Article

Coal and Gas Outburst Prediction Model Based on Miceforest Filling and PHHO–KELM

Liangshan Shao [1,2] and Wenjing Chen [1,*]

1   College of Business Administration, Liaoning Technical University, Huludao 125105, China; lntushao@163.com
2   Liaoning Institute of Science and Engineering, Jinzhou 121013, China
*   Correspondence: chenwenjing97@163.com

Abstract: Coal and gas outbursts are some of the most serious coal mine disasters, and effective prediction of coal and gas outbursts can reduce the likelihood of accidents and fatalities. Previously conducted studies have established that machine learning has achieved results in the prediction of coal and gas outbursts, but there is a problem that the available accident data of coal and gas outbursts are diminished or missing. This paper proposes a prediction model based on multiple filling of chained equations for random forests (miceforest) and the Harris Hawk optimization algorithm with Piecewise chaos mapping (PHHO) to optimize the kernel extreme learning machine (KELM) to solve the problem of missing data in coal and gas outburst prediction and to improve prediction accuracy in the case of missing data. Firstly, the miceforest algorithm was adopted to fill missing values in the salient samples, and then the PHHO algorithm was used to optimize the parameters of KELM. Finally, the datasets before and after filling were input into the PHHO–KELM model for experimentation and comparison with other models. The results show that miceforest filling is effective in improving the salient sample accuracy and overall accuracy of predictions, but the improvement is not significant for non-salient samples. The use of the PHHO–KELM model can effectively avoid falling into a local optimum and further improve the prediction accuracy of the KELM algorithm. The salient sample accuracy and overall accuracy of the miceforest–PHHO–KELM model prediction are 96.77% and 98.50%. And an effective coal and gas outburst model has been proposed, which is the miceforest–PHHO–KELM model.

Keywords: coal and gas outburst prediction; missing data filling; multiple filling of chained equations for random forests; kernel extreme learning machine; Harris Hawk optimization algorithm with Piecewise chaotic mapping

1. Introduction

Coal is the largest energy source in China, and gas outbursts are an important source of danger in coal mining. From 2001 to 2020, there were 484 coal and gas outburst accidents and 3195 deaths in China [1]. Although the number of coal and gas outburst accidents and deaths have decreased in recent years in China, the situation of coal mine safety is still not optimistic. Therefore, we still need to pay great attention to the prevention and control of coal and gas outburst accidents, and to accurately and efficiently predict coal and gas outburst accidents.

Many scholars have carried out a lot of research on the coal and gas outburst problem [2–4]. With the development of science and technology, scholars have introduced machine learning algorithms into coal and gas outburst prediction, and have improved the accuracy of coal and gas outburst prediction by optimizing these machine learning algorithms. As a machine learning algorithm, support vector machine performs well in coal and gas outburst prediction; in order to further improve prediction accuracy, scholars have used the improved particle swarm algorithm [5], adaptive differential

evolution algorithm [6] and improved gravity algorithm [7] to improve and optimize accuracy and achieve a good prediction effect. And scholars have also used optimized neural network algorithms such as the Neuron-Evolution of Augmenting Topologies algorithm [8], the adaptive learning rate to improve BP neural network [9] and the quantum gate line neural network [10] to improve the accuracy of coal and gas outburst prediction. Ref. [11] proposed a gas outburst hazard prediction model based on an adaptive tensile whisker search algorithm to optimize an extreme learning machine, which effectively improves prediction accuracy. Ref. [12] constructed a dynamic prediction model with multiple algorithms and multivariate analysis, which provided a new way for the prediction of coal and gas outburst hazard levels. However, in machine learning, the prediction results of algorithms are more skewed to the category of a large amount of data, and the influence of data on prediction accuracy is much larger than the influence of algorithms. In order to solve this kind of problem, scholars have improved the effectiveness and accuracy of coal and gas outburst prediction by optimizing the known measurable data. Ref. [13] proposed for the first time to use the correlation coefficient to fill missing data in real time and then use the random forest algorithm to achieve coal and gas protrusion prediction. Ref. [14,15] proposed to use the multiple filling method to fill missing data and then use support vector machines to predict coal and gas outbursts, which effectively improve the accuracy of protrusion prediction. Furthermore, the missForest method [16,17] has been used to fill the missing data of coal and gas outbursts and has achieved a good data filling effect, which is of great significance for improving the accuracy of outburst prediction. Ref. [18] proposed a coal and gas outburst prediction model based on chained support vector machine multiple interpolation and a whale optimization extreme learning machine, and the MICE–SVM interpolation algorithm significantly improved the accuracy of protrusion prediction. Although the above model has had a significant effect in improving the accuracy of coal and gas outburst prediction, there is still room for progress, and the existing missing value filling algorithm can be further optimized to improve data accuracy in order to enhance prediction accuracy.

In view of this, the multiple filling of chained equations for random forests (miceforest) is used to fill missing values in datasets to improve data accuracy. In order to make the prediction model more effective, the Harris Hawk optimization algorithm with Piecewise chaotic mapping (PHHO) is used to optimize the parameters of the kernel extreme learning machine (KELM), and a coal and gas outburst prediction model based on miceforest–PHHO–KELM is established. Simulation experiments are carried out using a measured dataset of the Huainan Zhuji mining area, and comparison and analysis are carried out with other models.

