International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 7534-7544

https://doi.org/10.70517/ijhsa463646

Particle swarm optimization of three-dimensional residual neural networks for coal mill outlet pressure prediction with anomaly pattern recognition

Shiguang Guo[1,*]
[1] Beijing Information Science and Technology University, Beijing, 100025, China
Corresponding authors: (e-mail: xinY11272018@163.com).

Abstract The traditional pattern recognition method of abnormal coal mill system has low accuracy and is difficult to meet the industrial production demand. In this paper, we propose a coal mill outlet pressure prediction and anomaly pattern recognition method based on particle swarm optimization three-dimensional residual neural network. The method firstly adopts the 3σ criterion to eliminate abnormal data and performs mean filtering preprocessing, constructs a 3D residual neural network integrating soft thresholding sub-network and distraction mechanism, and optimizes the network hyper-parameters using particle swarm algorithm. Experiments show that, compared with the traditional method, this model is significantly better than the genetic algorithm in convergence speed, shortening the training time by more than 30%; it performs excellently in prediction accuracy, with the MAE value reduced to 21.473, the MAPE reduced to 0.0333, and the RMSE reduced to 20.069, which is reduced by 10.4%, 38.4%, and 15.3% compared with the traditional ResNet, respectively. In terms of anomaly pattern recognition, the model's recognition accuracy for current anomaly, temperature anomaly, pressure anomaly, flow anomaly, and rotational speed anomaly reaches 94.7%, 97.5%, 97.7%, 96.1%, and 98.5%, respectively, with an accuracy far exceeding that of traditional classification algorithms. The results confirm that the particle swarm optimized 3D residual neural network has significant advantages in the prediction of coal mill outlet pressure and identification of abnormal states.

Index Terms coal mill outlet pressure, anomaly pattern recognition, particle swarm optimization, three-dimensional residual neural network, prediction accuracy, distraction mechanism

I. Introduction

Coal mill is the main auxiliary equipment in coal-fired power plants, and its operating condition directly affects the unit load output, the safe and economic operation of the boiler, and the maintenance cost [1], [2]. As the coal mill failure is easily affected by external environment, coal quality and other factors, the complex working environment presents a certain challenge to the operation reliability of the coal mill [3], [4]. Therefore. Effective fault warning and diagnosis of coal mill to ensure the normal operation of thermal power plants is very necessary.

The pulverizing system is an essential part of the pulverized coal boiler system, and parameters such as the outlet pressure of the coal mill outlet air-powder mixture will directly affect the boiler combustion efficiency and safe operation [5]. Effective fault diagnosis and early fault warning can help operators to fully understand the coal mill operating status and propose corresponding measures in time [6], [7]. At the present stage, in most power plants, in the process of arriving at the furnace from the coal mill outlet, the primary air ducts have different lengths and resistances, which lead to a large deviation in the amount of air powder sent into the furnace [8]-[10]. Therefore, through the collection of coal mill DCS data, coal mill offline data, coal quality industrial analysis data, ash carbon content and other data, the establishment of appropriate data model to carry out the coal mill outlet pressure prediction can effectively realize the safe and economic long-cycle operation of the coal mill, which is of great significance to reduce production costs [11]-[14].

Thermal power generation is the main power supply method in China, and the stable operation of its equipment is crucial to the safety of power grid. Coal powder preparation system is one of the key equipments in thermal power plant, in which coal mill is responsible for grinding raw coal into fine powder for full combustion in boiler. As an important parameter characterizing the working status of the pulverized coal preparation system, abnormal fluctuations of the coal mill outlet pressure will not only lead to a reduction in the operating efficiency of the coal mill, but also may cause safety accidents. Currently, a simple threshold-based alarm method is often used in industrial sites to monitor the coal mill outlet pressure, which leads to a high false alarm rate and serious leakage. Although traditional machine learning methods such as Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) have achieved certain results in the field of anomaly detection, their feature extraction capability is limited, and it is difficult to adapt to the demand for accurate prediction and classification under complex working conditions. In recent years, deep learning has shown great potential in the field of industrial process monitoring and fault diagnosis, and especially convolutional neural networks have been widely used due to their powerful automatic feature extraction capability. However, standard convolutional networks still suffer from the problems of information loss and gradient vanishing when dealing with multi-scale nonlinear industrial data. Residual neural networks effectively alleviate the problem of difficult training of deep networks by introducing cross-layer connections, but how to determine the optimal network structure parameters is still a major challenge in practical applications. On the other hand, particle swarm optimization algorithm, as an efficient global optimization method, performs well in solving complex nonlinear optimization problems, and combining it with deep learning is expected to significantly improve the model performance. Based on this, this study proposes a particle swarm optimization 3D residual neural network model for coal mill outlet pressure prediction and anomaly pattern recognition. The model first preprocesses the real-time data collected from the industrial site to remove the anomalies and reduce the influence of random noise; then constructs a 3D residual neural network that integrates soft thresholding sub-network and distraction mechanism to enhance the ability of identifying local anomalous features; finally, the particle swarm algorithm is used to automatically optimize the network hyper-parameters, avoiding the problem of inefficiency of the traditional manual adjustment of parameters. This study aims to solve the problems of insufficient prediction accuracy of coal mill outlet pressure and low recognition rate of anomalous patterns. By combining traditional optimization algorithms with deep learning models, the potential laws of the data are mined to improve the prediction accuracy and classification accuracy, so as to provide theoretical guidance and technical support for the safe and stable operation of the coal mill system.

