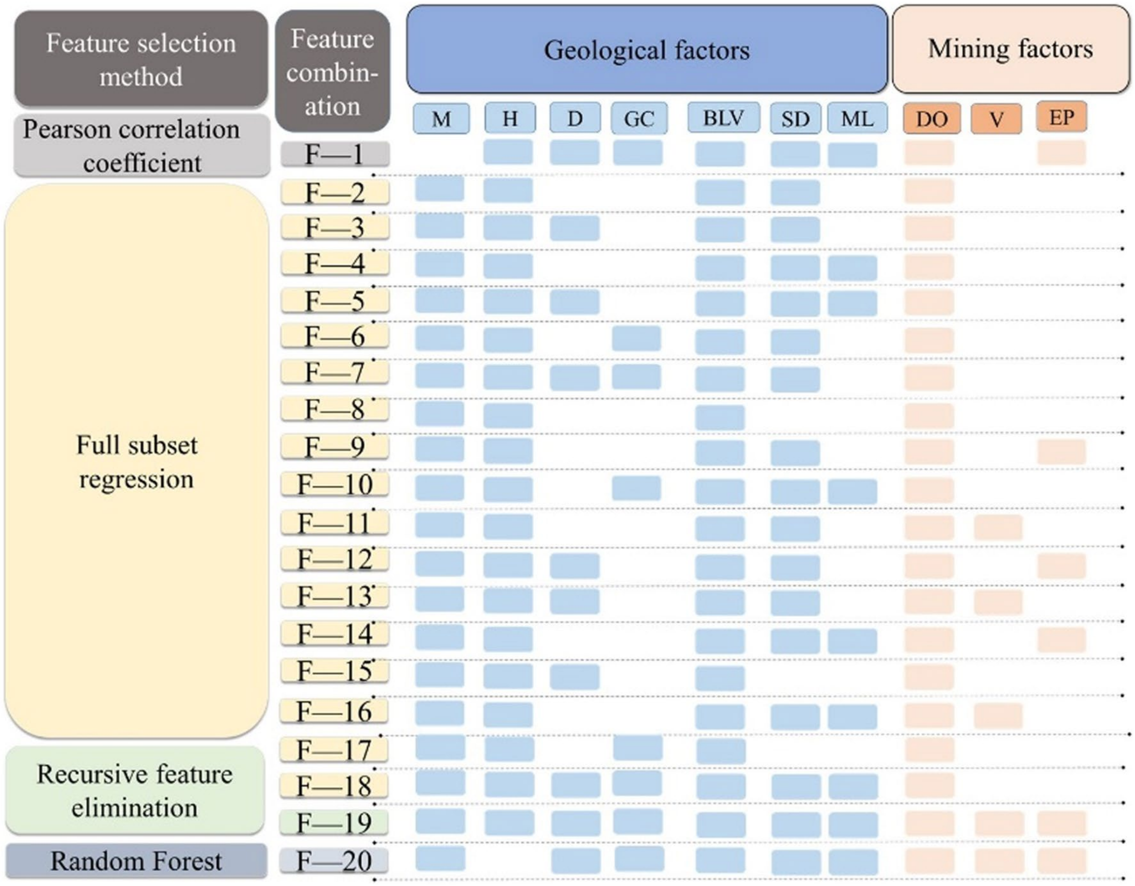


Fig. 4 Combination set of feature parameters affecting gas emission. *Notes*: F-1 represents the first feature combination, F-2 represents the second feature combination, and so on. The feature combination of the same color is selected by the corresponding feature selection method on the left

can be regarded as a time delay method that enables the network to have the function of dynamic memory.

(3) Integrated learning

In ensemble learning, a series of learners is used, and a certain rule is adopted to integrate various learning results to obtain significantly better generalization performance than a single learner. In this paper, in addition to ensemble learning, six single machine learning algorithms were also proposed to compare the ensemble algorithm and a single algorithm and adopt more comprehensive methods to establish a gas emission prediction model. The main methods in ensemble learning include boosting and bagging, and the combination rules of the two differ.