2. Selection of Characteristic Variables and Data Preprocessing

2.1. Selection of Characteristic Variables

The factors affecting coal and gas outbursts are complex and diverse, and prediction results derived from the selection of different influencing factors in modeling can be very different. A large amount of energy will be released when a coal and gas outburst accident occurs. Shi Haoyu et al. [19] analyzed the influencing factors of outburst through the ideal gas equation of state and pointed out that gas content and gas pressure have a great influence on the value of outstanding energy release. Wang Gang et al. [20] concluded that gas content had the greatest influence on protrusion through the Morris filtering method, followed by the gas diffusion coefficient, gas pressure and porosity. Zheng Xiaoliang pointed out that porosity, the coal seam coefficient of solidity and the initial velocity of gas dissipation have a greater influence on the desorption rate of gas and whether a certain pressure can be formed [14]. Therefore, combined with measured data from the Huainan Zhuji Mine [14], and with consideration that the gas diffusion coefficient is difficult to measure, five parameters in this dataset, namely, gas content, gas pressure, initial velocity of gas dissipation, coefficient of coal-bed solidity, and porosity of coal, are selected as

prediction indexes. There are 133 sets of data in this dataset, among which there are 62 sets of salient data and 71 sets of non-salient data, and only salient data are missing. The relevant statistics of the salient sample data are shown in Table 1.

Table 1. Relevant statistics of the salient sample data.

Parameter	Gas Content	Gas Pressure	Porosity of Coal	Coefficient of Coal-Bed Solidity	Initial Velocity of Gas Dissipation
Groups	62	62	48	47	51
Missing	0	0	14	15	11
Maximum	26.00	4.54	9.60	2.00	35.00
Minimum	7.12	0.28	2.94	0.12	5.00
Mean	12.15	1.86	5.70	0.55	9.90
Standard deviation	4.01	1.04	1.68	0.35	4.74

2.2. Multiple Filling of Chained Equations for Random Forests (Miceforest)

When it comes to scientific research, complete datasets are more valuable than incomplete ones. Missing data can result in less efficient predictions and tend to produce inaccurate results. Therefore, the problem of data accuracy is one of the main issues that need to be solved urgently to make effective predictions. Compared with other industries, it is difficult to collect data on coal and gas outburst accidents: there are fewer accident data available. Moreover, it is difficult to find relevant pre-accident parameters after an accident, and even if there is a record of parameters before the accident, the record is incomplete. As a result, prediction results tend to be biased towards non-accident data with more data volumes, and the accuracy of prediction is lower, which affects the safety of coal mine production. Therefore, filling in missing data, expanding available datasets and improving data quality cannot be ignored for improvement of the prediction accuracy of coal and gas outbursts.

Scholars at home and abroad have proposed data interpolation methods to fill missing values to solve the problem of missing data [21]. The more common filling methods mainly include regression filling, K-Nearest Neighbor filling (KNN) and random forest filling (RF). Multiple interpolation with chained equations is a series of iterative prediction models to fill missing data. Each iteration randomly selects other variables in the dataset to estimate each specified variable in the dataset, and the iteration is stopped when convergence is satisfied. There is a certain amount of randomness in the process of multiple interpolation of chained equations, and as the missing rate rises, it becomes less and less effective in filling in the missing data. The results obtained from random forest have randomness, which makes it more sensitive to missing values and more resistant to interference, higher interpolation accuracy and better robustness. The combination of random forest and multiple interpolation with chained equations can better improve the accuracy of filling in missing values. Therefore, the miceforest algorithm is chosen to fill missing values in the dataset. The filling process is shown in Figure 1. The specific filling steps are shown below:

- Use the miceforest algorithm to fill missing data of the original dataset m (default m = 4) times to obtain multiple complete datasets.
- Perform mathematical statistical analyses such as the mean change rate for each complete dataset, and tabulate the resulting results.
- In accordance with the principle of taking the optimum, select statistical results from step 2 for integration of the same columns to obtain the final filled complete dataset.
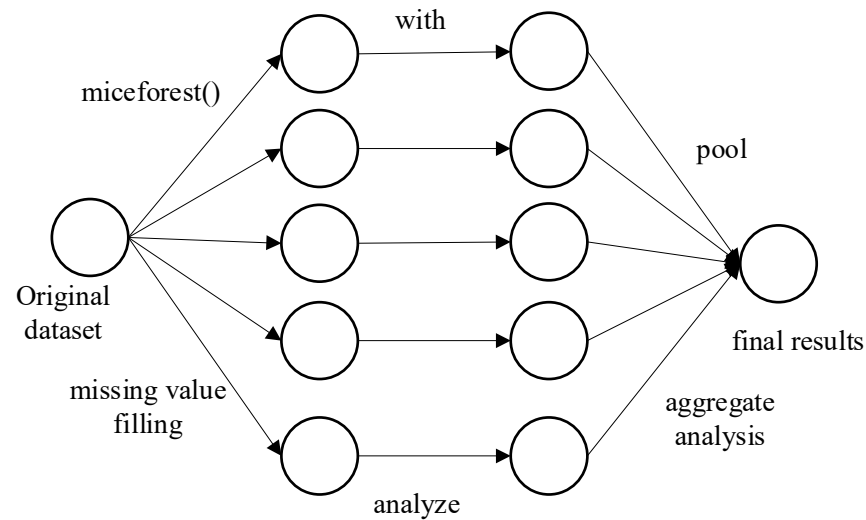
Figure 1. Diagram of the miceforest filling process.