II. Coal Mill Outlet Pressure Prediction and Anomaly Pattern Recognition Exploration

II. A. Data Acquisition and Preprocessing

II. A. 1) Data acquisition

The coal mill outlet pressure and abnormal mode data acquisition is measured by the sensors on the industrial site, and converted into 4-20m A standard analog signal input to the DCS (distributed control) system through the signal converter, with the help of the data acquisition module, the real-time data collection is realized, specifically the coal mill outlet pressure data and abnormal mode data (abnormal current, abnormal temperature, abnormal pressure, abnormal flow, abnormal speed).

II. A. 2) Data pre-processing

The accuracy of the sample data used for modeling directly affects its reliability and practicality, due to the impact of environmental factors in the industrial field so that the sensor measurement signal jitter noise, so the existence of errors in the field collected data is unavoidable, in order to reduce the error due to the perturbation brought about by the data, so before analyzing the data need to pre-processing of the data.

The data preprocessing algorithm designed through the MATLAB language includes (1) using the 3σ criterion to exclude abnormal data, and (2) filtering the data using mean filtering. Take 5000 sets of sample data in 10 minutes as an example of data preprocessing.

(1) Excluding abnormal data using 3σ criterion

Sample data $N = 1000$, the specific steps of the algorithm are as follows:

Step1: Average $\overline{x}$ over N data of variable x_i, see equation (1).

Step2: Calculate the deviation e_i of each item of x_i from the mean, see equation (2).

Step3: Calculate the standard deviation σ, see formula (3)

Step4: Exclude abnormal data points according to $e_i > 3\sigma$.

Can be derived:

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

$$e_i = x_i - \overline{x} \tag{2}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} e_i^2}{N-1}} \tag{3}$$

(2) Filtering by mean filtering

The sample data with outliers eliminated by 3σ is then filtered. The value of m in the mean value filter affects the sensitivity and smoothness of the data, when the value of m is larger, the sampled data is smoother, but its sensitivity is lower. m is smaller, the smoothness of the data is lower, but the sensitivity is higher. So for different types of sampled data, the value of m is different. The formula is shown below:

$$x_{i'} = \frac{1}{m} \sum_{j=i-m}^{i} x_j \tag{4}$$

II. B.Modeling

II. B. 1) Three-dimensional residual neural networks

The structure of the convolutional neural network used for training in deep learning can be divided into four parts: input layer, convolutional layer, pooling layer, and fully connected layer. Compared with general neural networks, the most prominent feature of convolutional neural networks is the addition of convolutional layers and pooling layers, and the composition of the remaining layers is basically the same as other neural networks. In the training process of neural networks, it is often accompanied by a large number of computational parameters. In the convolutional neural network, the structure of the convolutional layer is paired with the pooling layer, which can effectively realize the downsampling operation, thus reducing the parameters in the training and making the network more efficient in the process of backpropagation.

(1) Input layer

Generally, data such as pictures are used as input, and in this paper, the preprocessed data above is used as input, and the dimension of the input layer is 3.

(2) Convolutional layer

The convolution operation is the basis of the convolutional neural network model, which is the operation of doing the inner product of the window data and filter matrix of different images. Each convolutional layer contains multiple convolutional kernels, the significance of setting multiple convolutional layers is to extract feature information from shallow to deep, the deeper the convolutional kernel can be extracted to deeper features. The formula for the convolution operation is shown in equation (5):

$$Conv = \sum_{i=-\infty}^{\infty} x(i)h(n-i) = x(n)h(n) \tag{5}$$

And the overall formula for the convolution operation is shown in equation (6):

$$\begin{cases} Conv = \sum_{j} w_j y_j \\ z = g(Conv + a) \end{cases} \tag{6}$$

where $Conv$ denotes the convolution operation, y_j is the input feature of the convolution layer, w_j is the convolution kernel parameter, and g is the nonlinear activation function, and the more commonly used activation function is ReLU, which possesses the advantages of fast convergence and simple gradient finding.