The main idea of boosting ensemble learning is to assemble diverse weak classifiers into a strong classifier and then combine them linearly through additive model. GBRT (Zhou et al. 2020a, b; Persson et al. 2017) is a kind of boosting, and each calculation reduces the last residual error and builds a new model. In another integrated learning method called bagging, no strong dependence is observed among individual learners. RF (Lu et al. 2016) refers to an evolutionary version of the bagging algorithm. In the randomly selected sample features, an optimal feature is selected to divide the left and right subtrees of the decision tree and further enhance the generalization ability of the model.

Through the 20 feature combinations in Fig. 3 and eight different supervised learning algorithms, 160 kinds of gas emission prediction models in the working face are constructed. These prediction models are used to verify 20 groups of data in the verification set randomly, and the R^2 is shown in Table 2. R^2 is calculated using Eq. (5) as follows.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \tag{5}$$

where, y_i is the true value, and $i \in [1, n]$; $\hat{y}_i$ is the predicted value, $i \in [1, n]$.

The R^2 of the prediction model ranges from 0.255 to 0.999, among which the average judgment coefficient of LS-SVM (0.936), GBRT (0.932), MLP (0.901), and RF (0.803) are all greater than 0.800. The LS-SVM has low dependence

Table 2 R^2 of various algorithms using different parameter combinations

Feature combination	Supervised learning								Combination mean value
	GPR	SVM	LS-SVM	GBRT	RF	MLP	BP	Elman	
F-1	0.653	0.876	0.946	0.958	0.705	0.999	0.593	0.500	0.779
F-2	0.558	0.738	0.903	0.966	0.835	0.982	0.652	0.560	0.774
F-3	0.626	0.854	0.907	0.961	0.966	0.975	0.798	0.531	0.827
F-4	0.638	0.780	0.903	0.958	0.940	0.862	0.737	0.565	0.798
F-5	0.493	0.820	0.907	0.963	0.774	0.255	0.774	0.583	0.696
F-6	0.604	0.652	0.946	0.964	0.762	0.994	0.783	0.533	0.780
F-7	0.554	0.823	0.951	0.960	0.846	0.939	0.603	0.682	0.795
F-8	0.719	0.703	0.900	0.968	0.650	0.453	0.873	0.569	0.729
F-9	0.664	0.655	0.974	0.955	0.837	0.995	0.876	0.645	0.825
F-10	0.449	0.656	0.946	0.966	0.761	0.993	0.600	0.622	0.749
F-11	0.614	0.796	0.917	0.969	0.835	0.999	0.779	0.693	0.825
F-12	0.583	0.825	0.975	0.958	0.906	0.999	0.843	0.549	0.830
F-13	0.621	0.809	0.918	0.962	0.607	0.983	0.890	0.634	0.803
F-14	0.620	0.729	0.974	0.965	0.764	0.915	0.825	0.692	0.810
F-15	0.660	0.827	0.899	0.963	0.803	0.692	0.798	0.646	0.786
F-16	0.575	0.740	0.917	0.961	0.777	0.990	0.641	0.662	0.783
F-17	0.588	0.722	0.922	0.960	0.804	0.998	0.785	0.582	0.795
F-18	0.557	0.654	0.951	0.963	0.837	0.988	0.581	0.674	0.775
F-19	0.558	0.672	0.986	0.639	0.792	0.999	0.917	0.602	0.771
F-20	0.606	0.879	0.986	0.692	0.859	0.999	0.790	0.598	0.801
Algorithm mean value	0.597	0.760	0.936	0.932	0.803	0.901	0.757	0.606	

on feature combination (the range of R^2 is 0.899–0.986), followed by GBRT (the range of R^2 is 0.639–0.969) and RF (the range of R^2 is 0.607–0.966), whereas the MLP fluctuates greatly (the range of R^2 is 0.255–0.999). Except for the first four algorithms, the average judgment coefficient of the other algorithms is less than 0.800, among which BPNN has the largest fluctuation, with R^2 ranging from 0.581 to 0.917 (Table 2).

