Research on Coal and Gas Outburst Warning from Missing of Samples

Qinghua Chen, Lizhu Yue, Bencong Kou*

School of Business Administration, Liaoning Technical University, Huludao 125105, Liaoning, China

**Corresponding Author.*

Abstract:

In order to improve the prediction accuracy of coal and gas outburst in the case of missing data, a coal and gas outburst prediction model based on MICE_NN interpolation algorithm and improved Pathfinder Algorithm (IPFA) optimized Extreme Learning Machine (ELM) is proposed. Firstly, the correlation analysis of various indicators affected by coal and gas outburst is carried out, and the MICE_NN algorithm is used to interpolate the missing values, which is easy to obtain more sufficient information from incomplete data sets and improve the prediction effect and accuracy of the model. Secondly, the Pathfinder algorithm is jointly improved by introducing the evolutionary boundary constraint processing scheme, Levy flight strategy and group fitness variance strategy to improve its global optimization ability, so as to optimize the relevant parameters of ELM and construct the coal and gas outburst prediction model. Finally, the measured data interpolated by MICE_NN are used as samples for experimental verification, and the proposed algorithm is compared with single machine learning and ensemble algorithms. The results show that the data quality based on MICE_NN interpolation is significantly better than the data without interpolation. The classification accuracy, recall rate and of IPFA_ELM model based on MICE_NN interpolation are significantly higher than those of other comparison models. It provides a new idea and method for coal and gas outburst prediction, and provides a strong reference basis for the next step of gas outburst prevention and control.

Keywords: coal and gas outburst prediction, missing data, multiple imputation by chained neural networks method, improved pathfinder algorithm, extreme learning machine

INTRODUCTION

Coal and gas protrusion refers to a complex, nonlinear protrusion and uncontrolled evolution of catastrophic behavior in which a large amount of coal and rock in the mining space carrying a large amount of gas suddenly rushes into the mining space [1, 2], and its frequent occurrence in China's coal industry has become the number one problem of production safety [2]. With the increasing depth and intensity of mining, the upper plate of the positive fault has become a key factor in the occurrence of coal and gas protrusion [3], which poses a serious threat to the personal safety of coal miners and the production stability of the mine. Therefore, scientific and accurate prediction of coal and gas protrusion accidents is currently a key factor in the production safety of China's coal industry.

So far, numerous academics, both domestically and internationally, have extensively researched coal and gas protrusion prediction, and the coal and gas protrusion prediction methods can be broadly divided into two categories: traditional mathematical models and machine learning models. Traditional mathematical models generally use Fisher's criterion [4, 5], hierarchical analysis [6, 7] and fuzzy comprehensive evaluation method [8] to quantify, assign weights to selected attributes and calculate the correlation degree of protrusion level, so as to carry out comprehensive evaluation of protrusion level in the target area. These methods have more subjective factors and are easily affected by the original data. Among the machine learning methods, neural networks [9] and support vector machines [10] are extensively researched in the realm of coal and gas prominent forecasting. The precision of predicting coal and gas protrusion is enhanced by these algorithms, but the model training process needs to set the parameters and models artificially based on experience, and the operation is cumbersome.

In recent years, scholars have begun to use intelligent optimization algorithms combined with machine learning algorithms aiming to enhance the precision of predicting coal and gas protrusion while minimizing manual procedures. Li et al [11] combined fruit fly algorithms and random forest algorithms to construct protrusion prediction models; Wen &Su [12] used chained SVM to interpolate the missing data of coal and gas protrusion and optimized the limit learning machine using whale optimization algorithms to achieve protrusion intensity prediction; Sun [13] used rough set to approximate attributes and particle swarm algorithm to optimize SVM to establish a prediction model; Wu et al [14] combined genetic algorithm and simulated annealing algorithm to optimize BP neural network to achieve protrusion intensity prediction.

The above research results reveal that coal and gas protrusion is a complex nonlinear problem from different methods, and improve the prediction accuracy of coal and gas protrusion to a certain extent. However, after the occurrence of coal and gas herniation accidents, the detection equipment is destroyed, resulting in a small amount of accident data and sometimes missing. The small amount of this data leads to overfitting of the prediction model and the inherent defects of the model: the genetic algorithm and the whale algorithm need to regulate more parameters; the generalization ability of the SVM is weak; the FOA and PSO algorithms converge quickly and are prone to fall into the local optimum, etc. All of them will affect the prediction accuracy and convergence performance to a certain extent.

