Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.085749

ARTICLE

An Enhanced Osprey Optimization-Based Interpretable Deep Learning Framework for Predicting Coal Spontaneous Combustion Temperatures

Rui Yan[1], Botao Fan[2], Xueqi Qu[2], Cuihuan Ren[1,*] and Xu Zhou[1,*]

[1]College of Science, North China University of Science and Technology, Tangshan, China
[2]School of Emergency Management and Safety Engineering, North China University of Science and Technology, Tangshan, China
*Corresponding Authors: Cuihuan Ren. Email: rench@ncst.edu.cn; Xu Zhou. Email: zhouxu@ncst.edu.cn

Received: 17 May 2026; Accepted: 24 June 2026; Published: 13 August 2026

ABSTRACT: Accurate prediction of coal spontaneous combustion (CSC) temperatures is crucial for safe coal mine production. To further improve the accuracy of CSC temperature prediction and the interpretability of the model, this study proposes an interpretable Chebyshev chaotic mapping-Lévy flight-Enhanced pinhole imaging inverse learning-Adaptive weighted Osprey Optimization Algorithm optimized Bidirectional Long Short-Term Memory (CLEA-OOA-BiLSTM) framework for predicting CSC temperatures. First, we optimized the Osprey Optimization Algorithm (OOA) by incorporating the Chebyshev chaotic map, Lévy flights, enhanced pinhole imaging backpropagation, and an adaptive weighting strategy, thereby developing the CLEA-OOA algorithm. Through comparative experiments using eight benchmark test functions and four heuristic algorithms, we verified that CLEA-OOA achieves superior convergence accuracy and convergence speed. Subsequently, using data from the Dongtan Coal Mine as the subject of study, we employed CLEA-OOA to perform adaptive optimization of the BiLSTM hyperparameters. Using Spearman's correlation analysis, C_2H_4/C_2H_6, CO, C_2H_4, $CO/\Delta O_2$, and O_2 (%) were identified as key indicators for predicting the CSC temperatures. The results show that the CLEA-OOA-BiLSTM model achieves an R^2 of 0.98, which is higher than that of all comparison models. The model's MSE, RMSE, MAE, and MAPE are 11.26%, 3.36%, 2.73%, and 2.74%, respectively, demonstrating the model's excellent error control capabilities. The results of the global and local SHapley Additive exPlanations (SHAP) interpretability analysis indicate that C_2H_4/C_2H_6 and CO are key contributing factors to the model's decision-making, consistent with the oxidation mechanisms of CSC. The model was validated using coal mine data from multiple locations in Inner Mongolia, Shanxi, and Anhui, with R^2 consistently reaching 0.985, demonstrating strong cross-regional generalization capabilities and practical engineering value. This framework provides a new method for CSC early warning and the intelligent development of mines.

KEYWORDS: Coal spontaneous combustion; coal mine safety; interpretable deep learning; temperature prediction; BiLSTM; improved osprey optimization algorithm; SHAP analysis

1 Introduction

Coal is a primary energy source in China and plays a crucial role in social production. However, during mining operations, coal spontaneous combustion (CSC) may occur, potentially leading to mine fires and severe mining disasters [1]. CSC not only severely hinders safe and sustainable mining operations and result in significant economic losses, but also generate greenhouse gases that threaten environmental sustainability and endanger the lives of workers [2]. Therefore, timely prediction of the degree of CSC is crucial for safe coal mine production.

Currently, the primary methods for predicting CSC include infrasound wave monitoring and warning, indicator gas analysis [3], temperature sensing, tracer gas detection, fuzzy cluster analysis, and CSC tendency assessment. Many scholars have used various experimental studies and statistical methods to analyze how indicator gases change with coal temperature during the spontaneous combustion process, thereby enabling early warning for CSC. Both Yan et al. and Guo et al. [4,5] measured the parameter data for indicator gases, and the latter constructed a logistic function to fit the data curves and classified the spontaneous coal combustion warning into seven stages. Zhang et al. [6] combined practical experience with SHAP (Shapley Additive Explanations) theory to select the optimal indicator gas, and experimental results confirmed that the Random Forest (RF) model performed best after the indicator gas was selected. Muduli et al. [7] used machine learning and data mining methods based on a collected dataset to predict CSC. Kamran and Shahani [8] studied fire indicators at the Adularya coal mine in Turkey. They constructed a database using the t-SNE dimension reduction method combined with the k-means clustering algorithm, and developed a support vector classification (SVC) model to predict fire severity levels. Thuppari et al. [9] utilized IOT-based detection technology and employed active learning (AL) and semi-supervised learning (SSL) to improve logistic regression models, thereby enabling the prediction of coal mine fires. In addition, Kumar et al. [10] emphasized the importance of using sensor technology for monitoring CSC and described the various stages of this process. Yetkin et al. [11] conducted a systematic study of the causes of CSC using the Fault Tree Analysis (FTA) method, established a framework for predicting the risk of CSC, and assessed the probability of such risks under various operating conditions. Liu et al. [12] and Lei et al. [13] used gases such as CO, O_2, CO_2, and C_2H_4 to develop a model for the CSC temperatures, thereby improving the reliability of early warnings.

The above research represents traditional mathematical and statistical methods for predicting CSC, which are widely used in this field. The traditional monitoring, statistical analysis, and risk assessment methods described above provide an important theoretical basis for early warning of CSC; however, most of them struggle to fully capture the complex nonlinear relationships among multiple factors during the spontaneous combustion process, and thus still have certain limitations in terms of predictive accuracy and adaptability. With advances in computer science and technology, numerous scholars have applied machine learning algorithms to early warning of CSC, continuously enhancing the accuracy of prediction results.

In recent years, Dogra et al. [14] has proposed applying machine learning techniques to the field of CSC. Zhang et al. [15] categorized the indicators into five groups by analyzing CH_4, C_2H_2, CO, and Graham coefficients, and compared RNN, GRU, and LSTM prediction models. The results showed that under data imbalance conditions, the GRU model performed best. Pan et al. [16] improved prediction performance by using a multi-layer GCN to capture gas concentrations and their interrelationships. Shukla et al. [17] used five machine learning models, including Support Vector Regression (SVR), RF, and XGBoost, to predict indicators of CSC susceptibility, and found that ensemble tree models such as RF and XGBoost performed best. Said et al. [18,19] employed artificial neural networks (ANNs) to predict the liability for CSC based on data from the Witbank coalfield in South Africa. Dogra et al. [20] proposed an AdaBoost-based method for predicting CSC, applying machine learning algorithms to this field and comparing other machine learning models. Vo Thanh et al. [21] compared four machine learning models—ET, XGBoost, AdaBoost, and RF—based on five indicator gases. Among these, XGBoost performed exceptionally well on both the test and training datasets. To overcome the limitations of single-model prediction accuracy, Zhao et al. [22] developed a new framework for predicting CSC based on rough sets (RS)-Stacking-SHAP.

The forecasting accuracy of the above machine learning models for CSC temperatures is influenced by model hyperparameters. Traditional manual hyperparameter tuning struggles to find the globally optimal combination of hyperparameters. Consequently, many scholars have applied intelligent optimization algorithms to the task of hyperparameter optimization for machine learning models, offering new approaches to solving hyperparameter optimization problems.

Wu et al. [23] used the SSA to optimize a RF model and compared it with five other models, including PSO-RF and SSA-SVR, achieving high predictive accuracy. Lawal et al. [24] applied the Spotted Hyena optimization algorithm to optimize an ANN network, constructing a SHO-ANN model to predict the risk of CSC. Ding et al. [25] collected coal sample data, conducted a study of the temperature-dependent behavior of CSC indicator gases, and compared multiple models, including BO-LSTM, LSTM-GRU, and BO-XGBoost. From the results, BO-LSTM demonstrates excellent performance and application potential. Kong et al. [26] and Bai et al. [27] proposed the Modified Whale Optimization Algorithm (MWOA) and the Improved Particle Swarm Optimization (IPSO), respectively, to optimize backpropagation networks (BP), demonstrating outstanding performance in CSC early warning. Shao et al. [28,29] proposed two variants of the Improved Tornado Optimization with Coriolis force (ITOC) method to optimize the CNN-BiGRU-CBAM and Kernel Extreme Learning Machine (KELM) models, respectively, for accurately predicting the CSC temperature. Most swarm intelligence algorithms struggle to balance convergence efficiency and accuracy during iterations, and improvements are needed to enhance optimization performance.

Overall, significant progress has been made in existing research on CSC prediction. However, traditional methods have limitations in identifying complex nonlinear characteristics, the predictive performance of machine learning models is highly influenced by hyperparameter settings, and most swarm intelligence optimization algorithms still suffer from issues such as getting stuck in local optima and insufficient convergence accuracy. Therefore, it is necessary to design more efficient optimization strategies to further improve the predictive performance and generalization ability of deep learning models in the field of CSC temperature prediction.

With the advancement of swarm intelligence optimization algorithms, an increasing number of novel optimization methods are being applied to the field of hyperparameter tuning for machine learning models. Among these, the Osprey Optimization Algorithm (OOA) has attracted the attention of researchers due to its simple structure and few parameters. Based on the hunting behavior of ospreys, Dehghani and Trojovský [30] OOA in 2023; however, this algorithm also suffers from poor optimization performance. Therefore, this paper proposes the CLEA-OOA algorithm, which integrates that integrates Chebyshev chaotic mapping, Lévy flight, enhanced pinhole imaging inverse learning, and adaptive weighting strategies to optimize the BiLSTM. We then construct the CLEA-OOA-BiLSTM CSC temperature prediction model to improve the model's predictive accuracy. In addition, this study conducts a SHAP explainability analysis on the CLEA-OOA-BiLSTM model to validate the rationality of its decision-making. We apply CLEA-OOA-BiLSTM to coal seam data from other regions, such as Inner Mongolia and Shanxi, to validate the model's generalizability and engineering value. The following are the contributions and innovations of this research:

(1) Using Spearman's correlation coefficient to analyze the complex coupling relationships between CSC temperatures and indicator gases, and to screening key features to eliminate redundancy.
(2) This paper combines the Chebyshev chaotic map, Lévy flights, an enhanced pinhole imaging back-propagation learning method, and an adaptive weighting strategy to enhance OOA, thereby developing a novel optimization algorithm called CLEA-OOA. Through comparative experiments using benchmark functions, we have verified that CLEA-OOA achieves superior convergence accuracy and convergence rates.
(3) We propose a CLEA-OOA-BiLSTM framework for predicting CSC temperatures, which automates the optimization of BiLSTM hyperparameters and effectively improves the accuracy of these predictions. Conduct global and local interpretability analyses of the model using SHAP theory to identify the

key factors influencing the model's decisions. In addition, applying the CLEA-OOA-BiLSTM-SHAP framework to coal mine data from other regions, such as Inner Mongolia and Shanxi, demonstrates the framework's applicability and engineering value, thereby contributing to safe production and intelligent development in the coal mining industry.

2 Algorithm Basics

2.1 Optimization Algorithm

The OOA is derived from the osprey's hunting behavior on the water. The algorithm generates a suitable optimization solution by having individual ospreys search the solution space. The position of each osprey in the OOA population in the exploration space directly maps to specific variable values, thus representing a possible solution to the problem. Initially, randomly initialize the osprey individuals, as expressed by the following formula:

$$x_{i,j} = b_{l,j} + rand \cdot \left(b_{u,j} - b_{l,j}\right) \tag{1}$$

In the formula, $x_{i,j}$ denotes the jth problem variable of the ith osprey, N represents the number of ospreys, $b_{u,j}$ and $b_{l,j}$ are the boundaries of the search space.