4 Optimization and verification of gas emission prediction model

4.1 Gas emission prediction model optimization

4.1.1 Determination of optimal prediction algorithm and feature combination

The average judgment coefficients of seven feature parameter combinations, such as F-3, F-9, F-11, F-12, F-13, F-14, and F-20, under various algorithms are all greater than 0.800. The NMSE (Das et al. 2020), MAPE, and TIC of 28 types of prediction models under the four algorithms and seven feature parameter combinations are calculated. The calculation is shown in Eqs. (6) to (8), and the results are shown in Table 2 and Fig. 5.

$$\mathrm{NMSE} = \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)^2} \tag{6}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{7}$$

$$\mathrm{TIC} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i^2} + \sqrt{\frac{1}{n}\sum_{i=1}^{n}y_i^2}} \tag{8}$$

The values of NMSE, MAPE, and TIC of the LS-SVM are less than 0.1, 0.040–0.084, and 0.024–0.063, respectively. The values of NMSE, MAPE, and TIC of GBRT are 0.031–0.308, 0.049–0.166, and 0.037–0.133,

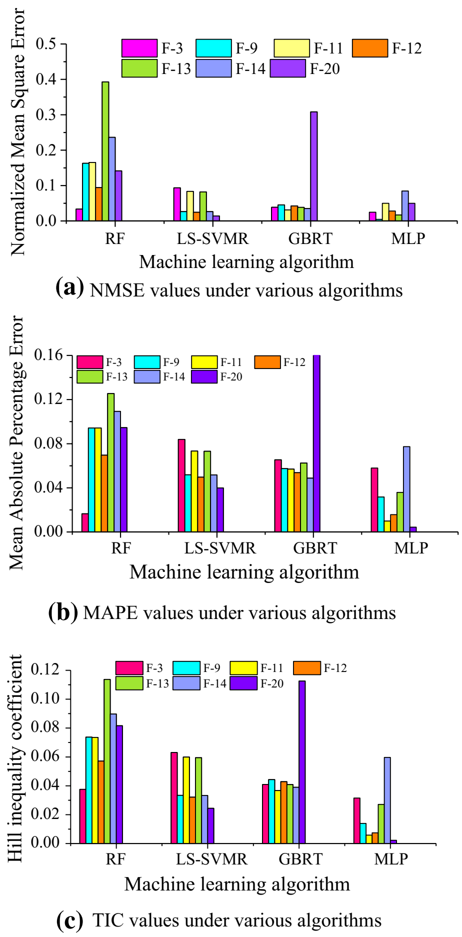


Fig. 5 Evaluation index of various algorithms using different parameter combinations

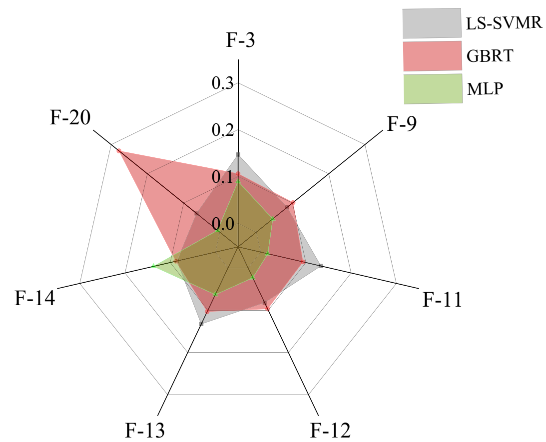


Fig. 6 Data distribution of added MAPE and TIC

respectively. The values of NMSE, MAPE, and TIC of RF are 0.034–0.393, 0.016–0.125, and 0.037–0.114, respectively. The values of NMSE, MAPE, and TIC of MLP are 0.001–0.085, 0.004–0.077, and 0.002–0.060, respectively. Overall, except for the low accuracy of RF, the prediction results of LS-SVM, GBRT, and MLP are ideal regardless of the combination of accuracy and volatility (Fig. 5).