In view of this, this paper proposes a Multiple Imputation by Chained Neural Networks (MICE_NN)-improved pathfinder optimization algorithm (IPFA) optimization Extreme Learning Machine (ELM) for coal and gas protrusion prediction model. Firstly, correlation analysis and MICE_NN interpolation are performed on the missing samples to obtain the input variables; secondly, to address the problems of the PFA algorithm's individual transgressions and the tendency to fall into the local optimum, the transgressions are handled by the evolutionary boundary constraints processing mechanism, and the Lévy flights and the population fitness variance strategy are used to help the algorithm jump out of the local optimum and improve the global optimization seeking ability of optimization and the improved IPFA is used to optimize the ELM's hyper-optimization algorithm. IPFA optimizes the hyperparameters of ELM, and constructs the coal and gas protrusion prediction model of IPFA_ELM. Example verification and comparison with other model analysis, the predictive model introduced in this document exhibits greater precision.

MODEL DESIGN AND IMPLEMENTATION

Mice_Nn

Missing data is an important issue in data quality research [15], especially for coal and gas protrusion data, due to the small sample size, accident data is difficult to obtain makes the interpolation of missing data becomes particularly important. In this paper, through correlation analysis, we found that there are some nonlinear relationships between coal and gas protrusion data, and the correlation coefficients of its factors are shown in Table 1.

Table 1. Pearson correlation coefficient of each factor

	gas content	gas pressure	porosity	Coal seam firmness factor	Gas Discharge Initial Velocity
gas content	1	0.531	0.102	-0.373	0.675
gas pressure	0.531	1	0.004	-0.117	0.432
porosity	0.102	0.004	1	-0.119	0.106
Coal seam firmness factor	-0.373	-0.117	-0.119	1	-0.433
Gas Discharge Initial Velocity	0.675	0.432	0.106	-0.433	1

Neural networks have strong nonlinear fitting ability and perform well in geological missing data interpolation [16]. Aiming at the characteristics of coal and gas protrusion data, this paper proposes a chain neural network multiple interpolation method, the whole interpolation process is shown in Figure 1, which contains three core steps of transmission, training and fitting and complementation, the specific process:

Step 1: MiceImputer() function passes data multiple times.

Step 2: MLPRegressor() The function is trained to fit the observable data using a single hidden layer forward neural network.

Step 3: Imputer(). fit_transform() The function fills the missing matrix by integrating the results obtained in the previous step into a set of results according to the principle of optimality.

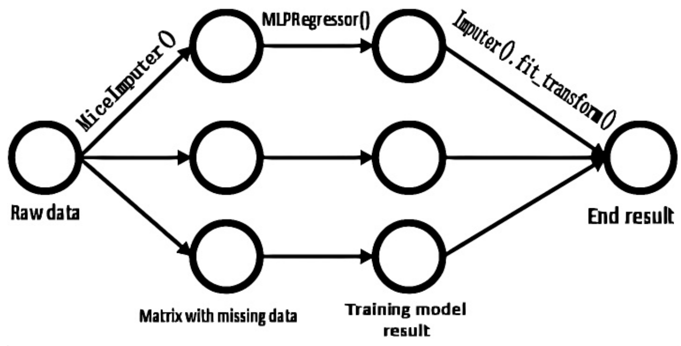

Figure1. Interpolation process of MICE_NN method

ELM

ELM functions as an algorithm for learning neural networks with a feed-forward approach that consists of input, hidden and output layers. Compared with traditional feed-forward neural networks, ELM is randomly initialized at the connection weights and hidden layer thresholds, and adopts the interval-minimizable activation function to compute the hidden layer outputs, and obtains the connection weights through simple matrix operations. The benefits of ELM include its straightforward design and rapid learning capabilities, and this document suggests employing the ELM model for forecasting coal and gas outflow. By introducing the improved PFA algorithm to solve the ELM's and the problem of weak model generalization ability and unstable prediction effect caused by random initialization, and then improve its prediction accuracy in coal and gas herniation.

PFA

The PFA algorithm is a novel heuristic algorithm inspired by the behavior of group animal communities and their leadership systems [17], which performs well in terms of global and local search capabilities by collaboratively searching for a globally optimal solution through communication between two roles, the pathfinder and the follower. The algorithm has the advantages of being easy to understand, high performance, and easy to operate and implement, and has certain advantages in parameter optimization applications. However, PFA adopts the treatment of pulling back the boundary for the transgressing individuals, and this operation causes the transgressing individuals to gather at the solution boundary, which is prone to affect the algorithm's population diversity and reduce the speed of convergence; the follower tends to learn from the pathfinder in PFA, which is prone to fall into the local optimum problem.