This paper uses a population matrix to represent the population members in the OOA algorithm, where a row in the matrix represents a candidate solution. Using matrix modeling, we obtain:

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} = \begin{bmatrix} x_{1,1}, x_{1,2} \cdots x_{1,m} \\ \vdots \qquad\qquad \vdots \\ x_{N,1}, x_{N,2} \cdots x_{N,m} \end{bmatrix}_{N\times m} \tag{2}$$

where, X represents the position matrix, X_i denotes the candidate solution.

The objective function for this problem can be expressed as a vector:

$$F = \begin{bmatrix} f_1 \\ \vdots \\ f_N \end{bmatrix} = \begin{bmatrix} f\left(X_1\right) \\ \vdots \\ f\left(X_N\right) \end{bmatrix} \tag{3}$$

where, f_i or $f\left(X_i\right)$ denotes the objective function value of the ith candidate solution.

OOA simulates the osprey's natural hunting behavior. This hunting process consists of two stages: ① the osprey detects the fish's location and hunts it (the exploration stage); ② it carries the fish to a favorable spot to consume it (the exploitation stage).

2.1.1 Positioning and Fishing (Exploration Phase)

For each osprey, the locations within the exploration space where the objective function takes on a better value are considered a "School of Fish underwater", as expressed below:

$$FP_i = \left\{X_k \,|k \in \{1, 2, \ldots, n\} \wedge f_K < f_i\right\} \cup \left\{X_{best}\right\} \tag{4}$$

where X_{best} denotes the optimal position and FP_i represents the set of positions. The osprey randomly catches a fish and continuously updates its position during the hunt based on the formulas below:

$$x_{i,j}^{P1} = x_{i,j} + rand \cdot \left(SF_{i,j} - I_{i,j} \cdot x_{i,j}\right) \tag{5}$$

$$x_{i,j}^{P1} = \begin{cases} x_{i,j}^{P1}, b_{l,j} \le x_{i,j}^{P1} \le b_{u,j} \\ b_{l,j}, x_{i,j}^{P1} < b_{l,j} \\ b_{u,j}, x_{i,j}^{P1} > b_{u,j} \end{cases} \tag{6}$$

$$X_i = \begin{cases} X_i^{P1}, f_i^{p1} < f_i \\ X_i, else \end{cases} \tag{7}$$

where, $SF_{i,j}$ is a fish selected randomly from its school, $I_{i,j}$ donates a number in the set $\{1, 2\}$, $rand$ is a random number within the range $[0, 1]$, X_i^{P1} is the new position in the first phase, $x_{i,j}^{P1}$ is the value of the jth dimension, and f_i^{p1} is the objective function value. If the new candidate position has a better fitness value, the position is updated according to Eq. (7).

2.1.2 Positioning Fish in Appropriate Locations (Exploitation Stage)

Ospreys carry the fish they have caught to a suitable spot to eat. This natural behavior is translated into the OOA algorithm, causing the osprey's position to undergo slight changes. In the algorithm, use Eqs. (8) and (9) to calculate a new location, and the position is updated using Eq. (10).

$$x_{i,j}^{P2} = x_{i,j} + \frac{b_{l,j} + rand \cdot \left(b_{u,j} - b_{l,j}\right)}{t} \tag{8}$$

$$x_{i,j}^{p2} = \begin{cases} x_{i,j}^{P2}, b_{l,j} \le x_{i,j}^{P2} \le b_{u,j} \\ b_{l,j}, x_{i,j}^{P2} < b_{l,j} \\ b_{u,j}, x_{i,j}^{P2} > b_{u,j} \end{cases} \tag{9}$$

$$X_i = \begin{cases} X_i^{P2}, f_i^{p2} < f_i \\ X_i, else \end{cases} \tag{10}$$

In the formula, T denotes the maximum iteration count, X_i^{P2} is the ith osprey's new position in the development phase, f_i^{p2} is its objective function value.

2.2 BiLSTM Model

BiLSTM is a network based on the ideas proposed by Schuster and Paliwal [31] that incorporates bidirectional LSTM units. LSTM is an improvement over RNN. The enhancement is achieved through a three-stage gating mechanism. This enables effective information retention over extended periods, endowing the network with memory and the ability to store historical information. Consequently, it resolves the gradient vanishing and gradient explosion issues inherent in RNNs. Fig. 1 illustrates the three-gate structure of an LSTM. The following are the formulas and calculation process for LSTM.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{11}$$

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{12}$$

In the equation, f_t and i_t represent the forget gate and input gate, respectively; $\tilde{C}_t$ is the state of the new candidate cell; σ is the sigmoid activation function; W_f, W_i, and W_c are the weights of the gate and candidate cells, respectively, and b_f, b_i, and b_c are the corresponding bias terms; x_t is the input value; and h_{t-1} is the hidden state at the previous time step.

$$\tilde{C}_t = \tanh\left(W_c \cdot [h_{t-1}, x_t] + b_c\right) \tag{13}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{14}$$

$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{15}$$

$$h_t = o_t \cdot \tanh\left(C_t\right) \tag{16}$$

where, f_t controls the discarding of memory units from the last time step, i_t acts on candidate units, represents the state of the memory unit at the current time step, o_t is the output gate, W_o and b_o are the corresponding weights and bias terms, and h_t represents the output.

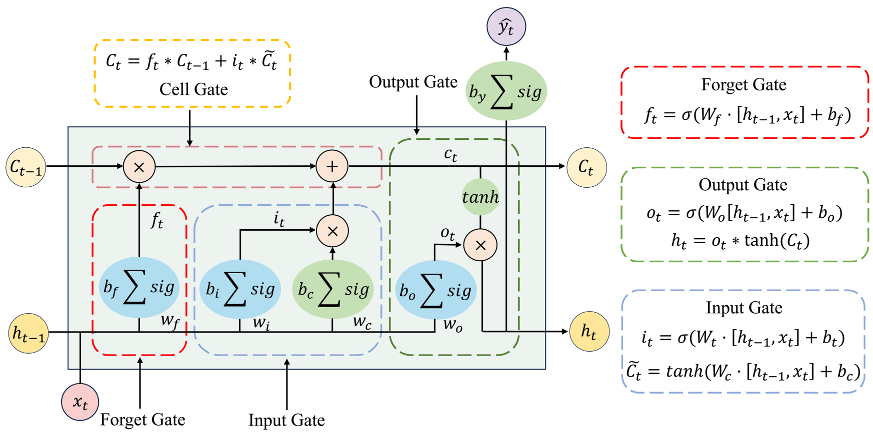

Figure 1: Structure of LSTM neural network.

BiLSTM is a modification of the standard LSTM architecture, combining forward and backward LSTMs into a single model. Within the BiLSTM structure, bidirectional LSTM units in each layer operate independently in a loop, then merge their output hidden states before passing them to the next layer. The BiLSTM architecture is as follows (Fig. 2):

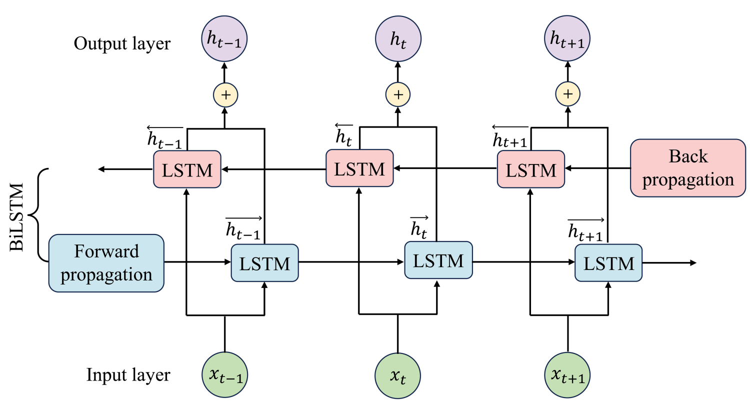

Figure 2: Structure of BiLSTM neural network.

This structure integrates past and future states through the operation of a bidirectional LSTM, providing more comprehensive information to the output layer. Therefore, this paper employs a BiLSTM to capture contextual dependencies and enhance the accuracy of predicting CSC temperatures.

3 Multi-Strategy Enhanced Osprey Optimization Algorithm (CLEA-OOA)

The standard OOA optimization model is relatively simple, and when dealing with complex problems, OOA suffers from poor search accuracy and slow convergence. To improve optimization performance, we incorporate Chebyshev chaotic maps, Lévy flights, an enhanced pinhole imaging backpropagation learning method, and an adaptive weighting strategy into the OOA to enhance the algorithm, thereby proposing an improved CLEA-OOA algorithm.

3.1 Improvement Strategies and the Stages of Improvement

3.1.1 Chebyshev Chaotic Mapping Strategy

In OOA, the initial positions of the population are assigned randomly, which limits the diversity of the population and affects search efficiency. Therefore, this paper proposes a Chebyshev chaotic mapping strategy to initialize the osprey's position and normalize the chaotic sequence, thereby enabling a more comprehensive exploration of the candidate solution space. The formula is as follows:

$$z_{i,j+1} = \cos\left(a \cdot \arccos\left(z_{i,j}\right)\right), z_{i,j}^{`} = \frac{z_{i,j} - \min Z_i}{\max Z_i - \min Z_i} \tag{17}$$

where, a takes on a value of 4, $Z_i = \left(z_{i,1}, z_{i,2}, \ldots z_{i,j}, \ldots, z_{i,m}\right)$ is the chaotic sequence of the i-th osprey generated by a Chebyshev mapping, $z_{i,j}^{`}$ is the normalization formula.

During population initialization, min-max normalized chaotic sequences are employed in place of random functions. This approach distributes individuals more evenly across the exploration space, thereby expanding the exploration space for finding optimal solutions. The formula is as follows:

$$x_i = b_{l,j} + z_{i,j}^{'} \cdot \left(b_{u,j} - b_{l,j}\right), i = 1, 2, 3, \cdots, N \tag{18}$$

3.1.2 Lévy Flight Strategy

The Lévy flight strategy is based on the Lévy distribution, which has a probability density that is discontinuous at the mean and exhibits long-tail properties. The expression is as follows:

$$Levy = 0.01 \cdot \left(\mu / |v|^{1/\beta}\right) \tag{19}$$

$$\sigma = \left(\frac{\Gamma\left(1+\beta\right)\sin\left(\pi\beta/2\right)}{\Gamma\left(\left(1+\beta\right)/2\right)\beta 2^{(\beta-1)/2}}\right)^{1/\beta} \tag{20}$$

where, Γ denotes the Gamma function, $v \sim N\left(0, 1^2\right)$, $u \sim N\left(0, \sigma^2\right)$, and here is set to $\beta = 1.5$. $Levy$ represents the Lévy flight function.

This paper employs this strategy to optimize the exploration phase. Modified formulas are as follows:

$$x_{i,j}^{P1} = x_{i,j} + Levy \cdot \left(SF_{i,j} - I_{i,j} \cdot x_{i,j}\right) \tag{21}$$

3.1.3 Enhanced Pinhole Imaging Backward Learning Strategy

Applying the pinhole imaging principle to OOA, we generate counterpart individuals for each population member through reverse learning. The fundamental definition for constructing reverse solutions is:

$$x'_{i,j} = \frac{b_{u,j} + b_{l,j}}{2k} - \frac{x_{i,j}}{k} + \frac{b_{u,j} + b_{l,j}}{2} \tag{22}$$

where, k denotes the ratio of the original image to the actual image. $x'_{i,j}$ is the corresponding inverse solution. To address the issue of blind search in traditional strategies, this study proposes incorporating the global optimal solution into the calculation of the backward solution. Apply the enhanced pinhole imaging reverse learning strategy to the exploration phase to increase population diversity and convergence speed. The modified formula is as follows:

$$X_i^{P1} = b_l + b_u - X_i + rand \cdot (X_{best} - X_i) \tag{23}$$

where X_{best} represents the global optimal solution. X_i^{P1} denotes the inverse solution for the i-th osprey individual.