4.1.2 Determination of the optimal collaborative forecasting model

MAPE and TIC have similar meanings, and the changes in MAPE and TIC are considered comprehensively. The MAPE + TIC value of the green area, where the MLP is located, is mostly less than 0.1, followed by the LS-SVMR and the GBRT. Finally, 13 prediction models are selected, and the evaluation indexes of the prediction results of each model are shown in Fig. 6.

To ensure the stability of prediction sequence, the prediction models with the maximum relative error (RE_{max}) less than 15% and mean relative error (MRE) less than 10% are selected as the optimal collaborative prediction models. The optimal collaborative prediction models are LS-SVM and F-20, GBRT and F-11, MLP and F-3, F-9, F-11, F-12, F-13, F-20 (Table 3).

4.2 Verification of optimal collaborative forecasting model

The average predicted data value of these eight optimized collaborative models is taken as the final predicted value. The predicted evaluation indexes are shown in Fig. 7 and Table 4. All the evaluation indexes of gas emission prediction results meet the requirements by optimizing the collaborative model. The absolute error (AE) and mean relative error (MRE) are calculated by Eqs. (9) and (10), respectively.

$$\mathrm{AE} = y_i - \hat{y}_i \tag{9}$$

$$\mathrm{MRE} = \frac{1}{n}\left|\left(\frac{y_i - \hat{y}_i}{y_i}\right)\right| \tag{10}$$

The maximum relative error (RE_{max}), the minimum relative error (RE_{min}), and the MRE of the predicted sequence in this paper are better than those in Ma (2017), Yan (2020), Wang et al. (2018), Jing et al. (2011) (Table 5).

Table 3 Evaluation indexes of model prediction results

Model	NMSE	MAPE	TIC	R^2	AE	MRE (%)	RE_{max} (%)
LSSVM combined with F-20	0.014	0.040	0.024	0.986	0.171	3.99	14.64
GBRT combined with F-11	0.031	0.057	0.037	0.969	0.254	5.70	13.14
MLP combined with F-3	0.025	0.058	0.032	0.975	0.255	5.80	14.19
MLP combined with F-9	0.005	0.032	0.014	0.995	0.255	5.74	9.06
MLP combined with F-11	0.001	0.010	0.006	0.999	0.254	5.70	3.28
MLP combined with F-12	0.001	0.016	0.007	0.998	0.244	5.37	6.60
MLP combined with F-13	0.017	0.036	0.027	0.983	0.273	6.25	10.88
MLP combined with F-20	0.001	0.004	0.002	0.998	0.016	0.43	1.42