IPFA Design

Evolutionary boundary constraint handling scheme

The PFA algorithm adopts the way of pulling back at the boundary to deal with the transgressing individuals, which causes the transgressing individuals to gradually gather at the solution boundary in the iterative process, so that the diversity of the population will be reduced and the rate at which the algorithm converges will likewise be influenced. In this paper, we introduce the evolutionary boundary constraint processing mechanism proposed in the literature [18] to process the transgressing individuals in order to improve the performance of the algorithm. The processing method is as follows:

$$X_i^{`} = \begin{cases} a_1 \times lb_i + (1 - a_1) \times X_p, X_i < lb_i \\ a_2 \times ub_i + (1 - a_2) \times X_p, X_i > ub_i \end{cases} \tag{1}$$

Where X_i' represents the individual's location post-border crossing and X_i is the current position of the individual that crossed the border, a_1 and a_2 represent arbitrary figures within the range [0,1], ub and lb represent the population's maximum and minimum limits, in that order, X_p is the current location of the pathfinder.

Lévy flight strategy

In the PFA algorithm, followers will gradually gather around the pathfinder due to its erroneous guidance, and fall into a local optimum. To avoid this, this paper introduces a random walk strategy with a Lévy distribution proposed in the literature [19]—the Lévy flight strategy—to perturb the position of the pathfinder and reduce the probability of the algorithm falling into a local optimum. The perturbation formula is as follows:

$$Levy(\beta) = 0.05 \times \frac{\mu}{|\nu|^{\frac{1}{\beta}}} \tag{2}$$

Among them, $\beta = 1.5$, and μ and ν obey a normal distribution, as expressed in Equation (3).

$$\mu \sim N(0, \sigma_x), \nu \sim N\left(0, \sigma_y\right)$$

$$\sigma_\mu = \left[\frac{\Gamma(1+\beta)\, sin\left(\frac{\pi\beta}{2}\right)}{\Gamma(\frac{1+\beta}{2}) \times \beta \times 2^{\frac{\beta-1}{2}}}\right]^{\frac{1}{\alpha}}, \sigma_\nu = 1 \tag{3}$$

Variance strategies for group fitness

In the PFA algorithm, followers only learn from the pathfinder, and the search is prone to local optima. To this end, this paper introduces the group adaptability variance indexσ^2proposed in the literature [20] during the IPFA iterative update phase, which is a metric for judging the search state of IPFA. The calculation formula of$\sigma^2$is as follows:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}[f_i - f_\alpha]^2 \tag{4}$$

Where: N represents the size of the group of pathfinders;$f_i$is the classification accuracy of individual i, %; and$f_\alpha$is the average classification accuracy of the group, %.

σ^2 can be used to determine the degree of fluctuation in the individual positions in the IPFA algorithm. A large fluctuation indicates that IPFA is in a global search state, while the opposite indicates that it is in a local or global convergence state. The IPFA search status is determined based on a comparison of the metric values$\sigma^2$and$\theta$with the thresholds, and the corresponding search operation is performed. If $\sigma^2 > \theta$, it is judged that the position fluctuation of each individual is relatively large, and IPFA will continue to perform a global search. If $\sigma^2 \leq \theta$, it is judged that the position fluctuation of each individual is almost zero, and IPFA will locally converge. The two-point crossover operation of the genetic algorithm will be introduced to locally update the position of each individual, so that IPFA can promptly and effectively escape from the premature convergence state.

Construct the IPFA_ELM Model

Input: training dataset, maximum number of iterations T, population size N, variance threshold for population adaptation θ, number of neurons in the hidden layer l and activation function g(x).

Output: IPFA optimizes ELM w and b.

(1) Randomly initialize the position of each individual according to the ELM-optimized object. The position is a K-dimensional vector with respect to w and b, and its value range is [-1, 1].

$$K = s \times l + l \tag{5}$$

Where: s is the number of nodes in the ELM input layer.

(2) Select the F1 index as the model fitness function as in Equation (6), record the location of the individual with the smallest fitness value and set it as the pathfinder.

$$Fitness = 2 - F1_{train} - F1_{test} \tag{6}$$

Where: $F1_{train}$is the score of the training set prediction model F1; $F1_{test}$is the score of the test set prediction model F1; and Fitness is the model fitness function.

(3) The pathfinder position is iteratively updated. The pathfinder position is updated according to Equation (7) and (8), and the out-of-bounds processing of the updated pathfinder position is performed using Equation (1).

$$X_p^{t+1} = X_p^t + 2r_1(X_p^t - X_p^{t-1}) + A + Levy(\beta) \tag{8}$$

Where: t represents the current iteration count; X_p^{t+1}, X_p^t, and X_p^{t-1} represent the updated position of the pathfinder, the current position, and the position of the previous generation of pathfinders, respectively;$r_1$is a random variable uniformly generated in the range [0, 1], representing the pathfinder's movement step factor; and A is a set of perturbation vectors, representing the randomness of the pathfinder's updated position.

$$A = u_1 \cdot e^{\frac{-2t}{T}} \tag{8}$$

where T is the maximum number of iterations and $u_1$is a random vector in the range [-1, 1].