(3) Pooling layer

The pooling layer is usually set between two adjacent convolutional layers, the role of the pooling layer is to take the regional average or maximum value, that is, the output data of the previous convolutional layer for the dimensionality reduction process, in order to ensure that the effective feature parameters are retained to reduce the parameters of the network as much as possible. The pooling operation is a convolution operation on the data using a special convolution kernel. The structure of the convolutional neural network used for training in deep learning can be divided into four parts: input layer, convolutional layer, pooling layer, and fully connected layer. Compared with general neural networks, the most prominent feature of convolutional neural networks is the addition of convolutional layers and pooling layers, and the composition of the remaining layers is basically the same as other neural networks. In the training process of neural networks, it is often accompanied by a large number of computational parameters. In the convolutional neural network, the structure of the convolutional layer is paired with the pooling layer, which can effectively realize the downsampling operation, thus reducing the parameters in the training and making the network more efficient in the process of backpropagation.

(4) Fully connected layer

The fully connected layer in the network, each of whose nodes is connected to each node in the previous layer, has more parameters in this network layer because it synthesizes the outputs of all previous network layers. This network layer can be viewed as a neural network structure with hidden layers. The convolutional neural network is a high-level abstract feature obtained by stacking the convolutional layer with the pooling layer after a number of convolutional and pooling operations. The input to the fully connected layer is an N-dimensional vector formed by the operations of convolution and pooling, and the matrix features are extracted again through the fully connected layer and output to the SoftMax layer to get the classification result. When the number of neural network layers is relatively high, it is difficult to be trained, and is prone to problems such as gradient explosion and gradient disappearance, resulting in a deep neural network that is difficult to converge during training, and the recognition accuracy is reduced. To address this problem, residual networks are proposed, which use shortcut connections to retain all the original information and reduce the network parameters. The emergence of residual networks makes it possible to train deep neural networks and can greatly improve the recognition and prediction accuracy of deep networks.

With the deepening research on residual networks, several improved residual network models have been derived from the original structure, which are widely used in various industries, and residual networks have become a milestone breakthrough in the field of deep learning with their good performance [15], [16]. The schematic diagram of the residual unit is shown in Fig. 1, $F(X)$ denotes the residual structure, and $H(X)$ denotes the output of the residual network unit, so $H(X)=F(X)+X$, X introduced by the residual network is a constant mapping, and the gradient can be transmitted back to the lower network through X, which is the reason why the residual network can improve the phenomenon of disappearing gradient, and at the same time, improve the arithmetic ability of the network.

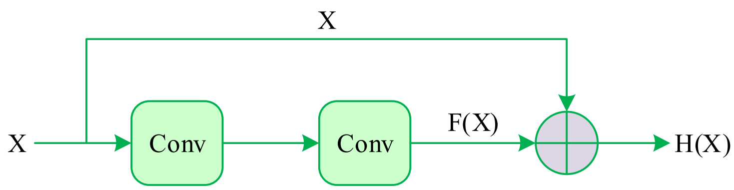


Figure 1: Schematic diagram of residual unit

II. B. 2) Particle Swarm Algorithm

Particle Swarm Optimization (PSO) algorithm is an optimization strategy based on group intelligence in the field of computational intelligence, in addition to traditional algorithms such as Genetic Algorithm and Fish Swarm Algorithm, PSO has been widely used due to its concise and efficient searching ability [17], [18]. In nature, bird flocks search for food sources efficiently by following their peers who are closest to the food, based on this phenomenon, PSO algorithm simulates the social behavior of bird flocks to find the optimal solution of the problem.

When a problem with multiple optimization objectives exists, the multi-objective optimization problem with n decision variables and its m objectives can be expressed as follows:

$$\min y = F(x) = \left(f_1(x), f_2(x), \cdots, f_m(x)\right) \tag{7}$$

$$s.t. \begin{cases} g_i(x) \geq 0 & i = 1,2,\cdots,p \\ h_j(x) = 0 & j = 1,2,\cdots,q \end{cases} \tag{8}$$

where $x = (x_1, x_2, \cdots, x_n) \in X \subset R^n$ is the n-dimensional decision space. $y = (y_1, y_2, \cdots, y_n) \in Y \subset R^m$ is the m-dimensional goal space. $g_i(x) \geq 0\left(i = 1,2,\cdots,p\right)$ are p inequality constraints. $h_j(x) = 0\left(j = 1,2,\cdots,q\right)$ is a q-equality constraint.

Optimization problems involve the selection of solutions and parameters to make the optimal value (maximization or minimization) of a particular performance metric under given constraints. The strategies used in this process are known as optimization methods. These methods are based on sound principles and mechanisms and are executed according to established rules with the aim of finding the best solution to the problem.

In the PSO algorithm, each particle represents a possible solution in the solution space, and these solutions are measured by the evaluation of the fitness function. The speed and direction of the particle's movement is determined by its historical behavior as well as the experience of the other members of the population, i.e., the optimal solution is found through individual and population learning.