Table 4 Optimization model evaluation indicators

NMSE	MAPE	TIC	R^2	AE	MRE	RE_{max}
0.003	0.022	0.010	0.998	0.082	2.20%	6.09%

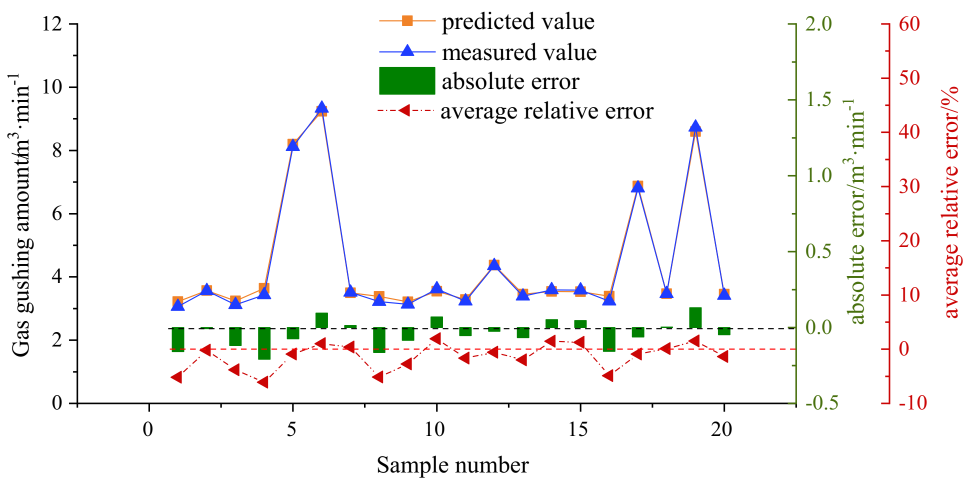


Fig. 7 Comparison of predicted value and original value

Table 5 Comparison of optimization models

Reference	RE_{max} (%)	RE_{min} (%)	MRE (%)
This paper	6.09	0.09	2.20
Ma (2017)	8.40	0.13	2.75
Yan (2020)	9.27	0.22	2.59
Wang et al. (2018)	15.32	1.89	6.94
Jing et al. (2011)	17.07	0.25	6.99

5 Conclusions

The use of data mining techniques is of great significance to analyze the rules between parameter combination and machine learning algorithm for the prediction of coal mine gas emission. Through the selection of feature parameter combination, establishment of prediction model, selection of collaborative model, and verification of the model, the latter realizes gas emission prediction under multiple characteristic parameters, algorithms, combinations, and judgment indicators. The main conclusions are presented as follows:

(1) A total of 20 combination sets of characteristic parameters of influencing factors of gas emission are established in the working face; one parameter combination is obtained by Pearson correlation coefficient method, full subset regression, and RF; and 17 parameter combinations are determined by recursive feature elimination.
(2) The R^2 of 160 kinds of gas emission prediction models with different combinations of algorithms and feature parameters are 0.255–0.999. Four algorithms, namely, LS-SVM, GBRT, RF, and MLP, have average judgment coefficients that are more significant than 0.800.
(3) Eight cooperative models, LS-SVM and F-20, GBRT and F-11, MLP and F-3, F-9, F-11, F-12, F-13, and F-20, can be used for predicting and optimizing gas

emission in the working face. The evaluation indexes of the final predicted value and the original value all meet the requirements.

A new gas emission prediction concept is proposed in this paper. Multi-parameter combination and multi-machine learning algorithm form a multi-prediction model group. In the future, based on the proposed collaborative prediction model of gas emission in the working face, this concept can be further verified by more gas emission influencing factors, algorithms, and sample data sets to optimize the prediction model further.

Acknowledgements This work was supported by National Natural Science Foundation of China (51734007); Outstanding Youth Program of Shaanxi Province, China (2020JC-48); Key Enterprise Joint Fund of Shaanxi Province, China (2019JLP-02).

Author contributions All authors read and approved the final manuscript.

Conflict of interest The manuscript is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously. All the authors listed have approved the manuscript that is enclosed. We declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ashis M, Abhijit M, Prabal K M, Debamalya B (2013) Comparative analysis of regression and ANN models for predicting drape coefficient of handloom fabrics. Indian J Fibre Text Res 37(4):313–320

Chen WH, Yan XH, Fu H (2015) On the innovated Elman neural network for forecasting the mining gas emission. J Saf Environ 15(3):19–24

Dai GL, Wang YQ, Zhang CR, Li QM, Shao GY (2007) Forecast of the gas effused from the face in protective seam. J China Coal Soc 32(4):382–385

Das P, Chanda K (2020) Bayesian Network based modeling of regional rainfall from multiple local meteorological drivers. J Hydrol 591:125563

Dominic E, Tamás F, Gábor J (2020) On relationships between the Pearson and the distance correlation coefficients. Statist Probab Lett 169:108960

Fu H, Xie S, Xu YS, Chen ZC (2014) Gas emission dynamic prediction model of coal mine based on ACC-ENN algorithm. J China Coal Soc 39(7):1296–1301