(4) Update the follower position according to Equations (9) to (11), and use Equation (1) to process the updated follower position for out-of-bounds processing.

$$X_i^{t+1} = X_i^t + \alpha \cdot r_2 \cdot (X_j^t - X_i^t) + \beta \cdot r_3 \cdot (X_p^t - X_i^t) + \varepsilon \tag{9}$$

where X_i^{t+1}, X_p^t, X_i^t, and X_j^trepresent the updated position of follower i, the current position of the pathfinder, the current position of follower i, and the current position of follower j, respectively; $r_2$and$r_3$are random variables uniformly generated in the range [0, 1], which represent the step length factors for moving with other followers other followers and the pathfinder; $\alpha$and $\beta$are the interaction coefficient between followers and the attraction coefficient of the pathfinder to the followers, both of which are random numbers in the interval [1, 2]; ε is the perturbation vector, which provides random movement for all followers.

$$\varepsilon = (1 - \frac{t}{T}) \cdot u_2 \cdot D_{ij} \tag{10}$$

where $u_2$is a random vector in the interval [-1, 1] and $D_{ij}$is the distance between followers i and j.

$$D_{ij} = \left\| X_i - X_j \right\| \tag{11}$$

Where$X_i$represents the position of follower i and X_j represents the position of follower j.

(5) Calculate the group fitness variance σ^2. If $\sigma^2 \leq \theta$, use the genetic algorithm's two-point crossover operator to locally update the positions of each individual. Otherwise, continue the global search. Use the updated positions of each individual as the initial positions of each individual for the next iteration. Repeat steps 3-5).

(6) Output the PFA-optimized w and b and train the ELM to train and fit the coal and gas outburst data.

Figure 2 displays the IPFA_ELM process for predicting coal and gas outbursts, utilizing MICE_NN.

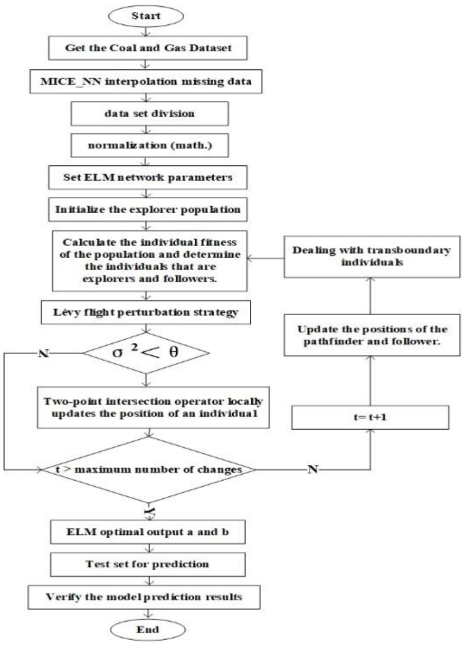

Figure 2. Forecasting methods for coal and gas eruptions

CASE STUDY

Reflecting the actual situation of the Huainan mining area, combined with the research results of previous generations [10, 13]and following the principles of scientificity, systematicness and feasibility, the following five factors were selected as the parameters for predicting coal and gas outbursts: X1: gas content, in m3/t; X2: gas pressure, in is MPa; X3: porosity, unit is %; X4: coal seam strength coefficient; and X5: initial gas release rate, unit is mL/s. In addition, the severity of coal and gas eruptions is categorized into two tiers: 0 (absence of outburst) and 1 (exposure).

The study chooses 133 distinct sets of sample data concerning coal and gas eruptions, including 71 groups of non-accident data (no missing values) and 62 groups of accident data (including missing values). Table 2 displays a statistical overview of the accident data.

Table 2. Statistics on description of accident data

Indicator	Number of groups	Missing number	Min	Max	Average value	Standard deviation
X1	62	0	7.12	26.00	12.15	4.04
X2	62	0	0.28	4.54	1.86	1.05
X3	48	14	2.94	9.60	5.70	1.70
X4	47	15	0.12	2.00	0.55	0.35

Data Interpolation

Data interpolation is required when the missing rate in the data sample exceeds 15% [12]. If only complete data samples are used for training and prediction, this action will diminish the volume of crucial data present in the initial dataset, thereby affecting the training and prediction accuracy of the model. As can be seen from Table 2, the missing rates of the X3, X4, and X5 indicators are 22.59%, 24.19%, and 17.74%, respectively. Therefore, the MICE_NN method proposed this study utilized interpolated data to supplement the absent information on coal and gas eruptions, which were then employed in training and evaluating the predictive model. Table 3 displays the detailed statistical analysis of the completed data set.