2.3. Filling Missing Data Values with the Miceforest Algorithm

There were 27 groups of 62 coal and gas salient samples with missing data for some factors. There were mainly missing data for the three influencing factors of porosity, coal seam solidity coefficient and initial velocity of gas discharge, and the index of coal seam solidity coefficient had the most amount of missing data. When the missing data rate is greater than 15%, data filling is required. The missing rate for this dataset is 24.19%, which is obviously greater than 15%. If only complete data are used for prediction, model training will be insufficient due to the small amount of data, which will affect the accuracy of recognition whether it is prominent. The miceforest algorithm was used to fill in the missing data, and the datasets before and after filling were used to train and test the prediction model, respectively, to compare prediction results.

In order to verify the advantages of miceforest filling, K-Nearest Neighbor, regression filling and random forest were selected to fill missing values and were compared with the data after miceforest filling. Analysis of the data before and after filling is shown in Table 2. The results show that the degree of change in the mean and standard deviation of the three missing data indicators after miceforest filling is relatively low.

Table 2. Analysis before and after filling missing values.

Missing Value Parameter		Porosity of Coal	Coefficient of Coal-Bed Solidity	Initial Velocity of Gas Dissipation
Missing		14	15	11
Mean	Original data	5.70	0.55	9.90
	Regression	5.70	0.56	9.93
	KNN	5.76	0.54	9.86
	RF	5.75	0.53	9.67
	miceforest	5.83	0.53	9.66
Standard deviation	Original data	1.68	0.35	4.74
	Regression	1.48	0.31	4.32
	KNN	1.53	0.31	4.32
	RF	1.52	0.31	4.37
	miceforest	1.68	0.31	4.37

The root mean square error (RMSE) and the coefficient of determination (R^2) were used to evaluate the overall filling effect. The root mean square error reflects the degree of deviation between the predicted and actual values. The smaller its value, the better the fit.

The numerator of the coefficient of determination is the sum of the squares of the difference between the predicted and real values, and the denominator is the sum of the squares of the difference between the mean and real values. The closer the value of R^2 is to 1, the better the fit of the model. The missing values in the dataset are filled with 0 and the RMSE and R^2 values of different models are calculated; a comparison of the overall filling effects of the different models is shown in Table 3. The two indexes filled by the miceforest algorithm have the best results, which indicates that the algorithm is able to realize compensation for the missing data to meet the system's requirements for data reliability and is able to improve the quality of the data and data accuracy.

Table 3. Comparison of overall filling effect of different methods.

Evaluation Index	Regression	KNN	RF	Miceforest
RMSE	2.264	2.227	2.235	2.136
R^2	0.584	0.599	0.619	0.651

3. Coal and Gas Outburst Prediction Model

3.1. Kernel Extreme Learning Machine

The extreme learning machine (ELM) is a feed-forward neural network algorithm. Although the overall structure of the ELM model is relatively simple, and it can be trained without repeated iterations, which has the advantages of shorter training time and higher generalization ability, the stability and generalization ability of ELM is drastically reduced due to the random generation of input weights and thresholds. Therefore, researchers spend more effort on adjusting weights and thresholds to make the model more accurate. Therefore, Huang et al. [22] introduced the concept of the kernel function into the ELM model and proposed the kernel limit learning machine (KELM) model. This does not need to randomly generate input weights and thresholds, but only needs to be given the kernel function to train and then obtain classification and recognition results. In order to avoid local optimality when predicting coal and gas outbursts, it is proposed to optimize the penalty coefficients and thresholds of KELM with the Harris Hawk optimization algorithm of Piecewise chaotic mapping (PHHO), so as to further improve the effect of KELM on the prediction of coal and gas outbursts.

3.2. Harris Hawk Optimization Algorithm

The Harris Hawk optimization algorithm (HHO) [23–25] is a population-based and gradient-free metaheuristic algorithm proposed by Heidari et al. in 2019. The algorithm simulates the hunting behavior of a hawk flock to search for an optimal solution and consists of three phases: a global search phase, a transition phase and a local development phase.

- Global search phase

The hawks move apart to expand their search range and increase the likelihood of finding prey. Individuals will perch on any location, and the two selection strategies for perching locations are as follows:

$$X(t+1) = \begin{cases} X_{rand}(t) - r_1|X_{rand}(t) - 2r_2X(t)|, & q \geq 0.5 \\ (X_{rabbit}(t) - X_m(t)) - r_3(LB + r_4(UB - LB)), & q < 0.5 \end{cases} \tag{1}$$

where, r_1, r_2, r_3, r_4 and q are all random numbers in the interval [0, 1], X_{rand} denotes a random individual position in the current hawk flock, X_{rabbit} denotes the prey position, UB is the upper limit of the search range, LB is its lower limit, X_m denotes the average position of all the individuals in the current hawk flock, $X_i(t)$ denotes the position of each hawk and N denotes the number of populations, as shown in Equation (2).

$$X_m(t) = \frac{1}{N}\sum_{i=1}^{N} X_i(t) \tag{2}$$

- Transition phase

The hawks choose different hunting strategies depending on changes in the prey's escape energy. At this time, the HHO algorithm will transition from the search phase to the exploitation phase with the following escape energy formula.

$$E = 2E_0(1 - \frac{t}{T}) \tag{3}$$

where E_0 denotes that the escape energy is in the initial state and is a random number in $[-1, 1]$, t denotes the number of current iterations and T denotes the total number of iterations.

- Local development stage

The hawks adopt different pursuit strategies depending on the escape route of the prey. The HHO algorithm uses the following four behavioral strategies to simulate the hawks' roundup behavior. The probability of the prey escaping is denoted by r, where $r < 0.5$ means success and $r \geq 0.5$ means failure.