Guo D, Lv P, Zhao J et al (2020) Research progress on permeability improvement mechanisms and technologies of coalbed deep-hole cumulative blasting. Int J Coal Sci Technol 7(2):329–336

Gao K, Li SN, Han R, Li RZ (2020) Study on the propagation law of gas explosion in the space based on the goaf characteristic of coal mine. Saf Sci 127:104693

Guo H, Liang Y, Shen BT (2012) Mining-induced strata stress changes, fractures and gas flow dynamics in multi-seam longwall mining. Int J Rock Mech Min Sci 54:129–139

Hu K, Wang SZ, Han S, Wang S (2017) Gas emission quantity prediction of working face based on TLBO-LOIRE method. J Basic Sci Eng 25(5):1048–1056

Jin HC, Hyun JH, Lee W, Bhang BG (2020) Power performance of high density photovoltaic module using energy balance model under high humidity environment. Sol Energy 219(5):50–57

Jing GX, Xu SM, Heng XW, Li CQ (2011) Research on the prediction of gas emission quantity in coal mine based on grey system and linear regression for one element. In: The first international symposium on mine safety science and engineering

Karacan CZ (2008) Modeling and prediction of ventilation methane emissions of U.S. longwall mines using supervised artificial neural networks. Int J Coal Geol 73(4):371–387

Karacan CZ, Goodman GVR (2012) A CART technique to adjust production from longwall coal operations under ventilation constraints. Saf Sci 50(3):510–522

Karacan CZ, Olea RA (2014) Inference of strata separation and gas emission paths in longwall overburden using continuous wavelet transform of well logs and geostatistical simulation. J Appl Geophys 105:147–158

Ke Y, Zhang D (2015) Feature selection and analysis on correlated gas sensor data with recursive feature elimination. Sens Actuat B Chem 212:353–363

Liu P, Wei HZ, Jing JB, Yang YY (2019) Predicting technology of gas emission quantity in coal mine based on enhanced CART regression algorithm. Coal Sci Technol 47(11):116–122

Liu A, Liu S, Liu P et al (2021) Water sorption on coal: effects of oxygen-containing function groups and pore structure. Int J Coal Sci Technol 8(5):983–1002

Long H, Lin HF, Yan M, Bai Y, Xiao T, Kong XG, Li SG (2021) Adsorption and diffusion characteristics of CH_4, CO_2, and N_2 in micropores and mesopores of bituminous coal: molecular dynamics. Fuel 292:120268

Lu PC, Qiu JL, Bian CF, Chen LL, Chen X (2016) Clustering-based under-sampling ensemble method for software defect prediction. Comput Eng Design 37(7):1805–1810

Lv F, Liang B, Sun WJ, Wang Y (2012) Gas emission quantity prediction of working face based on principal component regression analysis method. J China Coal Soc 37(01):113–116

Ma YY (2017) Research of the forecast gas emission based on factor analysis and kalman filter. Xian, China, Xi'an university of science and technology

Mahmoodzadeh A, Mohammadi M, Daraei A, Rashid TA (2021) Tunnel geomechanical parameters prediction using Gaussian process regression. Mach Learn Appl 3(15):10020

Noori M, Hassani H, Javaherian A, Amindavar H, Torabi S (2019) Automatic fault detection in seismic data using Gaussian process regression. J Appl Geophys 163:117–131

Persson C, Bacher P, Shiga T, Madsen H (2017) Multi-site solar power forecasting using gradient boosted regression trees. Sol Energy 150(7):423–436

Qian M, Ma XP, Zhou Y (2014) Forecasting of coal seam gas content by using support vector regression based on particle swarm optimization. J Natl Gas Sci Eng 21:71–78

Seçkin K, Aytaç A, Stelios B, Wasim A (2020) A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series. Energy 212(1):118750

Shanjun M (1998) Research in data model of coal mine GIS. Acta Geodaetica Et Cartographica Sinica 27(4):52–58