3.1.4 Adaptive Weighting Strategy

This study proposes an adaptive weighting strategy to dynamically adjust the stride length of individual ospreys, thereby balancing the exploration and exploitation phases. The formula is as follows:

$$w(t) = w_{start} - \frac{t \cdot (w_{start} - w_{end})}{T} \tag{24}$$

In the formula, the initial weight is set to $w_{start} = 0.8$, the final weight to $w_{end} = 0.3$.

During the exploitation phase, a dynamic adaptive weighting strategy was incorporated, whereby the weights gradually decrease as the number of iterations increases, thereby enhancing its applicability. The formula is:

$$x_{i,j}^{P2} = x_{i,j} + w(t) \cdot \left[\frac{b_{l,j} + rand \cdot (b_{u,j} - b_{l,j})}{t} \right] \tag{25}$$

3.2 Implementation Steps of the CLEA-OOA Algorithm

Based on Sections 3.1.1–3.1.4, we propose a multi-strategy improved CLEA-OOA algorithm. Fig. 3 illustrates the principle and implementation process of the CLEA-OOA algorithm.

Step 1: Initialize the CLEA-OOA. Set the maximum iterations $T_{\max}$, population size, hyperparameter dimensions. The population positions are initialized using the Chebyshev chaotic map, and the objective function values for each osprey are calculated to identify the globally optimal individuals, whose positions and fitness values are then recorded.

Step 2: In each iteration, traverse all individuals within the population. For each individual, screen its "submerged fish school." With a 50% probability, randomly choose either the exploration phase ($p < 0.5$) or the exploitation phase ($p \geq 0.5$) for position update.

- When $p < 0.5$ and $r < 0.2$, employ a Lévy flight strategy, the population search range is expanded, enhancing global search capabilities.

- When $p < 0.5$ and $0.2 \leq r < 0.5$, perform the pinhole imaging transformation mechanism, they incorporate positional information from the global optimum solution to generate inverse solutions, thereby enhancing positional diversity.
- When $p < 0.5$ and $r \geq 0.5$, directly use the original exploration position for position updates.
- When $p \geq 0.5$, applies an adaptive weighting strategy, it dynamically adjusts the movement step size of population individuals to enable fine-grained search during the later stages of the algorithm, thereby enhancing solution accuracy.

Step 3: Perform boundary checks on each individual's new position to ensure all updated positions remain within the exploration space. If the calculated new fitness is better, update the current individual's position.

Step 4: After completing the position update for an individual, if the fitness of the current individual is superior to the global optimum, then both the global optimum position and fitness value are updated. Stop iterating when $T_{\max}$ is reached and output the global optimal solution.

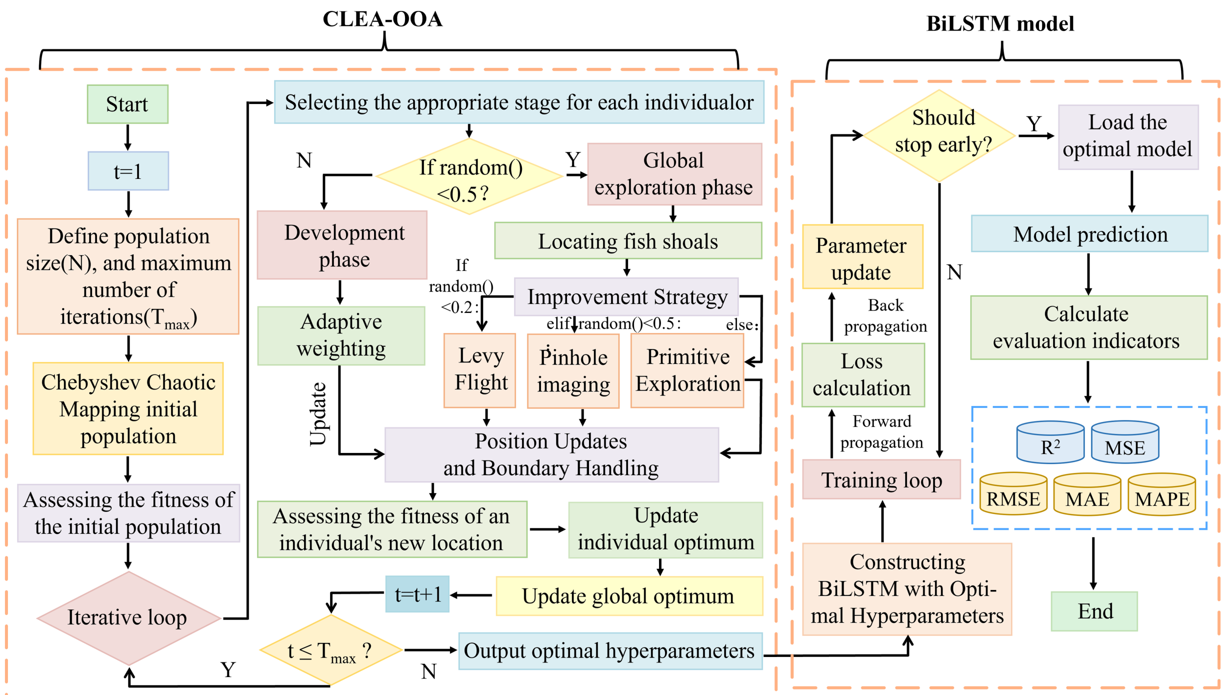

Figure 3: CLEA-OOA implementation flowchart.

Algorithm 1 is the pseudocode for implementing the CLEA-OOA algorithm.

Algorithm 1: CLEA-OOA algorithm

Input: hyperparameters: obj_func, dim, pop_size, max_iter, lb, ub, n_features
Output: GbestPosition, GbestScore
1: Initialize Pop via **Chebyshev chaotic mapping**
2: Fitness = [obj_func(ind, n_features) for ind in Pop]
3: GbestScore = min(Fitness); GbestPosition = Pop[argmin(Fitness)]
4: **for** t = 0 to maxiter − 1 **do**

(Continued)

Algorithm 1 (continued)

```
    current iter = t + 1
    for i = 0 to popsize − 1 do
        if p < 0.5 then
            FishSwarm = get_fish_swarm(i, pop_size, Fitness)
            FishPos = FishSwarm[randint(0, len(FishSwarm) − 1)]
            r = 𝒰(0, 1), s ∈ {1, 2}
            OriginalExplore = Pop[i] + r × (FishPos − s × Pop[i])
            if r < 0.2 then
                Apply Lévy flight strategy: LevyStep = levy_flight(dim)
                Update X_i: ExplorePos = Pop[i] + LevyStep × (FishPos − s × Pop[i])
            else if r < 0.5 then
                Apply small hole imaging strategy: ExplorePos =
                small_hole_imaging(OriginalExplore)
            else
                ExplorePos = OriginalExplore
            end if
            Apply boundary constraint: ExplorePos = clip(ExplorePos, lb, ub)
            ExploreFitness = obj_func(ExplorePos, n_features)
            if ExploreFitness < Fitness[i] then
                Update Pop[i] and Fitness[i]: Pop[i] = ExplorePos; Fitness[i] = ExploreFitness
            end if
        else
            Apply adaptive weight strategy: Weight = adaptive_weight(t)
            r_dev = 𝒰(0, 1)
            Update X_i: ExploitPos = Pop[i] + Weight × (lb + r_dev × (ub − lb))/current_iter
            Apply boundary constraint: ExploitPos = clip(ExploitPos, lb, ub)
            ExploitFitness = obj_func(ExploitPos, n_features)
            if ExploitFitness < Fitness[i] then
                Update Pop[i] and Fitness[i]: Pop[i] = ExploitPos; Fitness[i] = ExploitFitness
            end
        end
        if Fitness[i] < GbestScore then
            Update global best: GbestScore = Fitness[i]; GbestPosition = Pop[i]. copy()
        end
    end
end
Return GbestPosition, GbestScore
```

3.3 Performance Testing of CLEA-OOA

To validate the high performance of the proposed CLEA-OOA, we selected the original OOA, Pelican Optimization Algorithm (POA), Whale Optimization Algorithm (WOA), and Grey Wolf Optimizer (GWO) as comparison algorithms, and conducted comparative tests using four unimodal (F1–F4) and four multimodal (F5–F8) benchmark functions, respectively. During testing, the parameters of each optimization algorithm remained consistent, with 50 individuals and 500 iterations for each algorithm. Fig. 4 shows the test results.

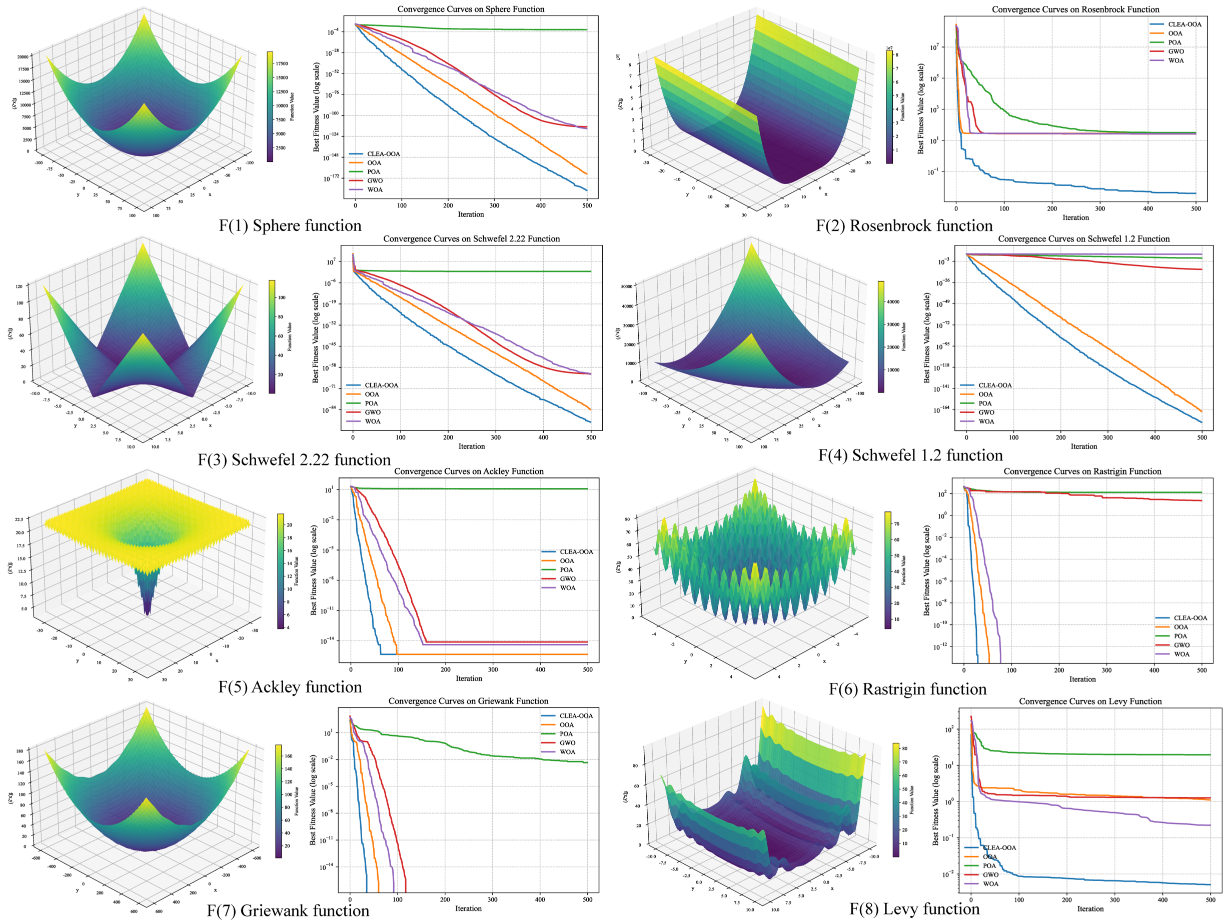

Figure 4: Benchmark function test results.