Soft roundup: when $r \geq 0.5$ and $|E| \geq 0.5$, the prey is energetic and has enough energy to escape, so the hawks use a soft strategy; the formula is as follows:

$$X(t+1) = \Delta X(t) - E|JX_{rabbit}(t) - X(t)| \tag{4}$$

$$X(t) = X_{rabbit}(t) - X(t) \tag{5}$$

where $\Delta X(t)$ denotes the difference between the prey position and the current individual position at iteration t, and J denotes the random escape intensity of the prey in the escape process, which is a random number in the range of $(0, 2)$.

Hard roundup: when $r \geq 0.5$ but $|E| < 0.5$, the prey is exhausted and the escape energy is low, so the hawks adopt a hard strategy; the formula is as follows:

$$X(t+1) = X_{rabbit}(t) - E|\Delta X(t)| \tag{6}$$

Gentle roundup with progressive fast dive: when $r < 0.5$ but $|E| \geq 0.5$, the prey is energetic and has a chance of successful escape, so the hawks adopt a gentle encirclement to launch a surprise attack. Levy flight (LF) was introduced to simulate the deceptive behavior and escape routes of prey during escape, and the position update strategy is shown below:

$$Y = X_{rabbit}(t) - E|JX_{rabbit}(t) - X(t)| \tag{7}$$

$$Z = Y + S \times LF(D) \tag{8}$$

$$X(t+1) = \begin{cases} Y, F(Y) < F(X(t)) \\ Z, F(Z) < F(X(t)) \end{cases} \tag{9}$$

where D denotes the dimension of the problem, S denotes a random vector of size $1 \times D$ dimension, $F(\cdot)$ denotes the fitness function and $LF(\cdot)$ denotes the Levy flight function.

Tough roundup with progressive fast dive: when $r < 0.5$ and $|E| < 0.5$, the prey does not have enough escape energy, so an asymptotically fast-dive tough roundup strategy is used.

$$Y = X_{rabbit}(t) - E|JX_{rabbit}(t) - X_m(t)| \tag{10}$$

$$Z = Y + S \times LF(D) \tag{11}$$

$$X(t+1) = \begin{cases} Y, F(Y) < F(X(t)) \\ Z, F(Z) < F(X(t)) \end{cases} \tag{12}$$

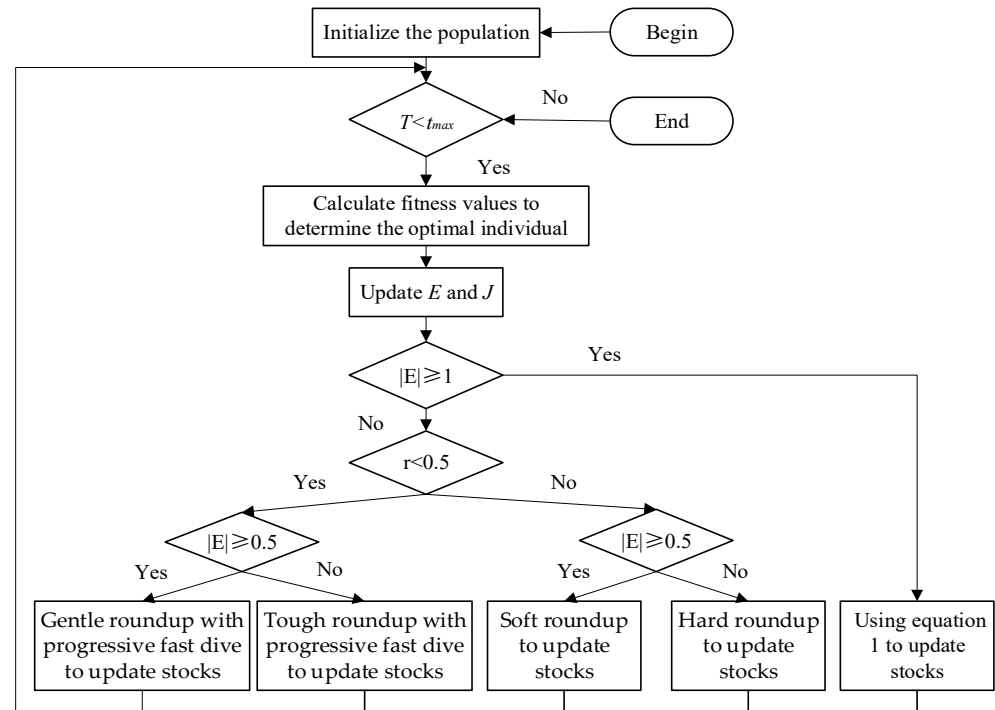To summarize, a HHO algorithm flowchart is shown in Figure 2.

Figure 2. Flowchart of HHO algorithm.

3.3. Piecewise Chaotic Mapping

Piecewise chaotic mapping [26] is a typical representative of chaotic mapping, which is more ergodic and randomized. In order to improve the population diversity of the HHO algorithm and enhance the probability of jumping out of a local optimum, Piecewise chaotic mapping is introduced to optimize the initialized population. The formula of Piecewise chaotic mapping is as follows:

$$x(t+1) = \begin{cases} \frac{x(t)}{p}, & 0 \leq x(t) < p \\ \frac{x(t)-p}{0.5-p}, & p \leq x(t) < 0.5 \\ \frac{1-p-x(t)}{0.5-p}, & 0.5 \leq x(t) < 1-p \\ \frac{1-x(t)}{p}, & 1-p \leq x(t) < 1 \end{cases} \tag{13}$$

3.4. Construction of Coal and Gas Outburst Prediction Model Based on Miceforest–PHHO–KELM

A miceforest–PHHO–KELM model is established, and coal and gas outburst data are used as a dataset for training and testing experiments; a flowchart is shown in Figure 3. The specific steps are as follows.