The algorithm first initializes the population, a step that involves setting the population size, the maximum number of iterations, and configuring the initial speed and position of the particles. It also involves setting parameters such as inertia weights and learning factors.

At the beginning of the algorithm startup, a group of particles will be randomly assigned initial positions and velocities to provide a starting point for the search process. As the algorithm iteratively proceeds, the particles adjust their velocities and positions according to the guidance of the individual optimum (Pbest) and the group optimum (Gbest) in order to search effectively in the solution space. In this process, the mechanism of updating the particle's velocity and position is the core of the algorithm, which determines how the particle uses the acquired information to optimize the search.

In a D-dimensional search space, a population of n particles can be represented as $X = (X_1, X_2, \cdots, X_n)$, where the position of each particle is described by a D-dimensional vector $X_i = (X_{i1}, X_{i2}, \cdots, X_{iD})^T$, representing a potential solution to the problem. Accordingly, the velocity vector $V_i = (V_{i1}, V_{i2}, \cdots, V_{iD})^T$ of each particle indicates the direction and magnitude of its movement in the solution space.

The algorithm adjusts the velocity and position of each particle in each iteration by the following update rule:

$$V^{k+1} = \omega V^k + c_1 r_1 \left(P_{id}^k - X^k\right) + c_2 r_2 \left(P_{gd}^k - X^k\right) \tag{9}$$

$$X^{k+1} = X^k + V^{k+1} \tag{10}$$

where ω - inertia weights, which determine the propensity of the particle to maintain the current direction of velocity.

$c1$, $c2$ - are the individual learning factor (or cognitive parameter) and the social learning factor (or social parameter), respectively, which determine the propensity of the particle to move toward the individual historical best position and the global best position.

$r1$, $r2$ - are random numbers in the range $[0,1]$, in order to increase the randomness of the search.

V^{k+1} - the velocity of the particle at the next iteration ($k+1$ th).

V^k - the velocity of the particle at the current iteration (k th).

P_{id}^k - individual particle to the best position found so far (individual best).

P_{gd}^k - the entire particle population to the best position found so far (global best).

X^k - the position of the particle at the current iteration.

X^{k+1} - the position of the particle at the next iteration.

In order to avoid too random search behavior of the particles, the range of positions and velocities is usually limited to a preset upper and lower bound range $[-X_{\max}, X_{\max}]$.

The design subtlety of the PSO algorithm is that it reduces the complex optimization problem to the result of group collaboration, where the behavior of each particle is both independent and interconnected, enabling the algorithm to find high-quality solutions in a wide range of applications. PSO has been proven to exhibit excellent performance in many engineering optimization problems, such as network layout, scheduling problems, and machine learning parameter optimization.

II. B. 3) Particle Swarm Algorithm for Optimizing 3D Residual Neural Networks

Although three-dimensional residual neural networks have achieved some success in gearbox fault classification, the following problems still exist:

(1) As the number of network layers deepens, a large number of stacked convolutional layers make the computation speed slow, which leads to a decrease in model performance.

(2) Training a neural network from scratch usually requires a large amount of data, while in real industrial production, the abnormal state of the equipment is often small, and a large amount of data cannot be obtained. At the same time, the arithmetic power and time required for training from scratch also cause the cost of anomaly diagnosis to rise.

(3) In a complete rotation cycle, abnormal features are often concentrated locally, but it is difficult for neural networks to focus on local abnormal features with limited computational resources.

(4) The hyper-parameters have an important influence on the neural network, but the manual adjustment of the hyper-parameters is inefficient and not easy to get the optimal solution. Aiming at the above problems, this chapter proposes a particle swarm optimization based three-dimensional deep residual network. Firstly, a soft thresholding sub-network is added to partially zero the multi-channel feature noise after convolution, which makes the network have good noise filtering performance. Secondly, group convolution is applied to a large number of stacked convolutional layers to reduce the number of parameters in the model, so as to improve the computational performance of the model. Then the network is pre-trained in the dataset first, so that it has the ability to recognize the general type of features, and then use the migration learning method to input the fault data set for training, which solves the problem of the lack of labeled data and the need to consume a lot of arithmetic power and time to train the neural network from scratch. The distraction mechanism of channel weighting is also used to enhance the network's ability to recognize local fault features. Finally, the particle swarm optimization method is used to optimize the network hyperparameters, which avoids the problem of inefficiency of manual parameter tuning.