Table 3. Statistical analysis of the data post-interpolation

Missing indicator	Treatment method	Number of groups	Average value	Standard deviation
X3	RD	48	5.70	1.70
	Mice_NN	62	5.40	1.70
X4	RD	47	0.55	0.35
	Mice_NN	62	0.59	0.30
X5	RD	51	9.90	4.70
	Mice_NN	62	10.00	4.40

A comparison with the raw data (RD) shows that the difference between the mean and standard deviation [21] of the data after MICE_NN imputation and the original data is almost negligible, and the distributions before and after imputation can be considered to be approximately the same.

Parameter Optimization

The PFA_ELM and IPFA_ELM models use 120, 60, 100 and θ uses 10-4 as parameters, and the activation function employs the Sigmoid function. In the ELM algorithm, optimizing the number of neurons in the hidden layer is crucial to improving the prediction effect. To this end, the author refers to the ideas of the “trial and error method” and the “grid search method” to select the number of neurons in the hidden layer of the ELM model between the intervals [1, 50]. To reduce the randomness of the optimization, the author compared the classification accuracy of each parameter setting through five repeated trials, and finally determined that the optimal number of neurons in the hidden layer is 20. The changes in the individual fitness of the populations of the PFA and IPFA algorithms during the optimization process are shown in Figure 3.

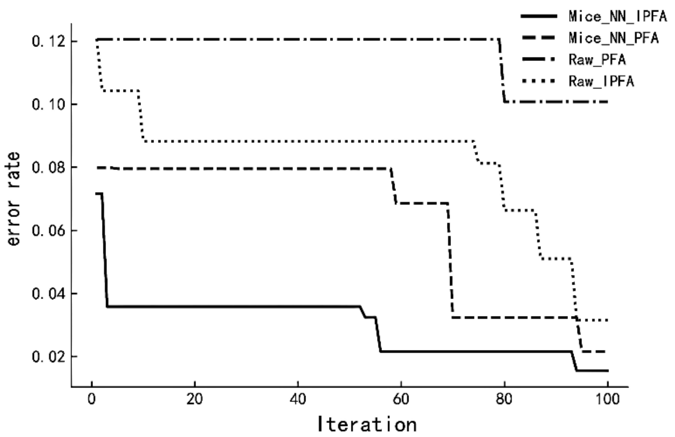

Figure 3. Adaptation change before and after interpolation

As can be seen from Figure 3, under the same experimental sample conditions, IPFA outperforms PFA in optimization efficiency and possesses a lower fitness score, suggesting IPFA's superior capacity to leap beyond local peaks compared to PFA. In addition, by observing the fitness of the four coupled models in Figure 3, it can be found that the IPFA algorithm with interpolated samples performs best, with the smallest fitness value of 0.015589 and a faster convergence rate. The findings indicate that the MICE_NN interpolation technique introduced in this research is capable of significantly enhancing the data volume in coal and gas outburst samples, and the IPFA optimization algorithm demonstrates superior overall optimization potential and fast convergence speed.

Model Evaluation Criteria

In this paper, the prominent and non-prominent categories in the dataset are marked using the positive sample (category 0) and negative sample (category 1), respectively. The classification is shown in Table 4. True positive (True Positive) indicates that the prediction result of coal and gas outburst and the real result are both the number of outburst; True negative (True Negative) indicates that the prediction result of coal and gas outburst and the real result are both the number of no outburst; False Positive (False Positive) refers to the count of non-outburst samples identified as outburst samples, whereas False Negative (False Negative) denotes the tally of outburst samples deemed non-outburst samples.

Table 4. Classification of coal and gas outburst prediction models

Prediction category	Real category	
	Protrusion	No protrusion
Projecting	True Positive (TP)	False Positive (FP)
No protrusion	False Negative (FN)	True Negative (TN)

To evaluate the performance of the prediction model for identifying coal and gas outbursts, the accuracy (Precision), recall (Recall), accuracy (Accuracy) and value F1 were chosen to serve as the evaluative metrics for forecasting coal and gas. The calculation formulas are as follows:

$$Precision = \frac{TP}{TP+FP} \tag{12}$$

$$Recall = \frac{TP}{TP+FN} \tag{13}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{14}$$

$$F1 = 2 * \frac{Recall * Precision}{Recall + Precision} \tag{15}$$

A higher precision indicates that the model has a stronger ability to predict samples without outliers; a higher recall indicates that the model has a stronger ability to predict samples with outliers; the accuracy intuitively reflects the quality of the classification effect; the F1 value is the result of comprehensively considering the precision and recall rates, and is usually used to evaluate the effectiveness of the classification model. This paper mainly uses it to characterize the learning effect of the model with the number of training iterations.