- Establish before- and after-filling datasets of coal and gas outbursts, respectively. The missing parts of the pre-fill dataset are deleted directly and only the complete parts are kept. The post-fill dataset is used to fill the missing parts of the data using the miceforest algorithm to ensure that all data are complete.

- Initialize all the parameters of the KELM and HHO models. Piecewise chaotic mapping is used to optimize and initialize the population; the number of populations is set to N and the number of iterations is set to T.
- Use the train set as the input vector of KELM to train KELM.
- Calculate the fitness value of each solution and evaluate the fitness to select the best fitness value.
- Update the prey's escape energy, probability of escape and random escape intensity according to the HHO algorithm formula.
- Update the search space selected by the hawks to implement the update adjustment to the individual positions, obtain the new search prey and selection area, and calculate and evaluate the adaptability of its corresponding deterministic solution.
- Find the optimal position and make a record of its corresponding fitness value, and change the number of iterations to t + 1.
- Determine whether the maximum number of iterations is reached. If it is satisfied, obtain the optimal penalty coefficient and threshold value; otherwise, go to step 4.
- Build the PHHO–KELM classification prediction model according to the optimal parameters of the final output, and import the test set to output the recognition results.
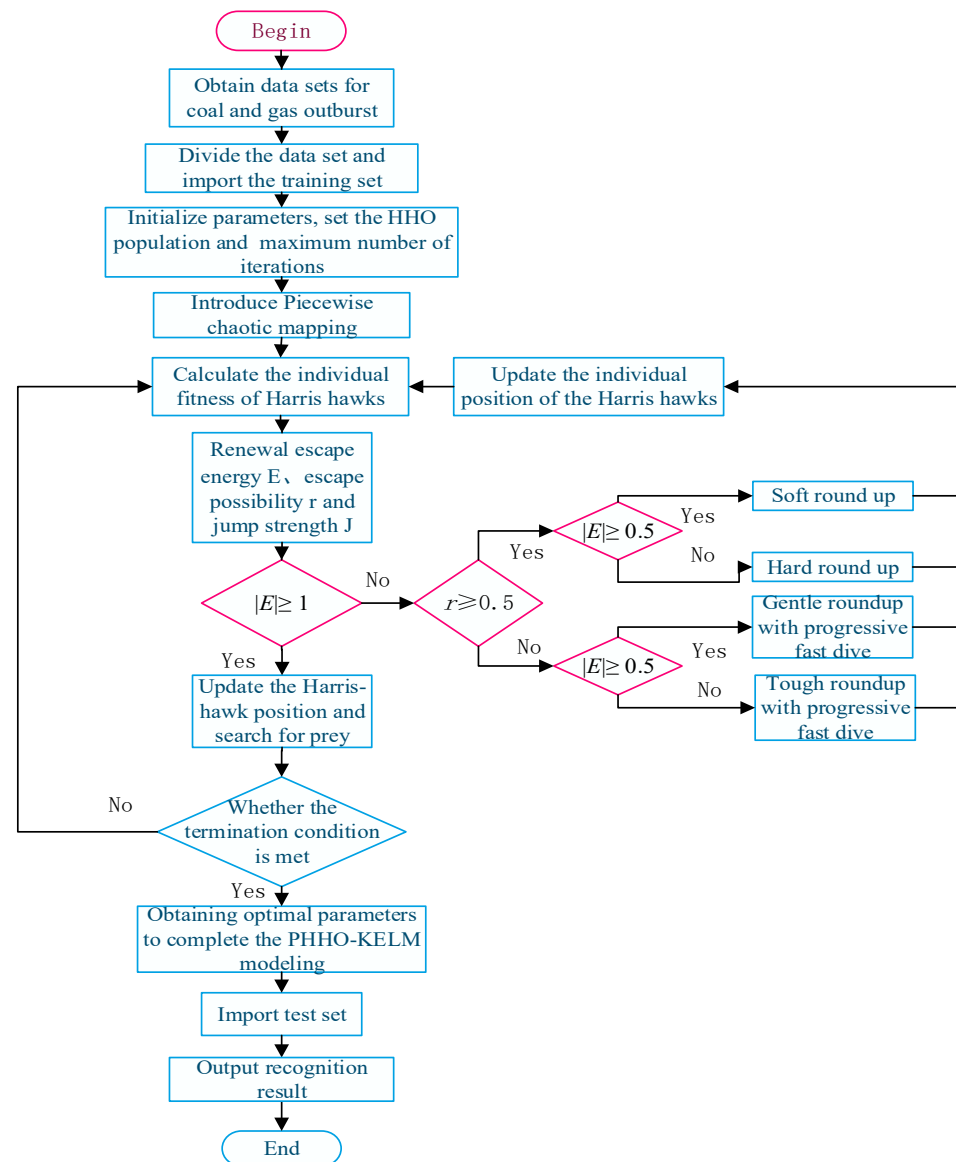

Figure 3. Flow chart of coal and gas outburst prediction by miceforest–PHHO–KELM.