Fig. 4 shows that the iteration convergence curves of the CLEA-OOA are steeper for all functions, requiring fewer iterations to converge, and that the optimal fitness values at convergence are consistently lower than those of other optimization algorithms, indicating that this algorithm can reduce optimization time and improve accuracy. These results further confirm that the proposed strategies have indeed improved the OOA algorithm. The proposed CLEA-OOA algorithm demonstrates significant advantages across all key performance metrics in function testing.

4 The CLEA-OOA-BiLSTM Framework Based on Interpretability

4.1 BiLSTM Hyperparameter Optimization

To identify the optimal hyperparameters, this paper employs the proposed CLEA-OOA to optimize the BiLSTM. The BiLSTM input layer was set to 5 dimensions, and the output layer to 1 dimension. Four hyperparameters were selected for optimization, and their respective value ranges were defined. The ranges for these four hyperparameters are shown in Table 1.

Table 1: Hyperparameter range.

Hyperparameters	Range
Hidden layer neurons	[50–300]
Learning rate	[0.001–0.01]
Time step	[5, 20]
Epochs	[100, 350]

After determining a reasonable range for hyperparameter optimization, this paper uses the L_{MSE} function to quantify the model's bias, thereby evaluating the hyperparameter optimization process to identify the optimal combination of hyperparameters. The formula can be expressed as:

$$L_{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{26}$$

To verify the effectiveness of the proposed CLEA-OOA algorithm in optimizing BiLSTM hyperparameters, this paper compares CLEA-OOA with other optimization algorithms such as WOA and POA. Each optimization algorithm is combined with BiLSTM to predict coal self-ignition temperatures, using MSE as the objective function, and the performance of each algorithm is evaluated based on convergence curves. To ensure a fair comparison of the different optimization algorithms, all algorithms were run with the same population size of five individuals and 20 iterations. Additionally, each individual underwent a fitness evaluation only once per iteration, ensuring that the number of objective function evaluations remained consistent across all algorithms. The convergence curves of the model are shown in Fig. 5.

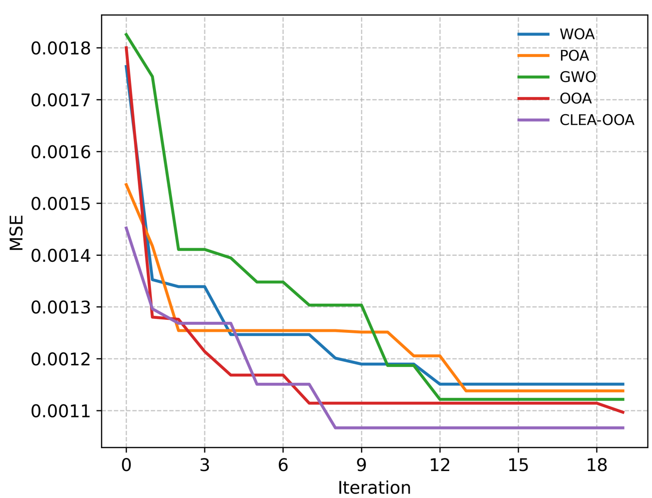

Figure 5: Training curves for different optimization algorithms.

Compared with the WOA, POA, GWO, and original OOA algorithms, CLEA-OOA exhibits a steeper descent rate in the early stages of the iteration and stabilizes after the 8th iteration, indicating that the algorithm is capable of quickly locating the region of high-quality solutions. At the same time, CLEA-OOA was the first to reach a converged state and achieved the lowest MSE value, indicating that it possesses stronger global search capabilities and optimization accuracy during the BiLSTM hyperparameter optimization process. These results demonstrate that the optimization accuracy of CLEA-OOA has been significantly improved, and it has exhibited superior computational capabilities and optimization performance in the task of predicting CSC temperatures.

4.2 *CLEA-OOA-BiLSTM-SHAP Coal Temperature Prediction Framework*

By optimizing the hyperparameters of the BiLSTM using the CLEA-OOA algorithm, we propose an interpretable CLEA-OOA-BiLSTM method for predicting CSC. This method consists of three main parts: the proposed CLEA-OOA algorithm, model building, and SHAP analysis. Fig. 6 illustrates the implementation process of this method:

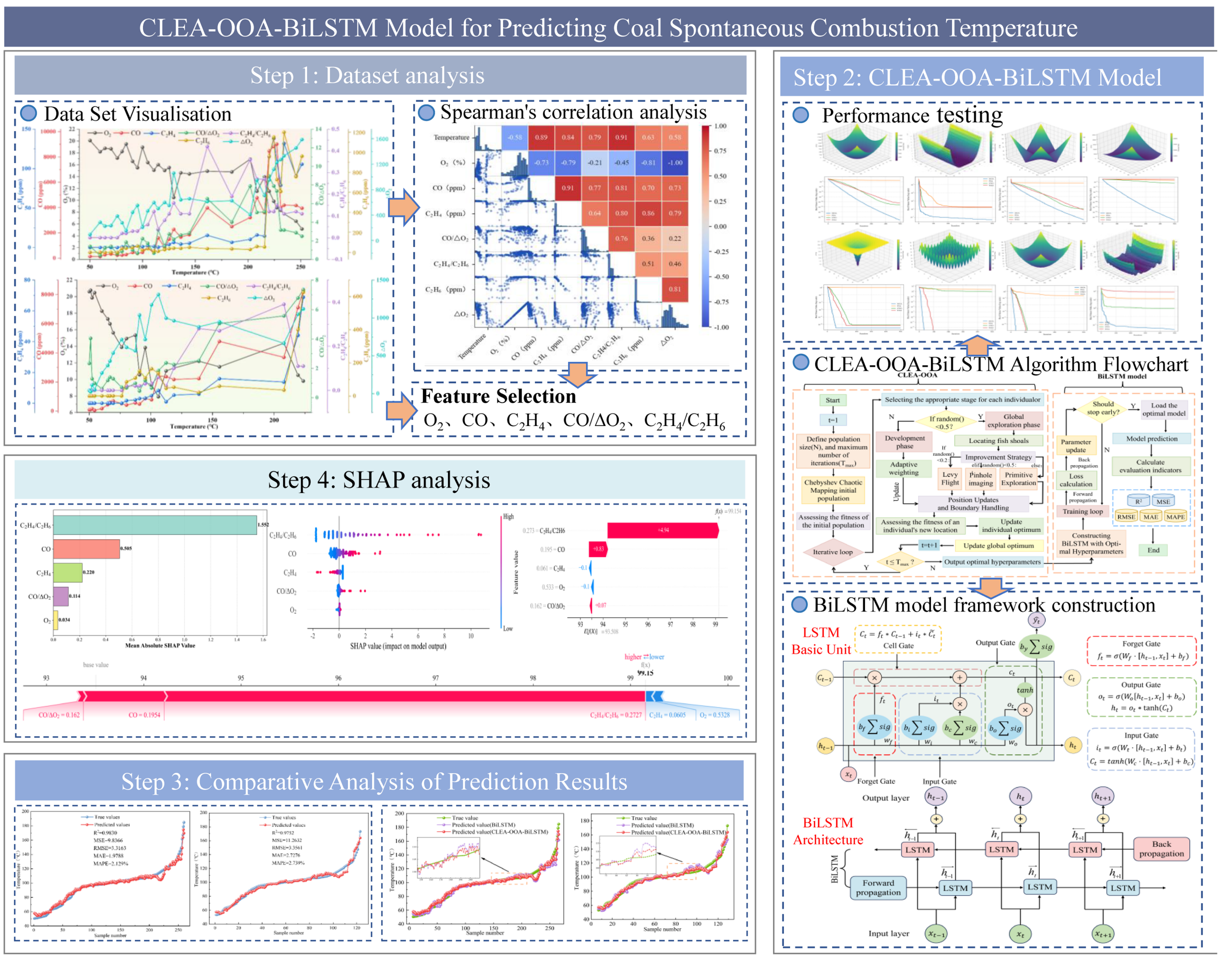

Figure 6: Method schematic diagram.

Step 1: Collect data and perform data analysis. Perform data preprocessing (including imputing missing values and data normalization), visualize the processed data, and then use Spearman's correlation analysis to identify key indicators for the coal temperature prediction model.

Steps 2 and 3: We enhanced the OOA using strategies such as Chebyshev chaotic maps, Lévy flights, and adaptive weights, and validated the performance of the CLEA-OOA through performance tests of basic functions. Apply the CLEA-OOA to optimize the BiLSTM, obtain the optimal hyperparameters, and build a predictive model. Apply the trained optimal BiLSTM model to predict the temperature of CSC, visualize the prediction results, and calculate metrics (R^2, RMSE, MAE, MAPE) to assess the model's performance.

Step 4: Global and local SHAP analyses explain the global influence mechanisms of input features and the prediction process for individual samples, respectively.

5 Experimental Design

5.1 Data Processing

Data Sources and Preprocessing

The data used in the paper is sourced from the publicly available dataset in Jiang's [32] literature, which was collected from Dongtan Coal Mine in China. During data collection, staff crushed a coal sample from a certain mine, screened out 200 g of coal particles of different sizes, and mixed these samples to form a composite sample. The average particle size of the coal sample was 4.18 mm. They then placed 1000 g of the composite sample into a temperature-programmed device. The composite sample was heated under conditions of a constant heating rate of 0.3°C/min, a bed voidage of 0.48% for the experimental apparatus, and an air flow rate of 120 mL/min, with gas products measured throughout the process. Heating was stopped when the temperature reached the predetermined value. The data includes temperature and gas composition data across multiple oxidation stages, encompassing single gases such as CO, CO_2, C_2H_4, C_2H_6, and composite gas indicators like C_2H_4/C_2H_6 and $CO/\Delta O_2$. The experiments yielded a total of 507 data samples, of which 337 were the training set and 170 were the test set. This experiment utilized mixed heterogeneous coal samples, collected characteristic data encompassing various stages of spontaneous combustion and oxidation, and verified the model's applicability to heterogeneous coal samples.

The data contains missing values and outliers. This study uses linear interpolation to impute a small number of missing values in order to ensure data completeness and reflect the true trend. The formula for linear interpolation is as follows:

$$y = y_1 + \frac{(y_2 - y_1)}{(x_2 - x_1)} \times (x - x_1) \tag{27}$$

In the formula, (x_2, y_2) and (x_1, y_1) are two known points, used to estimate the value of the y at point x between them.

The different units and value ranges of the features in the data can cause an imbalance in the model's weights, thereby reducing prediction accuracy. This study uses min-max normalization to process the data:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{28}$$

5.2 Feature Analysis

5.2.1 Mechanisms and Data Analysis of CSC

This paper analyzes the oxidation mechanisms underlying CSC and visualizes selected datasets to provide an in-depth explanation of the patterns of change in key indicator gases and coal temperature. The curves are shown in Fig. 7.

As shown in Fig. 7a,b, the O_2 concentration continuously decreases with increasing temperature, with the rate of decrease accelerating significantly after 200°C. This indicates that CSC is fundamentally an oxygen-consuming process, and the rate of oxygen consumption is continuously increasing. ΔO_2 represents the absolute change in oxygen consumption. As temperature increases, the ΔO_2 value continues to increase.