Table 1: Detailed architecture of each network layer of the model

Method	Network layer	Convolution kernel	Parameter quantity
Conv1	Conv2d BatchNorm2d ReLU	3×3,128 - -	1714 256 -
Conv2-1	Conv2d BatchNorm2d ReLU Shrinkage Split Attention	3×3,128 - - - -	2398 - - - -
Conv2-2	Conv2d BatchNorm2d ReLU Shrinkage Split Attention	3×3,128 - - - -	2398 - - - -
Conv3-1	Conv2d BatchNorm2d ReLU Shrinkage Split Attention	3×3,128 - - - -	1054 - - - -
Conv3-2	Conv2d BatchNorm2d ReLU Shrinkage Split Attention	3×3,128 - - - -	1054 - - - -
Conv4-1	Conv2d BatchNorm2d ReLU Shrinkage Split Attention	3×3,128 - - - -	1054 - - - -
Conv4-2	Conv2d BatchNorm2d ReLU Shrinkage Split Attention	3×3,128 - - - -	1054 - - - -
Conv5-1	Conv2d BatchNorm2d ReLU Shrinkage Split Attention	3×3,128 - - - -	1054 - - - -
Conv5-2	Conv2d BatchNorm2d ReLU Shrinkage Split Attention	3×3,128 - - - -	1054 - - - -
AvgPoolfc	AvgPool2d Linear	- -	- 4072

The coal mill outlet pressure prediction and anomaly pattern recognition model based on particle swarm optimization three-dimensional deep residual network is designed. Firstly, the pre-processed data are input into the improved deep residual network for pre-training, then the model output classification number is changed to the target classification number, and finally the network weights of the pre-trained model are migrated into the model. The model consists of six structural layers, structural layer 1 is a 64×3×3 convolution, structural layers 2-5 consist of convolutional layer, batch normalization layer, activation function layer, soft thresholding sub-network Shrinkage with distraction module Split Attention. The function of structural layer 6 is to classify the output, and each structural layer is connected by a cross-layer constant connection shortcut to avoid the gradient of the deep neural network disappears or explodes, and the three-dimensional residual neural network hyperparameters are determined by the particle swarm algorithm. Table 1 shows the detailed architecture of each network layer of the model, including the information of the structural and network layers, the size of the convolution kernel, and the number of parameters of each layer.

III. Model Example Analysis

III. A. Particle swarm algorithm optimization effect analysis

III. A. 1) Control algorithm setup

In order to verify the performance of particle swarm optimization algorithm (PSO) on three-dimensional residual neural network, the particle swarm optimization algorithm (PSO) proposed in this paper is compared with the traditional whale optimization algorithm (WOA) as well as more mainstream genetic algorithms (GA), Bayesian optimization (BO), and ant colony algorithm (ACO) in the experiments. The experiments are trained using the same data that have been preprocessed in the previous section (coal mill outlet pressure data and anomaly pattern data), and the maximum value of the test accuracy is taken as the result.

III. A. 2) Analysis of results

The accuracy comparison results of the training process of different optimization algorithms are shown in Fig. 2, and the comparison analysis of the convergence running time of different optimization algorithms is shown in Fig. 3. Based on the data in the figure, WOA, PSO and GA achieve 100% accuracy in the coal mill outlet pressure data, which is significantly better than ACO and BO, indicating that WOA, PSO and GA have obvious advantages in global search and avoiding local optimization. Although the accuracies of ACO and BO are close to the 100% optimal solution, they fail to reach the full optimization, probably due to the fact that they are trapped in the local optimal solution and the global optimization ability is weak, and the same is true for the anomalous mode data. The training time of ACO algorithm is the shortest among the three algorithms, which indicates that ACO algorithm has a certain advantage in convergence speed. WOA algorithm can find the optimal solution, but the training time is obviously longer than the other two optimization algorithms, which is not enough in terms of efficiency. The training time of PSO is in between that of ACO and WOA, which indicates that PSO guarantees to find the optimal solution and also optimizes the performance of the algorithm. find the optimal solution while optimizing the computational efficiency, shortening the training time relative to WOA.GA has the slowest running time among the five algorithms, and although the accuracy rate reaches 100%, its longer running time may limit the application in time-sensitive scenarios.BO has a running time of 39,118s, and while the accuracy rate reaches a better rate, the time performance is better than that of WOA, but slightly worse than that of PSO. In summary, the PSO optimization algorithm outperforms in terms of accuracy, indicating that it has advantages in global search and avoiding local optima. In terms of convergence time, PSO is significantly better than WOA, GA, and substantially closer to ACO, which indicates that PSO has excellent convergence speed while maintaining high accuracy, ensuring the optimization effect of particle swarm optimization algorithm on 3D residual neural networks.