4. Experimental Results and Analysis

Pre-fill and post-fill datasets for prediction testing were constructed; the division ratio of training set to test set was 8:2. PHHO–KELM model prediction was implemented in MATLABR2021b. The PHHO population size was set to 30, the maximum number of iterations was 100 and the radial basis 'RBF_kernel' was used as the activation function. After reading the dataset filled by the miceforest algorithm into the PHHO–KELM model for training, a fitness change curve was obtained, as shown in Figure 4. From the figure, it can be seen that the PHHO–KELM model converges faster than PSO–KELM and HHO–KELM.
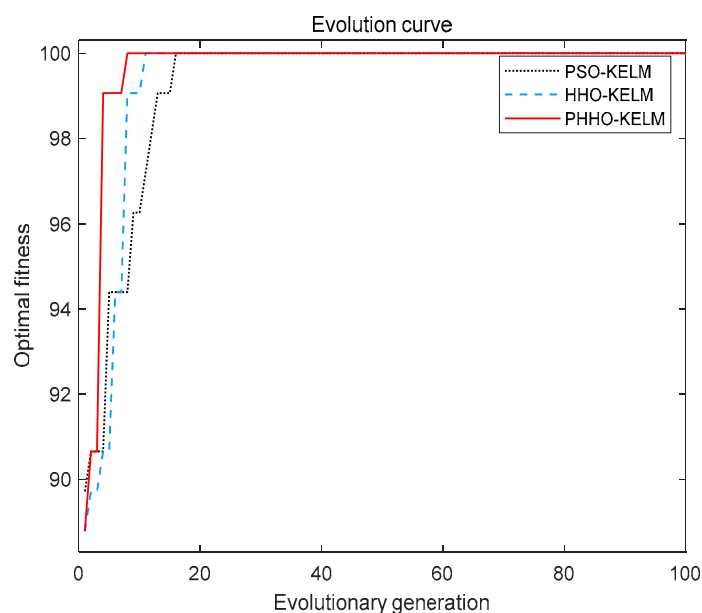

Figure 4. Adaptation change graph.

In order to verify the superiority of PHHO–KELM, the prediction results of SVM, KELM, PSO–KELM, HHO–KELM and PHHO–KELM models in the dataset before and after filling were compared. The SVM parameters were kernal = 'rbf', c = 1.0 and gamma = 0.2; the KELM activation function was 'RBF_kernel'. Test set prediction results for different models after miceforest filling are shown in Figure 5.

As can be seen from Figure 3, the test set prediction of PHHO–KELM was the best, with an accuracy rate of 25/26. The accuracy of SVM was 21/26, and the prediction of KELM was better than that of SVM, with an accuracy of 22/26. The accuracy rate also improved after optimization by PSO and HHO, but the accuracy of KELM optimized by PHHO was the most improved. A comparison of the prediction results of different models before and after filling is shown in Table 4.

Table 4. Comparison of prediction results of different models.

|	Model	SVM	KELM	PSO–KELM	HHO–KELM	PHHO–KELM
	Salient sample prediction accuracy/%	62.86	65.71	74.29	85.71	88.57
Pre-fill	Non-salient sample prediction accuracy/%	91.55	92.96	94.37	95.77	97.18
	Overall prediction accuracy/%	82.08	83.96	87.74	92.45	94.34
	Kappa coefficient	0.5732	0.6180	0.7124	0.8268	0.8702
	Salient sample prediction accuracy/%	83.87	85.48	88.71	93.55	96.77
Post-fill	Non-salient sample prediction accuracy/%	92.96	94.37	97.18	98.59	100
	Overall prediction accuracy/%	88.72	90.23	93.23	96.24	98.50
	Kappa coefficient	0.7722	0.8027	0.8633	0.9242	0.9698
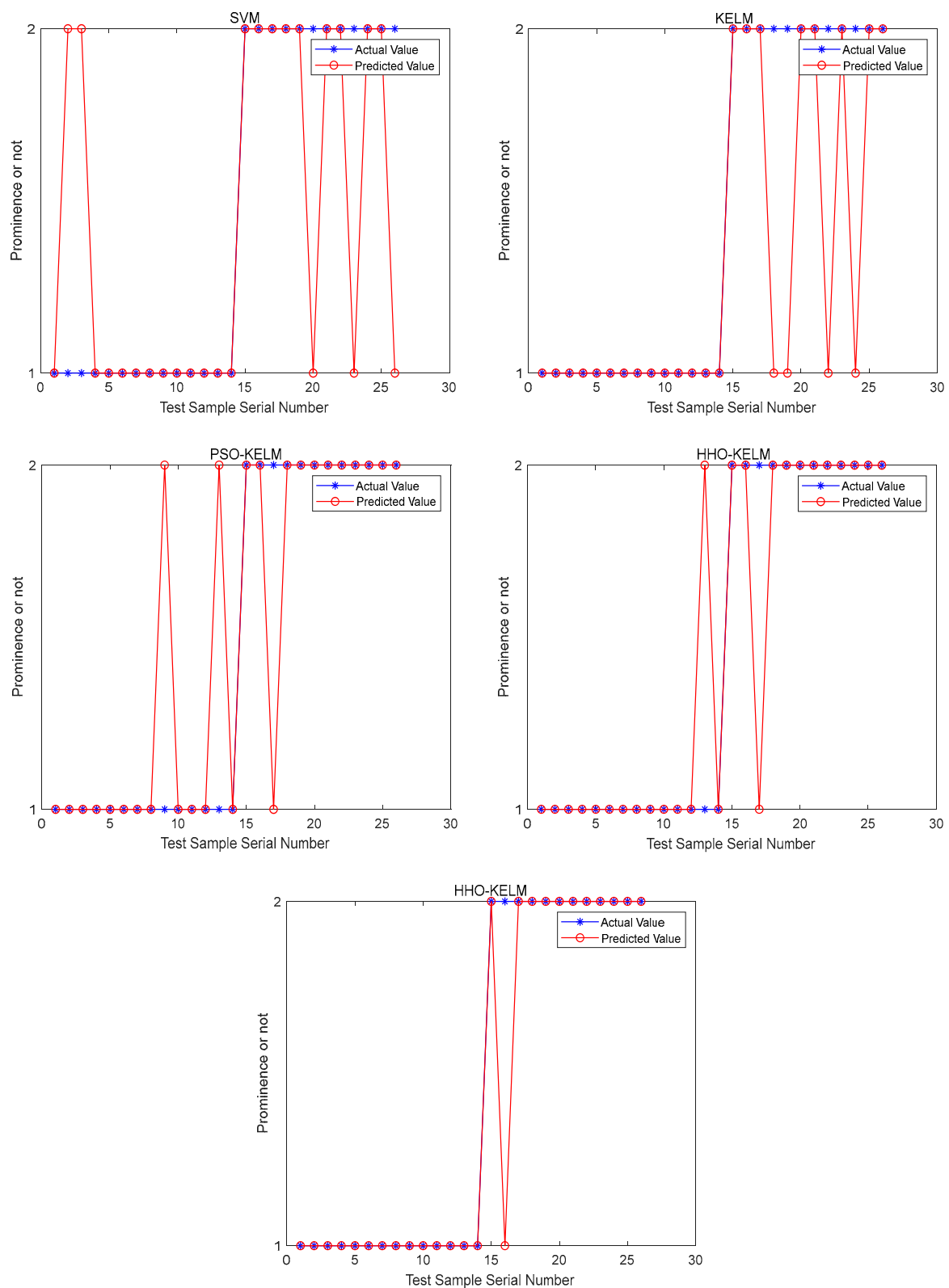
Figure 5. Test set prediction results for different models.