Speiser JL, Miller ME, Tooze J, Ip E (2019) A comparison of random forest variable selection methods for classification prediction modeling. Expert Syst Appl 134(11):93–101

Teresa BL, Wilson RO (2013) Particle swarm optimization of MLP for the identification of factors related to common mental disorders. Expert Syst Appl 40(11):4648–4652

Wang W, Cheng YP, Wang HF, Liu HY, Wang L (2015) Fracture failure analysis of hard–thick sandstone roof and its controlling effect on gas emission in underground ultra-thick coal extraction. Eng Fail Anal 54:150–162

Wang W, Peng L, Wang XC (2018) Prediction of coal mine gas emission quantity based on grey-gas geologic method. Math Probl Eng 4397237

Wang XL, Liu J, Lu JJ (2011) Gas emission quantity forecasting based on virtual state variables and Kalman filter. J China Coal Soc 36(1):80–85

Xiao P, Xie XJ, Shuang HQ, Liu CY, Wang HN, Xu JC (2020) Prediction of gas emission quantity based on KPCA-CMGANN algorithm. China Saf Sci J 30(5):39–47

Xie K, Yi H, Hu GY, Li LX, Fan ZY (2019) Short-term power load forecasting based on Elman neural network with particle swarm optimization. Neurocomputing 416:136–142

Xu YS, Qi CY, Feng SC (2019) Gas emission prediction model based on IGSA-BP network. J Electron Meas Instrum 33(5):111–117

Xue XH, Xiao M (2017) Deformation evaluation on surrounding rocks of underground caverns based on PSO-LSSVM. Tunn Undergr Space Technol 69:171–181

Yan H (2020) The study on prediction method of gas emission amount AQPSO-RBF in fully mechanized mining face ad its application. Xi'an, China, Xi'an University of Science and Technology

You W, Yang Z, Ji G (2014) Feature selection for high-dimensional multi-category data using PLS-based local recursive feature elimination. Expert Syst Appl 41(1):1463–1475

Yuan DC, Yue XG, Wang C, Zhang JF (2013) Gas emission prediction based on coal mine operating data. In: The 3rd international conference on green power, materials and manufacturing technology and applications (GPMMTA 2013)

Zhang XL, Shan JP, Peng SP (2009) Mathematical geology technique and method for prediction of gas content and emission. J China Coal Soc 34(3):350–354

Zhang YX, Cui NB, Feng Y, Guo DZ (2019a) Comparison of BP, PSO-BP and statistical models for predicting daily global solar radiation in arid Northwest China. Comput Electron Agric 164:104905

Zhang ZT, Han J, Wang XT, Chen HR, Wei GF, Yao ZH (2019b) Soil salinity inversion based on best subsets-quantile regression model. Trans Chin Soc Agric Mach 50(10):142–152

Zhang ZM, Zhang YG (2005) Three grades of gas-geological maps and their application to gas controlling. J China Coal Soc 30(4):455–458

Zhang ZX, Yuan CF (1999) Study on mathematical model of coalbed gas geology used to prediction of mine gas emission. J China Coal Soc 24(4):368–372

Zhao JQ, Yang DG, Wu JX, Meng XL (2021) Prediction of temperature and CO concentration fields based on BPNN in low-temperature coal oxidation. Thermochim Acta 695:178820

Zhou AT, Zhang M, Wang K, Elsworth D, Wang JW, Fan LP (2020a) Airflow disturbance induced by coal mine outburst shock waves: A case study of a gas outburst disaster in China. Int J Rock Mech Min Sci 128:104262

Zhou B, Xu J, Han F, Yan F, Jiao F (2020b) Pressure of different gases injected into large-scale coal matrix: analysis of time–space dependence and prediction using light gradient boosting machine. Fuel 279:118448

Zhou J, Li XB, Shi XZ (2012) Long-term prediction model of rockburst in underground openings using heuristic algorithms and support vector machines. Saf Sci 50(4):629–644