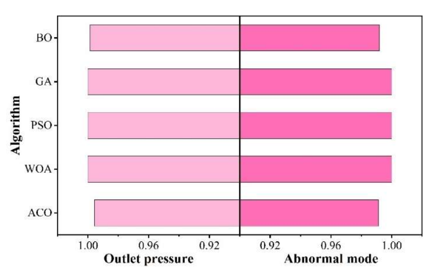


Figure 2: The accuracy comparison results of the training process

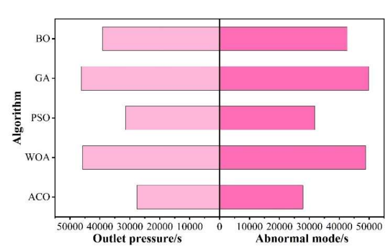


Figure 3: Comparative analysis of convergence running time

III. B. Modeled Coal Mill Export Pressure Forecast Analysis

III. B. 1) Evaluation indicators

The mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) are adopted as the evaluation indexes for comprehensively comparing the prediction accuracy of each model, which are calculated as shown in Eqs. (11), (12), and (13), respectively. Three evaluation index expressions:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left(\left|y_i - z_i\right|\right) \tag{11}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\left|\frac{z_i}{y_i}\right|\right)\times 100\% \tag{12}$$

$$RMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(y_i - z_i\right)^2} \tag{13}$$

where y_i denotes the predicted value, z_i denotes the actual value, and n denotes the number of times.

III. B. 2) Comparative analysis results

In order to verify the effectiveness of the coal mill outlet pressure prediction model based on particle swarm algorithm and three-dimensional residual neural network, this paper's model is compared and analyzed with ResNet, BPNN, KNN, LSTM, SVM in the prediction problem of coal mill outlet pressure, and the results of comparison of three kinds of evaluation indexes of the prediction model are shown in Fig. 4~Fig. 6. Comprehensive Fig. 4~Fig. 6 shows that compared with ResNet (MAE: 23.957, MAPE: 0.0541, RMSE: 23.693), BPNN (MAE: 25.654, MAPE: 0.0506, RMSE: 21.199), KNN (MAE: 28.317, MAPE: 0.0517, RMSE: 21.983), LSTM (MAE: 29.034, MAPE: 0.0521, RMSE: 21.265), SVM (MAE: 24.66, MAPE: 0.0518, RMSE: 23.565), and the prediction performance of the coal mill outlet pressure prediction model based on the particle swarm algorithm with the three-dimensional residual neural network was significantly better (MAE: 21.473, MAPE: 0.0333, RMSE: 20.069), it can be said that this paper obviously almost completely predicted the trend of coal mill outlet pressure changes, accurately grasped the inner law of coal mill outlet pressure changes, and confirmed the effectiveness of the particle swarm algorithm and the three-dimensional residual neural network in the prediction of coal mill outlet pressure.

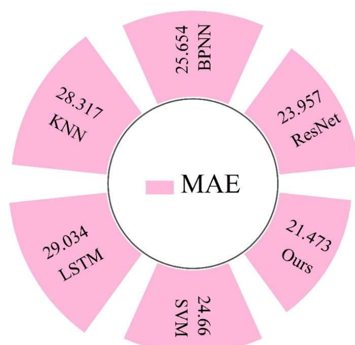


Figure 4: Comparative analysis of MAE results

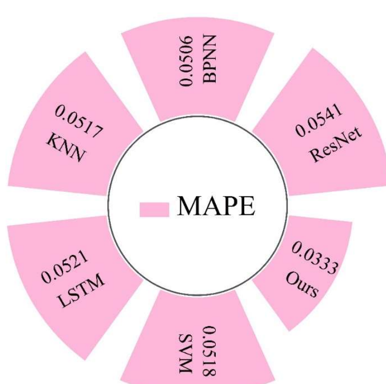


Figure 5: Comparative analysis of MAPE results

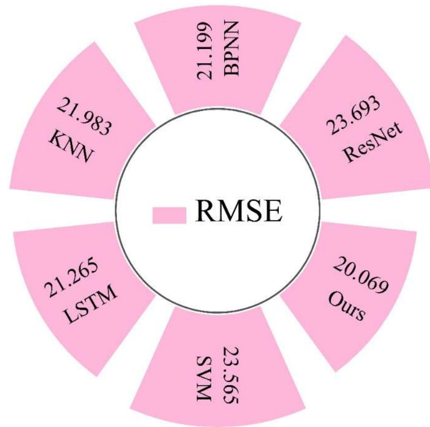


Figure 6: Comparative analysis of RMSE results

III. C. Anomalous pattern recognition analysis of the model

III. C. 1) Identifying Confusion Matrices

The experiment adopts the anomaly pattern data that has been preprocessed in the previous section, and the sample size of the data is 1,000, and the data set is divided into training set and test set according to 7:3, and 700 groups of data are randomly selected as the training set, and the remaining 300 groups of data are used as the test set for evaluating the effect of model recognition. The particle swarm algorithm optimized residual network (PSO-ResNet) is used for anomaly pattern recognition, and the confusion matrix of the recognition results is shown in Fig. 7, where C1~C5 denote current anomaly, temperature anomaly, pressure anomaly, flow anomaly, and rotational speed anomaly, respectively. PSO-ResNet has the lowest current anomaly (C1) identification accuracy of nearly 94.7%, presumably due to the small number of samples in the current anomaly (C1) class. For temperature anomaly (C2), PSO-ResNet has 97.5% recognition accuracy and performs well. Pressure anomaly (C3: 97.7%), flow rate anomaly (C4: 96.1%), and rotational speed anomaly (C5: 98.5%) have higher recognition accuracy because there are more data samples, and all of them have nearly 95% correct classification rate. It can be seen that the PSO-ResNet model has a better performance in coal mill anomaly patterns and can correctly classify most of the anomaly patterns.