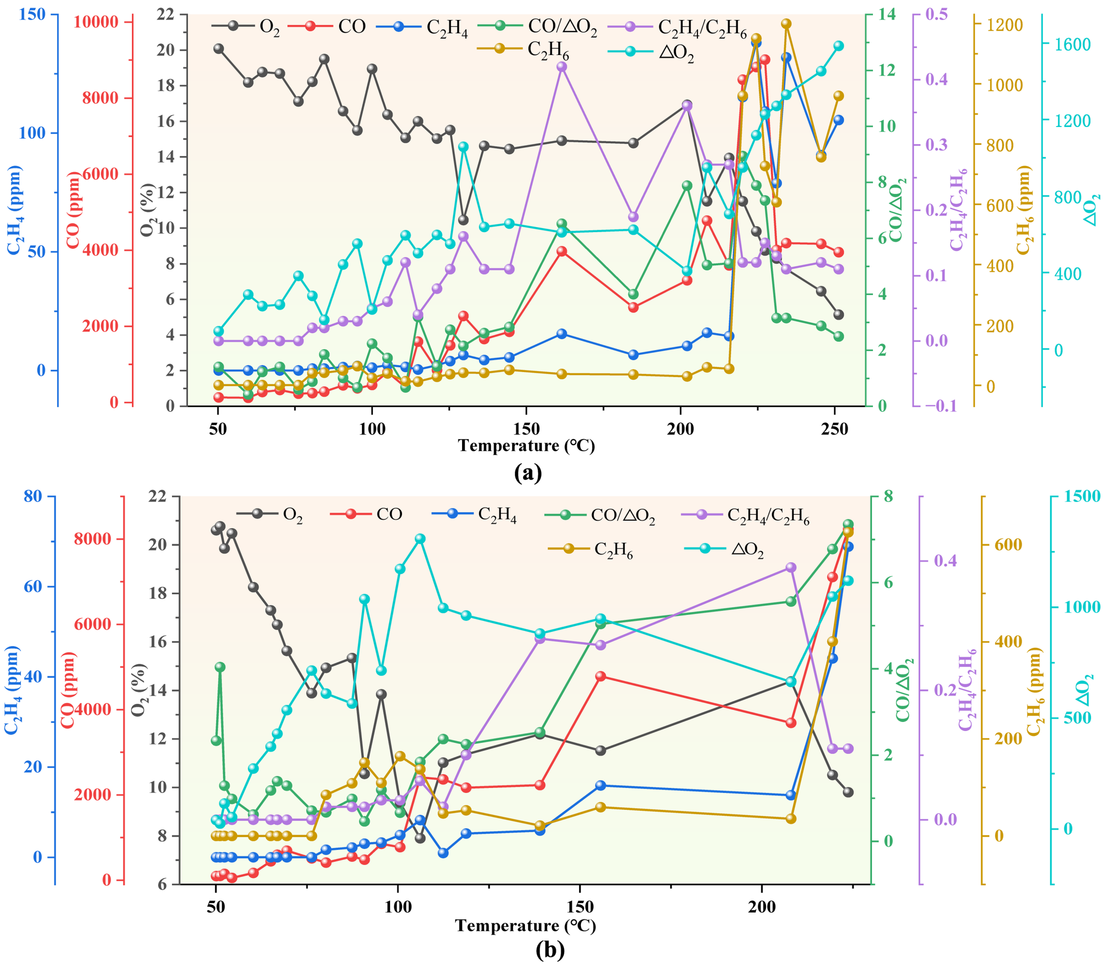

Figure 7: Raw data distribution chart. (**a**) Visualization of training set data; (**b**) visualization of test set data.

The CO rises slowly at the beginning, but increases sharply during the subsequent high-temperature rapid oxidation phase, with the rate of increase accelerating significantly after 100°C. Changes in CO concentration indicate that this indicator is highly sensitive to temperature variations in the high-temperature zone, serving as a crucial criterion for delineating CSC oxidation stages.

C_2H_4 is not present in coal seams. Significant C_2H_4 generation begins around 100°C, and the concentrations of C_2H_4 and C_2H_6 rise sharply after 200°C, exhibiting exponential growth. C_2H_4 is a product of high-temperature cracking of the side chains of coal molecules. Simultaneously, the concentration of C_2H_4 exceeds that of C_2H_6 during the high-temperature stage, demonstrating that C_2H_4 exhibits greater sensitivity to high temperatures than C_2H_6.

Changes in $CO/\Delta O_2$ precisely reflect the amount of CO produced per unit of oxygen consumed, exhibiting high sensitivity to temperature. The C_2H_4/C_2H_6 ratio significantly increases in the high-temperature zone, indicating that C_2H_4 is more sensitive to elevated temperatures and exhibits a faster formation rate, making it a key indicator of the stage transition in CSC. The indicator decreased after a period of time, with the formation rate of C_2H_4 gradually decreasing while the formation rate of C_2H_6 gradually increased. Meanwhile, the composite gas indicators $CO/\Delta O_2$ and C_2H_4/C_2H_6 are expressed as ratios, eliminating the effects of air leakage and initial concentrations. They can further serve as indicators for predicting CSC temperatures.

As shown in Fig. 7a,b, the curves of the two datasets exhibit highly consistent trends, which visually demonstrates the consistency of the data distributions and the reliability of the datasets.

5.2.2 Feature Selection

Based on the above analysis of the mechanisms and characteristics of CSC and oxidation, the mechanism by which gas parameters influence changes in coal temperature has been preliminarily identified. Furthermore, this study uses statistical methods to visually demonstrate the coupling relationship between gas indicators and coal temperature. We employ Spearman's correlation analysis to identify key features in this non-normally distributed data.

The Spearman correlation coefficients among the attributes and the corresponding scatter plot matrix are shown in Fig. 8. The lower left corner displays a scatter plot of variables, the diagonal represents the data distribution of corresponding variables, and the upper right corner shows the Spearman correlation coefficients between variables. As shown in the figure, temperature is strongly positively correlated with CO, C_2H_4, $CO/\Delta O_2$, and C_2H_4/C_2H_6, with their correlation coefficients being 0.89, 0.84, 0.79, and 0.91, respectively. C_2H_4/C_2H_6 exhibits the highest correlation with temperature, indicating that C_2H_4/C_2H_6 can sensitively reflect temperature trends during coal oxidation. Furthermore, composite indicator gases in ratio form can eliminate wind speed effects, enhance model generalization capabilities, and improve model stability. Simultaneously, C_2H_4 and CO also exhibit a significant positive correlation, indicating that these two gases follow identical release patterns during the heating process, which is consistent with the chemical mechanism of CSC. O_2 (%) is perfectly negatively correlated with ΔO_2 and is also the only variable that shows a negative correlation with temperature; this is consistent with the logic that oxygen is continuously consumed during the CSC. However, using both as input variables would cause feature redundancy and reduce model performance, so O_2 (%), which is negatively correlated with temperature, was chosen as the input feature. Meanwhile, the $CO/\Delta O_2$ ratio reflects how the rate of CO production varies with the intensity of oxidation, and it also provides a clearer picture of the oxidation process.

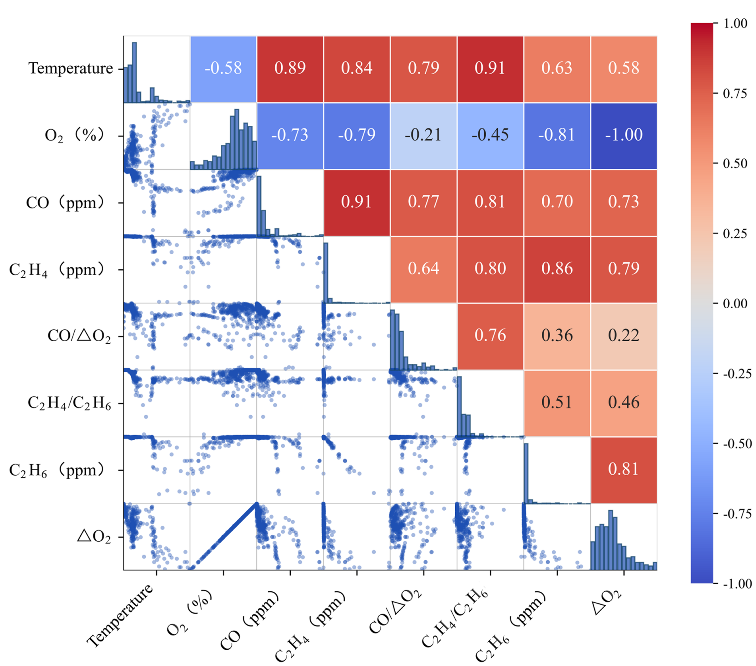

Figure 8: Spearman correlation analysis results.

To further assess whether there were serious multicollinearity issues among the selected features, a Variance Inflation Factor (VIF) analysis was conducted on the features included in the final model. The results of the VIF analysis are shown in Table 2.

Table 2: VIF for features.

Feature	C_2H_4/C_2H_6	CO	C_2H_4	$CO/\Delta O_2$	O_2 (%)
VIF	1.5119	6.6692	1.9655	4.0822	3.2846

The VIF values for all features are below 10, indicating that there is no serious multicollinearity issue among the features. In Table 2, the VIF for CO is 6.6692, which is slightly higher than that of the other indicators. CO concentration directly reflects the oxidation intensity of CSC and is a key indicator for predicting the CSC.

By analyzing the oxidation mechanism of CSC and evaluating the Spearman correlation coefficients and VIF values among the feature variables, this study ultimately selected C_2H_4/C_2H_6, CO, C_2H_4, $CO/\Delta O_2$, and O_2 (%) as the input features for the CLEA-OOA-BiLSTM CSC temperature prediction model.

5.3 Results and Comparison

5.3.1 Evaluation Metrics for the Model

To evaluate the performance of the predictive model, this paper employs five performance metrics—MSE, RMSE, coefficient of determination (R^2), MAE, and MAPE—to assess the model's forecasting performance. The following are the formulas for calculating the metrics:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(t_i - \hat{t}_i\right)^2 \tag{29}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(t_i - \hat{t}_i\right)^2} \tag{30}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|t_i - \hat{t}_i\right| \tag{31}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{t_i - \hat{t}_i}{t_i}\right| \times 100\% \tag{32}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(t_i - \hat{t}_i\right)^2}{\sum_{i=1}^{N}\left(t_i - \bar{t}_i\right)^2} \tag{33}$$

where, N represents the number of samples, $\hat{t}_i$ denotes the predicted temperature value, t_i indicates the actual temperature value, and $\bar{t}_i$ signifies the average of the actual temperatures.

5.3.2 Analysis of Results

The preprocessed training and test sets were input into the constructed CLEA-OOA-BiLSTM model for coal temperature prediction. Fig. 9 shows the prediction results and evaluation metrics.

As shown in Fig. 9, the R^2 values for the train set and test set are 0.9830 and 0.9752, both of which are close to 1. The model can accurately predict the CSC temperature and precisely capture the underlying patterns of how this temperature varies with different factors. The model achieved low RMSE and MAE values on both datasets (training set: RMSE = 3.3163, MAE = 1.9788; test set: RMSE = 3.3561, MAE = 2.7276), with MAPE values of 2.129% and 2.739%, respectively—both below 5%—demonstrating excellent generalization ability and stability. The proposed CLEA-OOA effectively optimizes the hyperparameters of BiLSTM, and the CLEA-OOA-BiLSTM model performs exceptionally well in the task of predicting CSC temperatures.

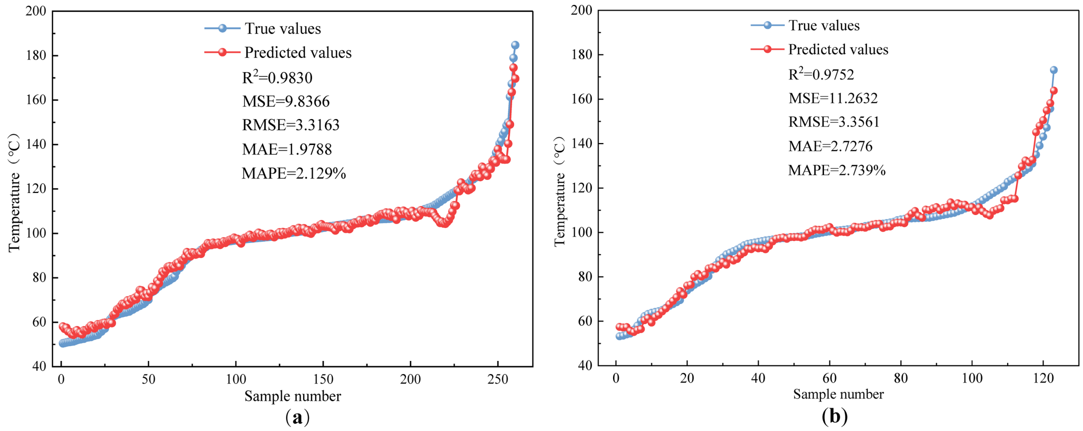

Figure 9: CSC temperature prediction results based on the CLEA-OOA-BiLSTM. (**a**) Train set; (**b**) test set.