Before filling, there were 35 sets of salient samples and 71 sets of non-salient samples, and the number of non-salient samples was significantly larger than the number of salient samples. As can be seen from Table 4, in the prediction of salient sample accuracy, KELM improved from 65.71% before filling to 85.48%, and PHHO–KELM improved from 88.57%

to 96.77%. Miceforest filling improved the salient accuracy by at least 8.2%. The prediction accuracies for the non-salient samples were all higher, above 90.00%, and the gap in non-salient accuracies between the models changed even less in relative terms, improving by 1.41–2.82%. The prediction results of the pre-filled dataset are clearly inclined to non-salient samples with a large sample size, which made the overall prediction accuracy higher. For overall prediction accuracy, KELM improved from 83.96% to 90.23% and PHHO–KELM improved from 94.34% to 98.50%. Miceforest filling improved the overall accuracy by 3.79–6.64%. The Kappa coefficient of each model after miceforest filling was also higher than that before filling.

The prediction effect of KELM was better than SVM; the salient sample and overall accuracies of KELM after filling were 85.48% and 90.23%, which were higher than those of SVM (83.87% and 88.72%); and the Kappa coefficient of KELM was 0.0305 higher than that of SVM. Compared with other models, the PHHO–KELM model had the best prediction effect. After filling, the salient sample and overall prediction accuracies of the PHHO–KELM model were 96.77% and 98.50%, respectively, with a Kappa coefficient of 0.9698.

It can be seen that the accuracy and Kappa coefficient of each model improved after miceforest filling. Miceforest filling was better, which greatly improved the accuracy of the protruding samples of each model. The prediction effect of KELM was better than that of SVM, and PHHO significantly improved the prediction performance of KELM. The salient sample accuracy of miceforest–PHHO–KELM was 96.77%, the overall accuracy was 98.50% and the Kappa coefficient was 0.9698. The miceforest–PHHO–KELM model gives a better prediction of coal and gas outbursts and has better prediction accuracy and generalization ability for coal and gas outburst prediction.