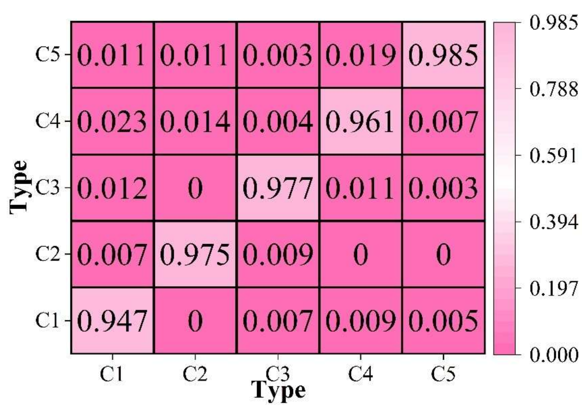
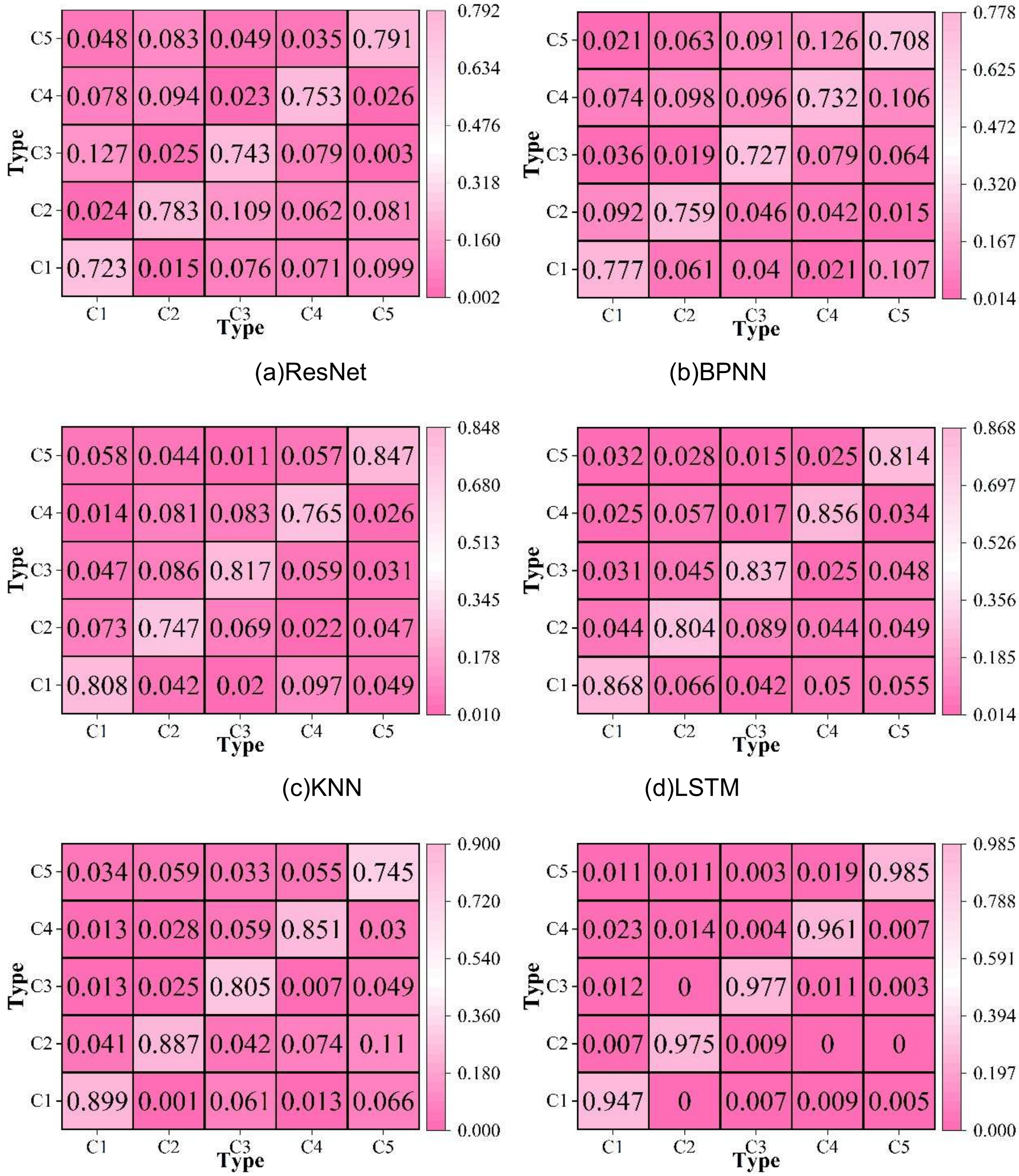