To verify the validity and practicality of the proposed model, the final experimental Precision, Recall, Accuracy and F1 are the average values of the classification results of 20 coal and gas outbursts, and multiple experiments are performed to reduce the noise impact of an abnormal experiment.

IPFA_ELM Model Prediction Results

To verify the effectiveness of the MICE_NN interpolation method and IPFA-optimized ELM, the author constructed interpolated and un-interpolated samples of the dataset. Both experimental specimens were arbitrarily split into training and testing groups in a ratio of 7 to 3. The IPFA_ELM model and PFA_ELM model were used to perform parameter optimization and compare the prediction results under the two sample conditions. The comparison of the model prediction results is shown in Table 5.

Table 5. Predicted results before and after interpolation

Model	Data set	Accuracy/%	Precision/%	Recall/%	F1 value/%	Accuracy of prediction with or without protrusion/%	
						Protrusion	No protrusion
PFA_ELM	Before interpolation	90.63	90	81.82	85.71	81.82	95.24
	After interpolation	92.5	94.44	89.47	91.89	89.47	95.24
IPFA_ELM	Before interpolation	93.75	100	81.82	90	81.82	100
	After interpolation	97.5	100	94.74	97.3	94.74	100

Table 5 shows that after MICE_NN interpolation, the average prediction accuracy of the PFA_ELM and IPFA_ELM algorithms for the overall coal and gas data sets is improved by 1.87% and 3.75%, respectively; the values are improved by 7.65% and 7.3%, respectively; and the prediction accuracy for the prominent sample levels is significantly improved. Findings indicate the efficacy of MICE_NN interpolation in enhancing the forecasting accuracy of the prediction model.

Furthermore, by comparing the performance of the IPFA_ELM and PFA_ELM models in predicting coal and gas outburst data samples, it was found that the ELM model optimized by IPFA outperforms the PFA optimization algorithm in several performance metrics. This shows that the MICE_NN-IPFA_ELM model has better prediction capabilities for coal and gas outburst risk prediction.

To further verify the advantages of IPFA, the prediction results of the IPFA_ELM model and the PFA_ELM model were compared and analyzed under the condition of data set interpolation. The prediction results of the two models are shown in Figure 4.

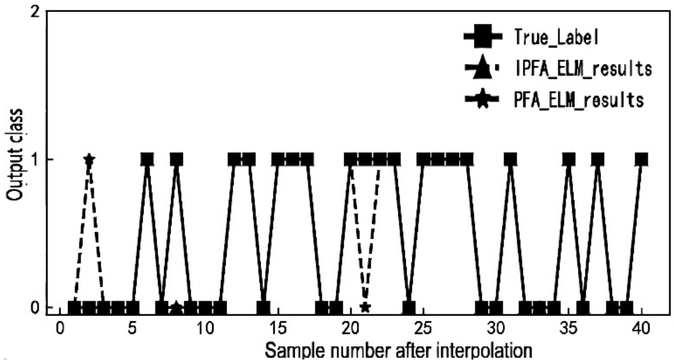

Figure 4. Prediction effects of different models

As can be seen from Figure 4, using the IPFA_ELM model to predict the data of the 40 test sets, only one sample does not match the actual situation, while the PFA_ELM model has three samples that do not match the prediction results. The research results show that the optimization method of IPFA significantly enhances the predictive ability of the ELM algorithm. Compared to the PFA_ELM model, the IPFA_ELM model demonstrates superior predictive performance and overall better results.

IPFA_ELM Model Prediction Results

In order to further demonstrate the superiority of the IPFA_ELM model in predicting coal and gas outburst risks compared to other models, we conducted a comparative analysis of the prediction performance of the IPFA_ELM model, SVM, ELM, IPSO_SVM, IPSO_ELM, and PFA_ELM models under the condition of dataset imputation. The number of misclassifications of outburst levels for different models can be found in Table 6, while the prediction results of different models can be viewed in Figure 5.

Table 6. The number of misjudgments of outstanding level

Model	Average accuracy/%	Number of false predictions for different levels of hazard	
		No protrusion	Protrusion
SVM	85	0	6
ELM	87.5	1	4
IPSO_SVM	87.5	1	4
IPSO_ELM	92.5	2	1
PFA_ELM	92.5	1	2
IPFA_ELM	97.5	0	1

Table 6 shows that the IPFA_ELM model has the fewest number of predictions with prominent numbers and the fewest number of prominent prediction errors. Among them, the number of prominent misjudgments is 1, which is 4 less than the average of the single machine learning algorithms (SVM, ELM) and 1 less than the average of the ensemble algorithms (IPSO_SVM, IPSO_ELM, PFA_ELM). The model in this paper has stronger robustness in predicting coal and gas outbursts.