5. Conclusions

- The miceforest method was proposed to fill in missing data values, and gave optimal results in terms of RMSE and R^2 evaluation compared to KNN, regression and RF. After miceforest filling, the accuracy of non-salient samples was improved to a certain extent, and the accuracy of salient samples and the overall accuracy were significantly higher than for the pre-filling dataset in each model. Miceforest filling improved the salient sample accuracy and overall accuracy by at least 8.2% and 3.79%, so it was an effective algorithm to fill in the missing values of the samples.
- Comparing the prediction effect of KELM and SVM, results showed that the salient sample accuracy, overall accuracy and Kappa coefficient of KELM before and after filling were significantly better than SVM, so the prediction effect of KELM was better than that of SVM.
- After miceforest filling, the optimal coal and gas outburst prediction miceforest–PHHO–KELM model was established by selecting PHHO to optimize the penalty coefficient and kernel function parameters of KELM. Compared with other models, miceforest–PHHO–KELM had higher prediction accuracy and precision, and its salient sample prediction accuracy, overall prediction accuracy and Kappa coefficient were 96.77%, 98.50% and 0.9698, respectively. These results verified that PHHO can effectively improve the prediction performance of KELM, and the miceforest–PHHO–KELM model had better prediction accuracy and recognition rate in the prediction of coal and gas outbursts.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Research data has been presented in the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, C.L.; Wang, E.Y.; Wang, Y.B.; Zhou, X.F. Spatial and temporal distribution of coal and gas protrusion accidents in China in the past 20 years and suggestions for prevention and control. *Coalf. Geol. Explor.* **2021**, *49*, 134–141.
2. Rudakov, D.; Sobolev, V. A mathematical model of gas flow during coal outburst initiation. *Int. J. Min. Sci. Technol.* **2019**, *29*, 791–796. [CrossRef]
3. Black, D.J. Review of coal and gas outburst in Australian underground coal mines. *Int. J. Min. Sci. Technol.* **2019**, *29*, 815–824. [CrossRef]
4. Kursunioglu, N.; Onder, M. Application of structural equation modeling to evaluate coal and gas outbursts. *Tunn. Undergr. Space Technol.* **2019**, *88*, 63–72. [CrossRef]
5. Li, Y.J.; Yang, Y.G. Research on coal and gas prominence prediction based on improved PSO optimized LS-SVM parameters. *Coal Technol.* **2017**, *36*, 129–131. [CrossRef]
6. Yan, Z.; Yao, K.; Yang, Y. A novel adaptive differential evolution SVM model for predicting coal and gas outbursts. *J. Differ. Equ. Appl.* **2016**, *23*, 238–248. [CrossRef]
7. Li, J. IGSA-SVM Prediction Model for Coal and Gas Outburst and Its Application. Ph.D. Thesis, Taiyuan University of Technology, Taiyuan, China, 2016.
8. Wu, Y.; Gao, R.; Yang, J. Prediction of coal and gas outburst: A method based on the BP neural network optimized by GASA. *Process Saf. Environ. Prot.* **2020**, *133*, 64–72. [CrossRef]
9. Xu, Y.S.; Cheng, Y.W. Coal and gas protrusion hazard prediction based on SKPCA and NEAT algorithm. *J. Saf. Environ.* **2021**, *21*, 1427–1433.
10. Fu, H.; Meng, T.R.; Yan, X.; Lu, W.J. Optimization of coal and gas protrusion prediction model for quantum gate line. *Control Eng.* **2021**, *28*, 1731–1737.
11. Wang, Y.; Meng, Y.; Fu, H.; Tu, N.; Xu, Y. Optimized limit learning machine for coal and gas prominence prediction method. *Control Eng.* **2022**, *29*, 2131–2137.
12. Lin, H.F.; Zhou, J.; Jin, H.W.; Yang, Z.Y.; Liu, S.H. Collaborative prediction method of coal and gas prominence hazard level based on feature selection and machine learning. *J. Min. Saf. Eng.* **2023**, *40*, 361–370.
13. Ru, Y.D.; Lyu, X.F.; Guo, J.K.; Hongquan, Z.; Lijuan, C. Real-Time prediction model of coal and gas outburst. *Math. Probl. Eng.* **2020**, *2020*, 2432806.
14. Zheng, X.L. Research on the key technology of coal and gas protrusion prediction based on gas content method. Ph.D. Thesis, Anhui University of Technology, Maanshan, China, 2018.
15. Zheng, X.L.; Lai, W.H.; Xue, S. Application of MI and SVM in coal and gas outburst prediction. *China Saf. Sci. J.* **2021**, *31*, 75–80.
16. Chen, L.-C.; Chen, J.-H. Research on the effect of coal and gas prominence prediction based on data filling-machine learning. *China Sci. Technol. Saf. Prod.* **2022**, *18*, 69–74.
17. Shao, L.S.; Zhan, S.F. Coal and gas protrusion missForest-EGWO-SVM prediction model. *J. Liaoning Univ. Eng. Technol.* **2020**, *39*, 214–218.
18. Wen, T.X.; Su, H.B. WOA-ELM coal and gas protrusion prediction model based on chain multiple interpolation. *China Saf. Prod. Sci. Technol.* **2022**, *18*, 68–74.
19. Shi, H.Y.; Ma, N.J.; Xu, H.T. Exploring the mechanism of coal and gas protrusion based on energy theory. *China Sci. Technol. Saf. Prod.* **2019**, *15*, 88–92.
20. Wang, G.; Wu, M.; Wang, H.; Huang, Q.; Zhong, Y. Sensitivity analysis of influencing factors of coal and gas outburst based on energy balance model. *J. Rock Mech. Eng.* **2015**, *34*, 238–248.
21. Zhang, J.; Wang, X.; Lu, L.; Niu, P. A comparative analysis study of several new intelligent optimization algorithms. *Comput. Sci. Explor.* **2022**, *16*, 88–105.
22. Huang, G.-B.; Zhou, H.; Ding, X.; Zhang, R. Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst. Man Cybern. Part B* **2011**, *42*, 513–529. [CrossRef]
23. Heidari, A.A.; Mirjalili, S.; Faris, H.; Aljarah, I.; Mafarja, M.; Chen, H. Harris hawks optimization: Algorithm and applications. *Future Gener. Comput. Syst.* **2019**, *97*, 849–872. [CrossRef]
24. Tang, A.D.; Han, T.; Xu, D.W. Chaotic elite Harris hawks optimization algorithm. *J. Comput. Appl.* **2021**, *41*, 2265–2272.
25. Song, L.; Wan, J.Z. A comparative study of missing data interpolation methods. *Stat. Decis. Mak.* **2020**, *36*, 10–14.
26. Yi, H.; Wang, Y.; Huang, J.; Song, W.; Li, L. Research on sparrow search algorithm based on hybrid optimization strategy. *Mechatron. Eng. Technol.* **2023**, *52*, 93–97+17.