To further validate the performance of the CLEA-OOA and demonstrate the superiority of CLEA-OOA-BiLSTM model in predicting CSC temperatures, this paper conducted comparative experiments involving multiple models. Four base models—RF, XGBoost, GRU, and BiLSTM—and four optimized models—CLEA-OOA-RF, CLEA-OOA-XGBoost, CLEA-OOA-GRU, and CLEA-OOA-BiLSTM—were selected. All models are fed with the selected indicator gases, and evaluation metrics are calculated for quantitative comparative analysis. Fig. 10 and Table 3 present the model's prediction results and evaluation metrics.

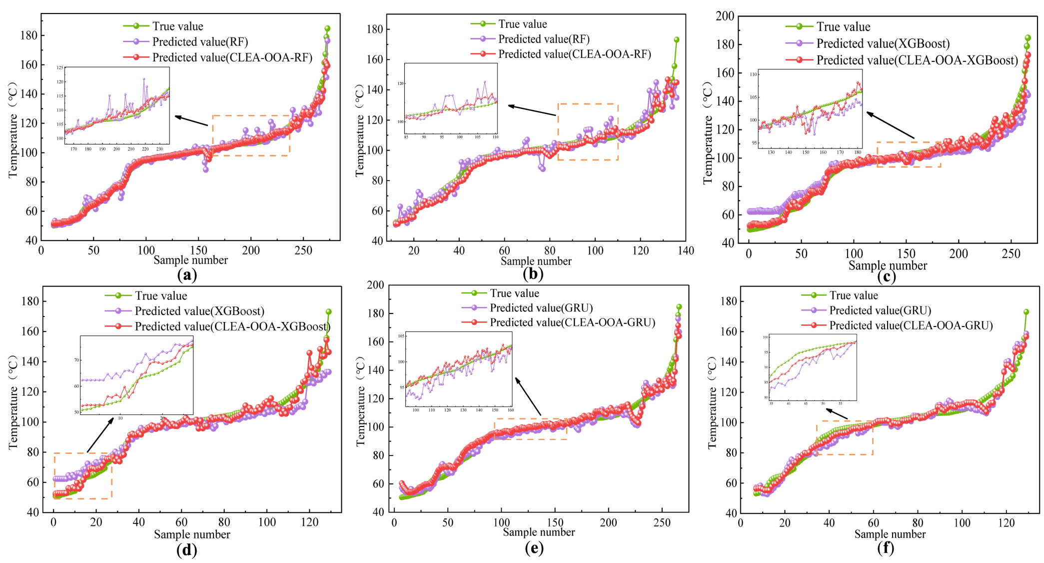

Figure 10: (Continued)

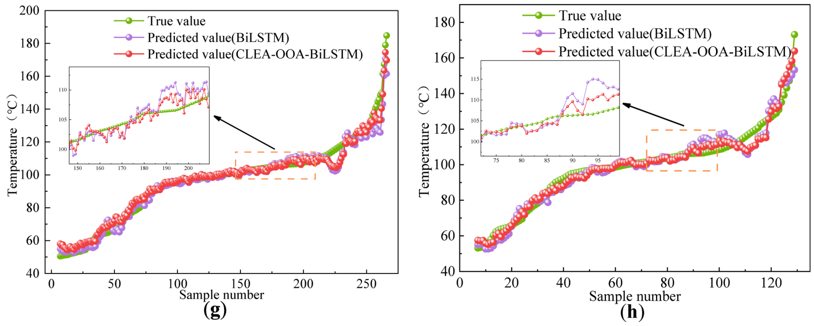

Figure 10: Comparison of model prediction results. (**a**) RF and CLEA-OOA-RF: train set; (**b**) RF and CLEA-OOA-RF: test set; (**c**) XGBoost and CLEA-OOA-XGBoost: train set; (**d**) XGBoost and CLEA-OOA-XGBoost: test set; (**e**) GRU and CLEA-OOA-GRU: train set; (**f**) GRU and CLEA-OOA-GRU: Test set; (**g**) BiLSTM and CLEA-OOA-BiLSTM: train set; (**h**) BiLSTM and CLEA-OOA-BiLSTM: test set.

Table 3: Model performance metrics.

Model	MSE		RMSE		MAE		MAPE (%)		R^2	
	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
RF	11.25	33.47	3.35	5.79	1.74	3.28	1.73	3.36	0.98	0.95
CLEA-OOA-RF	7.23	16.59	2.69	4.07	1.32	2.67	1.27	2.62	0.99	0.97
XGBoost	394.66	46.53	19.87	6.82	12.88	4.42	10.89	5.14	0.93	0.91
CLEA-OOA-XGBoost	7.71	17.56	2.78	4.19	1.98	2.62	2.15	2.78	0.99	0.97
GRU	21.88	25.12	4.68	5.01	3.11	3.84	3.67	4.30	0.96	0.95
CLEA-OOA-GRU	17.73	14.48	4.21	3.81	2.82	2.79	3.11	2.78	0.97	0.97
BiLSTM	22.09	22.48	4.70	4.74	2.80	3.70	2.97	3.94	0.96	0.96
CLEA-OOA-BiLSTM	9.84	11.26	3.14	3.36	1.99	2.73	2.13	2.74	0.98	0.98

As shown by the fitted curve in Fig. 10, the model curve exhibits reduced error and fluctuation after incorporating CLEA-OOA, with the prediction results generally demonstrating superior curve-fitting performance compared to the unoptimized model. The R^2 values for RF, XGBoost, GRU, and Bi-LSTM on the test set are 0.95, 0.91, 0.95, and 0.96, respectively. After optimizing the four baseline models by introducing CLEA-OOA, the R^2 values are 0.97, 0.97, 0.97, and 0.98, respectively, indicating an improvement in prediction accuracy across the board. Compared to the baseline model, the error metrics of the CLEA-OOA-optimized model have decreased significantly. Specifically, the MAPE of CLEA-OOA-GRU is 3.11%, representing a 15.3% reduction compared to the MAPE of 3.67% for GRU. An analysis of Table 3 and Fig. 10 shows that the CLEA-OOA algorithm, which integrates the Chebyshev chaotic mapping strategy, Lévy flights strategy, enhanced pinhole imaging backward learning strategy, and an adaptive weighting strategy, can effectively optimize model hyperparameters, thereby further improving predictive performance.

As shown in Table 3 and Fig. 10, the R^2 of the CLEA-OOA-BiLSTM reaches 0.98, approaching 1. This not only outperforms the four baseline models—RF, XGBoost, GRU, and LSTM—but also surpasses the CLEA-OOA-RF, CLEA-OOA-XGBoost, and CLEA-OOA-GRU models. On the test set, CLEA-OOA-BiLSTM achieved an R^2 of 0.98, which is 2.1% points higher than that of BiLSTM (R^2 = 0.96). CLEA-OOA-BiLSTM also performs exceptionally well on error metrics, with the following results: (Train set: MSE = 9.84,

RMSE = 3.14, MAE = 1.99, MAPE = 2.13%; Test set: MSE = 11.26, RMSE = 3.36, MAE = 2.73, MAPE = 2.74%), all of which are relatively low, demonstrating the model's error control and generalization capabilities.

5.3.3 Ablation Experiment

To further verify the contributions of various optimization strategies in the CLEA-OOA algorithm to the performance of CSC temperature prediction, this paper conducted ablation experiments focused on the task of predicting CSC temperatures. While maintaining consistent experimental conditions, we constructed corresponding ablation models by sequentially removing the Chebyshev chaotic map, the Lévy flight strategy, the enhanced pinhole imaging strategy, and the adaptive weighting mechanism, and compared the evaluation metrics of the models. Table 4 shows the prediction results of the different ablation models.

Table 4: Ablation experiment results.

Model	R^2	MSE	RMSE	MAPE (%)
CLEA-OOA-BiLSTM	0.9738	12.2407	3.4987	2.6571
Without Chebyshev	0.9682	14.451	3.8015	2.7774
Without Lévy Flight	0.964	16.3234	4.0402	2.9298
Without Pinhole Imaging	0.9688	14.1806	3.7657	2.8634
Without Adaptive Weight	0.9694	13.8729	3.7246	2.686

As shown in Table 4, the full CLEA-OOA-BiLSTM model achieved the best predictive performance, with an R^2 of 0.9738 and MSE, RMSE, and MAPE values of 12.2407%, 3.4987%, and 2.6571%, respectively. Performance declined most significantly after removing the Lévy flight strategy, with R^2 dropping from 0.9738 to 0.9640, a decrease of 1.01%. MSE increased from 12.2407 to 16.3234, a rise of 33.35%, while RMSE increased by 15.47% and MAPE by 10.27%. This demonstrates that the Lévy flight strategy plays a crucial role in enhancing the population's global search capabilities and preventing it from getting trapped in local optima. Furthermore, the removal of the Chebyshev chaotic map, enhanced pinhole imaging backpropagation, and the adaptive weighting strategy also led to a decline in the model's predictive performance, further validating the effectiveness of hyperparameter optimization for each of these improvement strategies. The synergistic effects of the various improvement strategies in CLEA-OOA collectively enhance the algorithm's optimization capabilities and the accuracy of CSC temperature predictions.

5.4 Model Interpretability Analysis with SHAP

SHAP is a model explanation method based on the Shapley value from game theory. SHAP calculates the contribution of each feature to the model's output to quantify the importance of the features. The formula for calculating the SHAP value can be expressed as:

$$\Phi_i = \sum_{S \in m/\{i\}} \frac{|S|! \cdot (|m| - |S| - 1)}{|m|!} \left[f\left(S \cup \{i\}\right) - f\left(S\right) \right] \tag{34}$$

In the formula, S denotes any subset of m that excludes feature i, $|m|$ and $|S|$ denote the number of features contained in sets m and S, respectively, ! is factorial, and $\left[f\left(S \cup \{i\}\right) - f\left(S\right) \right]$ represents the marginal contribution brought by the feature S when added to the subset.

The CLEA-OOA-BiLSTM model features a complex structure and is considered a "black-box model," lacking intuitive analysis of the interpretability of input features. Therefore, this research employs the

SHAP analysis method to interpret the model. This approach helps to reveal the coupling relationships among features and identify key influencing factors, thereby enhancing model transparency. To achieve a comprehensive interpretation of the CLEA-OOA-BiLSTM model, this paper conducts SHAP analysis from both global and local single-sample perspectives.

Figs. 11 and 12 provide a global interpretation of how input features influence the model's output. Fig. 10 is a bar chart of feature importance based on the average absolute SHAP value, which visually illustrates the overall contribution of input features and ranks the features in descending order of importance. An analysis of Fig. 11 reveals that the characteristic contribution value of C_2H_4/C_2H_6 is 1.552, ranking first and making it the core contributing variable in the model. This indicator reflects the coal pyrolysis process and reaction intensity in the form of a ratio and demonstrates the strongest predictive ability regarding temperature changes. The global SHAP value for CO is 0.505, making it the second most significant factor influencing the prediction results. This is because CO is a signature gas of the low-temperature oxidation stage of coal and has a significant impact on the prediction of the initial temperature of CSC.