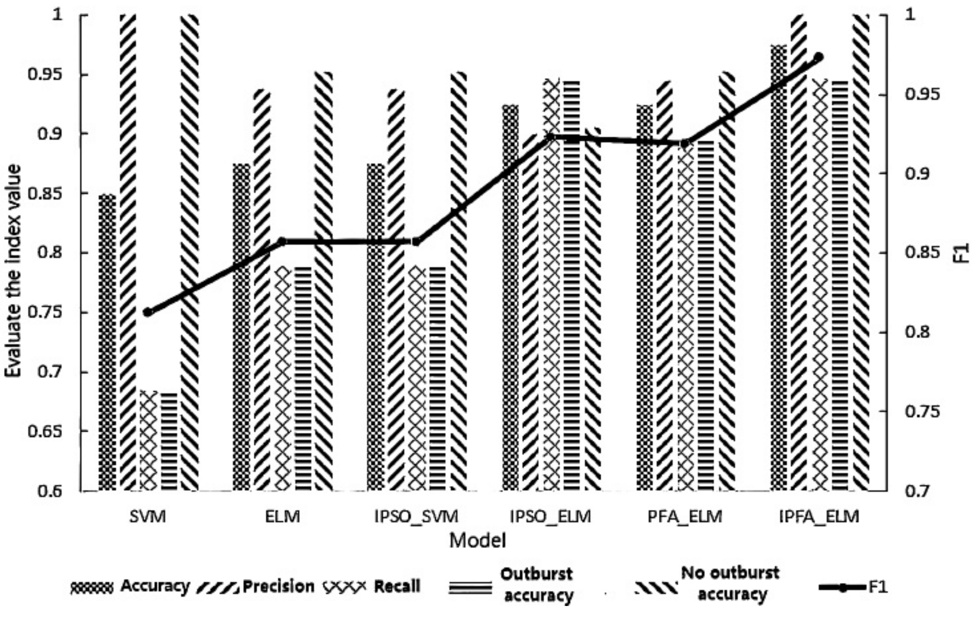

Figure 5. Comparison of prediction effects of various models

According to the results in Figure 5, the single machine learning algorithm performed poorly on the six indicators for coal and gas outburst prediction, while the integrated algorithms IPSO_SVM, IPSO_ELM, PFA_ELM and IPFA_ELM had better prediction results and generalization capabilities. This shows that the single machine learning algorithm is less adaptive in terms of coal and gas outburst prediction, and that the performance of the model can be improved by integrating intelligent optimization algorithms. Among the four integrated algorithms of IPSO_SVM, IPSO_ELM, PFA_ELM and IPFA_ELM, IPFA_ELM has the best prediction effect, and its results of the six comparison indicators are better than those of the other prediction models. Therefore, the IPFA-optimized ELM model based on MICE_NN interpolation has better stability and prediction effect in the prediction of coal and gas outbursts.

CONCLUSION

(1) The MICE_NN algorithm proposed in this study effectively addresses the issue of nonlinear missing values in coal and gas outbursts, significantly increasing the amount of information in coal and gas outburst sample data. At the same time, the algorithm avoids the problem of insufficient model training that could arise from relying solely on complete data samples, thus improving the accuracy of predictions for coal and gas outbursts, both with and without occurrences, as well as the overall prediction accuracy. The research findings reveal that the MICE_NN algorithm demonstrates a high level of effectiveness in handling missing data imputation issues.

(2) In the traditional PFA algorithm, the evolutionary boundary constraint processing scheme can effectively deal with individuals in the population that cross the boundary. At the same time, the Lévy flight strategy and the group fitness variance strategy introduced can effectively prevent the PFA algorithm from falling into “premature” convergence. Under the condition of data set interpolation, the IPFA_ELM model has significantly improved the six comparison indicators and has a better prediction effect than the PFA_ELM prediction model. The results show that IPFA-optimized ELM has high effectiveness and can effectively improve the prediction performance of the ELM algorithm.

(3) The IPFA_ELM model with and without prominent prediction of coal and gas with complex nonlinear characteristics is superior to traditional single machine learning and intelligent optimization integration algorithms. The research results reveal that the IPFA_ELM model based on MICE_NN can effectively enhance the prediction accuracy of coal and gas outburst events, while also demonstrating higher model generalization performance and more robust prediction outcomes.