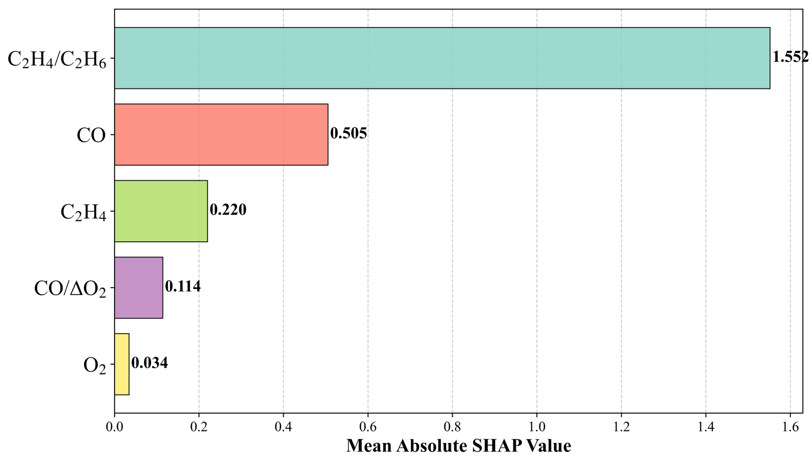

Figure 11: Global feature importance ranking of key gas indicators.

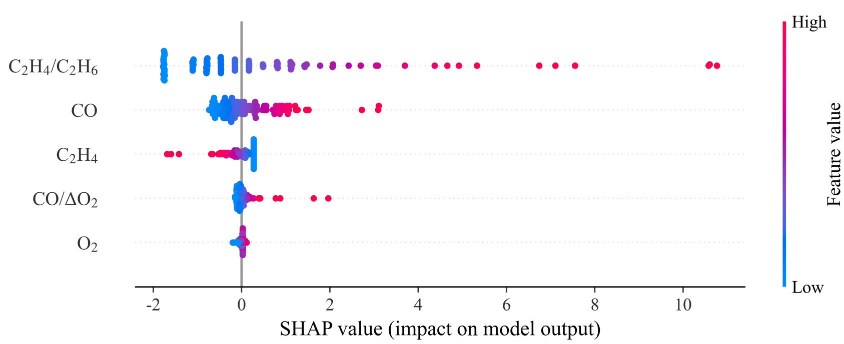

Figure 12: SHAP value distribution.

Fig. 11 depicts the SHAP summary swarm plot, illustrating the specific distribution of SHAP values across all samples. Y-axis and X-axis represent the names of the input features and their SHAP values, respectively, and the features have been sorted in descending order based on their average absolute SHAP values. In SHAP theory, a SHAP value greater than 0 indicates that the feature has a positive impact. A negative value, on the other hand, indicates a negative impact. In the figure, each data point represents a sample, and the scatter points transition from blue to red represents values ranging from low to high. The range of data points in each row illustrates the importance of that feature. In Fig. 12, the C_2H_4/C_2H_6 ratio exhibits the widest range of SHAP values; even slight changes in its numerical value can cause significant fluctuations in the SHAP score, making it the most important feature. Meanwhile, in Fig. 11, the C_2H_4/C_2H_6 ratio has the highest average absolute SHAP value, making it the most critical indicator gas for the model, which demonstrates consistency between the two figures. The red scatter points representing C_2H_4/C_2H_6 and CO are concentrated in the region where the SHAP value is greater than 0; and these gas concentrations have a positive influence. Analysis of Fig. 12 shows that the indicator gases are ranked in order of importance as follows: C_2H_4/C_2H_6, CO, C_2H_4, $CO/\Delta O_2$, O_2(%). Among these, C_2H_4/C_2H_6, CO, C_2H_4, $CO/\Delta O_2$ are positive factors influencing CSC, while O_2 exhibits a negative correlation with coal temperature. This aligns with the actual patterns of change in indicator gases during the oxidation process. Furthermore, this ranking is consistent with the order of importance derived from the Spearman correlation analysis, indicating that the selection of model features was appropriate.

In Fig. 11, the C_2H_4/C_2H_6 ratio contributes the most; this indicator eliminates the influence of underground air leakage through the use of a ratio. When the coal temperature exceeds 100°C, the molecular side chains undergo severe cracking, and the C_2H_4/C_2H_6 ratio increases exponentially after 200°C, making it sensitive to transformations during the high-temperature phase. CO is present throughout the entire CSC process and exhibits an upward trend, making it an important predictive indicator. Meanwhile, the $CO/\Delta O_2$ ratio reflects the pattern of CO production per unit of oxygen consumed, aiding in the prediction of the coal's spontaneous combustion temperature. C_2H_4 is a pyrolysis gas that requires a higher activation energy; its appearance is primarily concentrated in the medium-to-high temperature stages. O_2 (%) is a consumed reactant; this gas exhibits relatively minor fluctuations and makes a relatively small contribution. The ranking of the importance of these features is consistent with the mechanism of spontaneous combustion in coal.

Feature importance may be influenced by the specific way samples are partitioned. To verify the robustness of the SHAP interpretability method, we adjusted the ratio of the training set to the test set to 80%:20%, 70%:30%, and 75%:25%, respectively. Under different data partitions, the models underwent independent hyperparameter tuning and training, and the average absolute SHAP values were calculated for each scheme. The results are shown in Table 5.

Table 5: SHAP values for feature contributions.

Training Set: Test Set	C_2H_4/C_2H_6	CO	C_2H_4	$CO/\Delta O_2$	O_2 (%)
Scheme A: 80%:20%	1.552	0.505	0.220	0.114	0.034
Scheme B: 70%:30%	1.624	0.335	0.257	0.057	0.167
Scheme C: 75%:25%	1.770	0.401	0.151	0.146	0.032

The results in Table 5 show that the order of importance for the top three features—C_2H_4/C_2H_6, CO, and C_2H_4—remains consistent across all dataset partitioning schemes, indicating the global stability of the SHAP interpretation results. In Scheme B, the order of importance for the two least influential features—$CO/\Delta O_2$ and O_2 (%)—has changed compared to Schemes A and C. This is due to slight fluctuations resulting from the data distribution and the model training process. The order of importance of these three indicators aligns with the contribution mechanisms of key indicators in the CSC oxidation process and the oxidation mechanism of CSC.

In addition to performing global SHAP analysis on the CLEA-OOA-BiLSTM model, this paper also conducts local explanations for individual samples. The sample with a predicted temperature of 99.154°C is selected. Its SHAP waterfall plot and force plot are shown in Figs. 13 and 14. This temperature point marks the critical temperature at which the coal's spontaneous combustion process transitions from the slow oxidation stage to the rapid oxidation stage. At the same time, the concentrations of characteristic spontaneous combustion gases, such as CO and olefins, begin to rise sharply. A detailed analysis of the 99.154°C temperature point verified the model's predictive capability at this critical temperature and assessed the contribution of the indicator gas, thereby providing a better understanding of the oxidation mechanisms underlying CSC. In the SHAP waterfall plot shown in Fig. 13, the input feature labels are arranged vertically on the left axis. The red or blue bar charts for each feature represent that feature's contribution to the final prediction. The E[f(x)] label on the horizontal axis indicates the baseline value when predicting this sample. The baseline value for the coal temperature prediction result of this sample is 93.508°C. The features C_2H_4/C_2H_6, CO, and $CO/\Delta O_2$ exhibit positive contributions with SHAP values of 4.94°C, 0.83°C, and 0.07°C, respectively. Negative features include C_2H_4 and O_2, with SHAP values of −0.1°C and −0.1°C, respectively. The model's output is a linear combination of the marginal contributions of the positive and negative features and the baseline value.

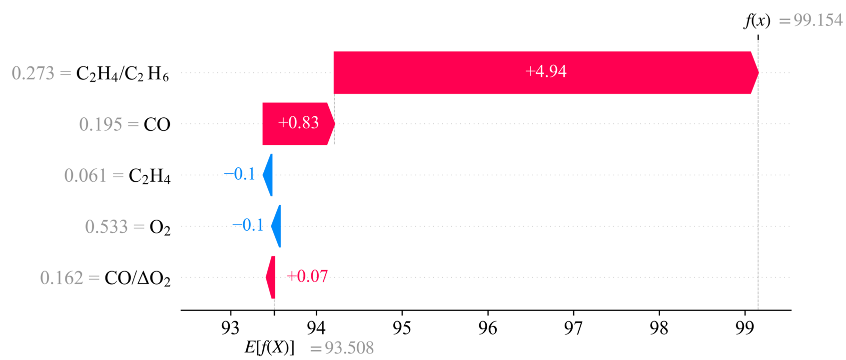

Figure 13: Waterfall plot.

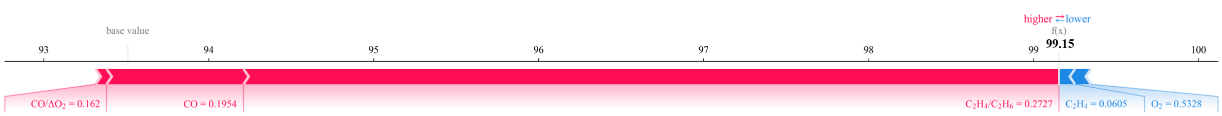

Figure 14: Force plot.

5.5 Model Validation and Application

The CLEA-OOA-BiLSTM model demonstrates excellent accuracy on the aforementioned dataset, while the model's generalization ability is crucial for the practical application of CSC early warning. This study selects programmed heating datasets from five coal mines: Lijiahao (LJH), Madiliang (MDL), Xingjiagou (XJG), Gaojialiang (GJL), and Luling (LL). There is limited data on temperature-programmed experiments, and small-sample statistics can lead to random variations in evaluation metrics. To increase the data sampling density while preserving the continuity and physical characteristics of the original curve, this paper employs Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) interpolation to enhance the original data. We also introduce weak Gaussian noise to simulate measurement errors and construct a new dataset. The augmented dataset is used solely for out-of-model validation and is not involved in the training, hyperparameter optimization, or parameter updating processes of the CLEA-OOA-BiLSTM model. Feed the new dataset into the pre-trained CLEA-OOA-BiLSTM model to predict CSC temperatures. Table 6 shows the model's prediction results.

Table 6: Prediction results of the model for different coal mines.

Coal Mine	R^2	MSE	RMSE	MAE	MAPE (%)
LJH	0.9853	37.5045	6.1241	5.0998	6.6499
MDL	0.9952	9.6259	3.1026	2.6893	2.8819
XJG	0.9967	6.1206	2.474	2.056	2.4462
GJL	0.9946	12.9967	3.6051	2.8197	3.4282
LL	0.9971	3.1281	1.7687	1.3512	1.4584

As shown in Table 6, the R^2 values of the CLEA-OOA-BiLSTM model for coal samples from the five mines are all greater than 0.985. The maximum RMSE is 6.1241, while the minimum RMSE is 1.7687, and the minimum MAPE is 1.4584%. These results indicate that the model has strong generalization capability across coal mine datasets from different regions and demonstrates broad application prospects in practical CSC early warning, providing technical support for mine fire prevention.

5.6 Discussion

The framework proposed in this paper performs exceptionally well in the field of CSC early warning, but it also has some shortcomings. The current models primarily rely on gas indicators to predict the temperatures of CSC, but they do not adequately consider changes in mine air leakage and water seepage, oxygen consumption, and geological pressure. Additionally, the current hyperparameter tuning is guided by a single objective, and incorporating multi-objective optimization to balance prediction accuracy, model robustness, and computational efficiency remains to be explored. In the future, this model can be used to integrate wind speed and oxygen concentration data to investigate the relationship between CSC temperatures and these physical parameters, thereby further enhancing the model's reliability and generalizability.

6 Conclusion

In response to the challenges of preventing CSC, this paper proposes a CLEA-OOA-BiLSTM-SHAP framework for predicting CSC temperatures and enhances model transparency through SHAP analysis. The model's generalizability was verified using data from five different coal mines. The conclusions reached are as follows:

(1) We performed a correlation analysis using the Spearman test and plotted a correlation scatter plot to identify C_2H_4/C_2H_6, CO, C_2H_4, $CO/\Delta O_2$, and O_2 (%) as the model's input features.
(2) An innovative CLEA-OOA algorithm is proposed, which integrates Chebyshev chaotic maps for population initialization, Lévy flight, enhanced pinhole imaging backpropagation, and an adaptive weighting strategy. A comparative experiment was conducted using four heuristic optimization algorithms based on eight benchmark functions. The results show that CLEA-OOA outperforms all other algorithms in terms of both convergence accuracy and convergence rate, effectively validating its ability to optimize hyperparameters.
(3) An CLEA-OOA-BiLSTM model for predicting CSC temperature was proposed and compared with models such as RF, XGBoost, GRU, and CLEA-OOA-RF. The CLEA-OOA-BiLSTM model had R^2 values of 0.98 on both the train and test datasets, both close to 1.0 and outperforming other benchmark models. During the testing phase, its MSE, RMSE, MAE, and MAPE were 11.26%, 3.36%, 2.73%, and 2.74%, respectively, demonstrating excellent predictive performance. Among all models, this model has the best prediction accuracy and the lowest error. To better explain the model's prediction process,

this study introduces SHAP analysis to quantify the contribution of individual features to the results. Among these, C_2H_4/C_2H_6 and CO are the core factors influencing the prediction results. C_2H_4/C_2H_6, CO, C_2H_4, and $CO/\Delta O_2$ exert positive effects, while O_2 (%) exerts a negative effect, which is consistent with the oxidation mechanism of CSC. The importance of indicator gases for predicting coal temperature is ranked as follows: C_2H_4/C_2H_6, CO, C_2H_4, $CO/\Delta O_2$, and O_2. This ranking is consistent with the results of the Spearman correlation analysis, which validates the scientific soundness of the model's decision-making logic. Furthermore, when the CLEA-OOA-BiLSTM framework was applied to data from five different coal mines in Inner Mongolia, Shanxi, and other regions, the model's R^2 values all exceeded 0.985, demonstrating its strong generalization capability and engineering value.

Acknowledgement: Not applicable.

Funding Statement: This research was supported by the Basic Research Fund for Provincial Universities in Hebei Province (No. JJC2024082) and the Provincial College Student Innovation and Entrepreneurship Training Program (No. S202510081070).

Author Contributions: Conceptualization, Rui Yan; methodology, Rui Yan; software, Rui Yan and Botao Fan; validation, Xueqi Qu and Cuihuan Ren; data curation, Rui Yan and Botao Fan; writing—original draft preparation, Rui Yan and Botao Fan; writing—review and editing, Rui Yan, Xueqi Qu and Xu Zhou; visualization, Rui Yan, Xueqi Qu and Cuihuan Ren; supervision, Xu Zhou; project administration, Xu Zhou. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The core dataset for model construction is included in Reference [32], and other data supporting the findings of this study are available from the Corresponding Author, Xu Zhou, upon reasonable request.

Ethics Approval: Not applicable.

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

References

1. Liu Y, Wen H, Chen CM, Guo J, Jin YF, Zheng XZ, et al. Research status and development trend of coal spontaneous combustion fire and prevention technology in China: a review. ACS Omega. 2024;9(20):21727–50. doi:10.1021/acsomega.4c00844.
2. Onifade M, Genc B. A review of research on spontaneous combustion of coal. Int J Min Sci Technol. 2020;30(3):303–11. doi:10.1016/j.ijmst.2020.03.001.
3. Anghelescu L, Diaconu BM. Advances in detection and monitoring of coal spontaneous combustion: techniques, challenges, and future directions. Fire. 2024;7(10):354. doi:10.3390/fire7100354.
4. Yan HW, Nie BS, Liu PJ, Chen ZY, Yin FF, Gong JG, et al. Experimental assessment of multi-parameter index gas correlation and prediction system for coal spontaneous combustion. Combust Flame. 2023;247(5):112485. doi:10.1016/j.combustflame.2022.112485.
5. Guo Q, Ren WX, Lu W. Risk evaluation of coal spontaneous combustion from the statistical characteristics of index gases. Thermochim Acta. 2022;715(1):179287. doi:10.1016/j.tca.2022.179287.
6. Zhang JL, Zhou X, Su JP, Xiao YL. An interpretable machine learning model for optimization of prediction index gases in coal spontaneous combustion. Alex Eng J. 2025;122(1):268–78. doi:10.1016/j.aej.2025.02.104.
7. Muduli L, Mishra DP, Jana PK. Wireless sensor network based underground coal mine environmental monitoring using machine learning approach. In: Chang X, editor. Proceedings of the 11th International Mine Ventilation Congress. Singapore: Springer; 2019. p. 817–26. doi:10.1007/978-981-13-1420-9_66.

8. Kamran M, Shahani NM. Decision support system for the prediction of mine fire levels in underground coal mining using machine learning approaches. Min Metall Explor. 2022;39(2):591–601. doi:10.1007/s42461-022-00569-1.
9. Thuppari C, Jannu S, Edla DR. A novel logistic regression-based fire detection model using IoT in underground coal mines. In: Fortino G, Kumar A, Swaroop A, Shukla P, editors. Proceedings of Third International Conference on Computing and Communication Networks. Singapore: Springer; 2024. p. 115–24. doi:10.1007/978-981-97-0892-5_11.
10. Kumar S, Mishra PK, Kumar M, Pratik, Kumar J. Coactive application of environmental sensors for detection and assessment of spontaneous combustion in underground coal mines. In: Proceedings of the 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS); 2017 Aug 1–2; Chennai, India. p. 631–6. doi:10.1109/ICECDS.2017.8389513.
11. Yetkin ME, Özfirat MK, Kun M, Pamukcu C. The prediction of the risks of spontaneous combustion in underground coal mines using a fault tree analysis method. MethodsX. 2024;13:102835. doi:10.1016/j.mex.2024.102835.
12. Liu XL, Wang H, Wang EY, Li ZH, Liu XF, Wang K. Research on evolution law of index gases produced by the coal spontaneous combustion based on wavelet analysis. Fuel. 2024;370(1):131859. doi:10.1016/j.fuel.2024.131859.
13. Lei CK, Feng QC, Zhu YQ, Cui CB, Bao RY, Deng CB. Multiple indicator gases and temperature prediction of coal spontaneous combustion oxidation process. Fuel. 2025;393(3):134991. doi:10.1016/j.fuel.2025.134991.
14. Dogra SK, Jayanthu S, Samal AK, Pramanik J, Pani SK. Machine learning approach to implement mine fire predicting for underground coal mines. In: Proceedings of the 2021 2nd Global Conference for Advancement in Technology (GCAT); 2021 Oct 1–3; Bangalore, India. p. 1–4. doi:10.1109/GCAT52182.2021.9587499.
15. Zhang LD, Luo ZM, Wang YC, Su B, Zhang H. Study on multi-monitoring point coal spontaneous combustion grading early warning system and GRU model. Fuel. 2026;405(1):136452. doi:10.1016/j.fuel.2025.136452.
16. Pan HG, Fan YB, Deng J, Shi KK, Wang CP, Lei XY, et al. GCN-based prediction method for coal spontaneous combustion temperature. Process Saf Environ Prot. 2025;196(1):106855. doi:10.1016/j.psep.2025.106855.
17. Shukla US, Mishra DP, Mishra A. Prediction of spontaneous combustion susceptibility of coal seams based on coal intrinsic properties using various machine learning tools. Environ Sci Pollut Res. 2023;30(32):69564–79. doi:10.1007/s11356-023-27248-y.
18. Said KO, Onifade M, Lawal AI, Githiria JM. An artificial intelligence-based model for the prediction of spontaneous combustion liability of coal based on its proximate analysis. Combust Sci Technol. 2021;193(13):2350–67. doi:10.1080/00102202.2020.1736577.
19. Said KO, Onifade M, Lawal AI, Githiria JM. Computational intelligence-based models for predicting the spontaneous combustion liability of coal. Int J Coal Prep Util. 2022;42(6):1626–50. doi:10.1080/19392699.2020.1741558.
20. Dogra SK, Pramanik J, Jayanthu S, Samal AK. Machine learning applications: coal fire prediction using graham's ratio. In: Sinha A, Sarkar BC, Mandal PK, editors. Proceedings of the 10th Asian Mining Congress 2023. Cham, Switzerland: Springer; 2023. p. 119–26. doi:10.1007/978-3-031-46966-4_13.
21. Vo Thanh H, Zhang HM, Xu YY, Wang BF, Wang YJ, Dai ZX. Advancing risk prediction of coal spontaneous combustion: intelligent gas index models for accurate self-ignition temperature forecasting towards low-carbon energy systems. Process Saf Environ Prot. 2025;202:107676. doi:10.1016/j.psep.2025.107676.
22. Zhao HM, Zhou X, Han JJ, Liu YX, Liu Z, Wang SS. Research on early warning model of coal spontaneous combustion based on interpretability. Sci Rep. 2025;15(1):18847. doi:10.1038/s41598-025-01154-4.
23. Wu C, Wang G, Song T, Xia T, Niu Y, Guo F, et al. Prediction of coal spontaneous combustion temperature under variable moisture contents: a study based on the random forest model optimized by the sparrow search algorithm. Process Saf Environ Prot. 2026;209:108586. doi:10.1016/j.psep.2026.108586.
24. Lawal AI, Onifade M, Abdulsalam J, Aladejare AE, Gbadamosi AR, Said KO. On the performance assessment of ANN and spotted hyena optimized ANN to predict the spontaneous combustion liability of coal. Combust Sci Technol. 2022;194(7):1408–32. doi:10.1080/00102202.2020.1815196.

25. Ding HL, Li LH, Jia GJ, Zhao CL, Li JX, Yan S, et al. Research on coal spontaneous combustion temperature prediction model based on BO-LSTM. Process Saf Environ Prot. 2025;201(4):107570. doi:10.1016/j.psep.2025.107570.
26. Kong B, Wan HJ, Zhu SX, Zhang WR, Song SL, Zhang XL, et al. Development and implementation of an intelligent early warning system for preventing environmental pollution from coal spontaneous combustion. Green Smart Min Eng. 2025;2(3):313–29. doi:10.1016/j.gsme.2025.09.005.
27. Bai G, Liang R, Wang W, Qi Y, Zhang XW, Li XM, et al. Dynamic early warning model for coal spontaneous combustion based on machine learning. Nat Resour Res. 2026;35(10):102–18. doi:10.1007/s11053-026-10649-4.
28. Shao XM, Liu WH, Bai G, Chen Y, Liu Y, Guang JH. Deep learning framework based on ITOC optimization for coal spontaneous combustion temperature prediction: a coupled CNN-BiGRU-CBAM model. Sci Rep. 2025;15(1):26700. doi:10.1038/s41598-025-11294-2.
29. Shao XM, Zhang XJ, Liu WH, Li YC, Jia JZ, Li ZX, et al. Multi-indicator feature extraction and temperature prediction model for spontaneous coal combustion and oxidation. Sci Rep. 2026;16(1):840. doi:10.1038/s41598-025-30395-6.
30. Dehghani M, Trojovský P. Osprey optimization algorithm: a new bio-inspired metaheuristic algorithm for solving engineering optimization problems. Front Mech Eng. 2023;8:1126450. doi:10.3389/fmech.2022.1126450.
31. Schuster M, Paliwal KK. Bidirectional recurrent neural networks. IEEE Trans Signal Process. 1997;45(11):2673–81. doi:10.1109/78.650093.
32. Jiang P. Research on coal spontaneous combustion temperature prediction model based on machine learning [master's thesis]. Xi'an, China: Xi'an University of Science and Technology; 2020. doi:10.27397/d.cnki.gxaku.2020.000257.