ACKNOWLEDGMENT

This research was funded by the National Natural Science Foundation of China, grant number 71771111.

REFERENCES

[1] Yunpei Liang, Menghao Zheng, Quangui Li, et al. A review on prediction and early warning methods of coal and gas outburst. Journal of China Coal Society, vol:48, no:08, pp:1-24, 2023.

[2] Quanbin Mou. Study status and prospects for regional prediction methods of coal and gas outburst in China. Coal Science and Technology, vol:42, no:11, pp:59-63, 2014.

[3] Yunxing Cao, Haiyang Zhang, Zhen Zhang, et al. Characteristics of coal and gas outburst and controlling mechanism of stress field in the hanging wall of normal faults. Coal Geology & Exploration, vol:52, no:4, pp:61-69, 2022.

[4] Changxing Li, Chengpeng Xin, Huigui Li, et al. Prediction of coal and gas outburst based on Fisher-stepwise discriminant analysis method. China Mining Magazine, vol:29, no:07, pp:106-111, 2020

[5] Jianping Chen, Jun Dong, Xiangwei Lyu. Research on risk level of coal and gas outburst prediction based on PCA-Fisher discriminant analysis model. Mining Safety & Environmental Protection, vol:45, no:03, pp:61-66, 2018.

[6] Hongwei Jin, Zhuoya Yang, Gang Xu. Evaluation of coal and gas outburst risk based on improved AHP-GRA evaluation model. Mining Safety & Environmental Protection, vol:47, no:05, pp:113-118, 2020.

[7] Enyu Xu, Xijian Li, Feng Xue. Assessment of coal and gas outburst by AHP-GT combined model. Safety in Coal Mines, vol:53, no:04, pp:172-177, 2022.

[8] Zequan Lan, Zhiliang Wang, Yuebing Zhang, et al. Study on coal and gas outburst risk fuzzy comprehensive evaluation method. Safety in Coal Mines, vol:47, no:03, pp:159-163, 2016.

[9] Zhijie Zhu, Hongwei Zhang, Jun Han, et al. Prediction of Coal and Gas Outburst Based on PCA-BP Neural Network. China Safety Science Journal, vol:23, no:04, pp:45-50, 2013.

[10] Xiaoliang Zheng, Wenhao Lai, Sheng Xue. Application of MI and SVM in coal and gas outburst prediction. China Safety Science Journal, vol:31, no:01, pp:75-80, 2021.

[11] Nengzeng L, Mei Y, Guanliang W. Application of FOA-RF model in prediction of coal and gas outburst intensity//Proceedings of the 2019 International Conference on Robotics, Intelligent Control and Artificial Intelligence, pp:379-383, 2019.

[12] Tingxin Wen, Huanbo Su. Woa-ELM prediction model of coal and gas outburst based on multiple imputation by chained equations. Journal of Safety Science and Technology, vol:18, no:07, pp:68-74, 2022.

[13] Liyuan Sun. Application resrarch on the coal and gas outburst prediction based on gray correlation analysis and PSO-SVM. Xuzhou: China University of Mining and Technology, vol:47, no:05, pp:113-118, 2020

[14] Yaqin Wu, Ronglei Gao, Jinzhen Yang. Prediction of coal and gas outburst: A method based on the BP neural network optimized by GASA. Process Safety and Environmental Protection, vol:133, pp:64-72, 2020

[15] Jianxin Deng, Lubao Shan, Deqiang He, et al. Processing method of missing data and its developing tendency. Statistic & Decision, vol:35, no:23, pp:28-34, 2019

[16] Lingling Zhang, Guoqing Li, Guangcheng Jiang, et al. A BP Neural Network based method for geological missing data processing. Gold Science and Technology, vol:23, no:05, pp:53-59, 2015.

[17] Yapici H, Cetinkaya N. A new meta-heuristic optimizer: Pathfinder algorithm. Applied soft computing, vol:78, pp:545-568,2019.

[18] Gandomi A H, Yang X S. Evolutionary boundary constraint handling scheme. Neural Computing and Applications, vol:21, pp:1449-1462, 2012.

[19] Mantegna. Fast, accurate algorithm for numerical simulation of Lévy stable stochastic processes. Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics, vol:49, no:05, pp:4677-4683, 1994.

[20] Dongping Tian, Tianxu Zhang. Adaptive particle s-warm optimization based on colony fitness variance. Computer Engineering and Applications, vol:46, no:18, pp:24-39, 2010

[21] Licheng Chen, Jianhong Chen. Study on prediction effect of coal and gas outburst based on data imputation and machine learning. Journal of Safety Science and Technology, vol:18, no:09, pp:69-74, 2022.