Figure 7: Identify the result confusion matrix

III. C. 2) Comparative test analysis

In order to further verify the effectiveness and recognition effect of particle swarm optimization 3D residual neural network on coal mill anomaly patterns, the recognition accuracy of different models is explored. A comparison of the recognition accuracy of different models is shown in Fig. 8, where (a)~(f) are ResNet, BPNN, KNN, LSTM, SVM, PSO-ResNet, respectively. based on the data performance in the figure, it can be seen that, compared with ResNet, BPNN, KNN, LSTM, SVM model, PSO-ResNet model is more preferred in the recognition of abnormal patterns of coal mill, and can meet the user's recognition needs. The PSO-ResNet model has a higher priority in coal mill anomaly pattern recognition, can meet the user's recognition needs, and has a certain reference value for the development of industrial intelligence.



(a)ResNet (b)BPNN

(c)KNN (d)LSTM

(e)SVM (f)PSO-ResNet

Figure 8: Comparison of recognition accuracy rates of different models

IV. Conclusion

Particle swarm optimization three-dimensional residual neural network shows excellent performance in coal mill outlet pressure prediction and anomaly pattern recognition. In the algorithm performance comparison, the PSO optimization algorithm not only maintains the same 100% accuracy as WOA and GA, but also significantly outperforms the latter two in terms of convergence speed, showing strong global search and local optimization capabilities. In the field of coal mill outlet pressure prediction, the model achieves excellent results with MAE value of 21.473, MAPE value of 0.0333, and RMSE value of 20.069, which are 10.4%, 38.4%, and 15.3% lower than that of the traditional ResNet model, respectively, which proves the model's ability of accurately grasping the trend of the change of the pressure at the outlet of the coal mill. In the field of anomaly pattern recognition, the model's recognition accuracy of five typical anomalous states exceeds 94%, of which the recognition rate of rotational speed anomalies is as high as 98.5%, which is significantly better than the traditional models such as BPNN, KNN, and LSTM. The experimental results show that the soft thresholding sub-network effectively filters out the data noise, the distraction mechanism enhances the recognition ability of local abnormal features, and the particle swarm optimization solves the problem of neural network parameter tuning. This model provides reliable technical support for the condition monitoring of coal mill system, which is of great practical value for improving energy utilization efficiency and reducing equipment failure downtime, and at the same time provides new research ideas in the field of intelligent diagnosis and prediction of industrial processes.

References

[1] Hu, Y., Ping, B., Zeng, D., Niu, Y., & Gao, Y. (2020). Modeling of coal mill system used for fault simulation. Energies, 13(7), 1784.

[2] Huang, C., Qu, S., Ke, Z., & Zheng, W. (2024). Dual fault warning method for coal mill based on Autoformer WaveBound. Reliability Engineering & System Safety, 245, 110030.

[3] Hong, X., Xu, Z., & Zhang, Z. (2019). Abnormal condition monitoring and diagnosis for coal mills based on support vector regression. IEEE Access, 7, 170488-170499.

[4] Jian, Y., Qing, X., Zhao, Y., He, L., & Qi, X. (2020). Application of Model‐Based Deep Learning Algorithm in Fault Diagnosis of Coal Mills. Mathematical Problems in Engineering, 2020(1), 3753274.

[5] Grzegorzek, W., Adamecki, D., Głuszek, G., Lutyński, A., & Kowol, D. (2021). Technique to Investigate Pulverizing and Abrasive Performance of Coals in Mineral Processing Systems. Energies, 14(21), 7300.

[6] Feng, L., Yang, F., Zhang, W., & Tian, H. (2019). Model predictive control of duplex inlet and outlet ball mill system based on parameter adaptive particle swarm optimization. Mathematical Problems in Engineering, 2019(1), 6812754.

[7] Eslick, J. C., Zamarripa, M. A., Ma, J., Wang, M., Bhattacharya, I., Rychener, B., ... & Miller, D. C. (2022). Predictive modeling of a subcritical pulverized-coal power plant for optimization: Parameter estimation, validation, and application. Applied Energy, 319, 119226.

[8] Zhu, P., Qian, H., & Chai, T. (2020). Research on early fault warning system of coal mills based on the combination of thermodynamics and data mining. Transactions of the Institute of Measurement and Control, 42(1), 55-68.

[9] Basu, S., & Cherian, S. (2021). Dynamic modelling and simulation of coal pulverizer. International Journal of Mechanical Engineering and Applications, 9(4), 58-71.

[10] Fan, W., Ren, S., Zhu, Q., Jia, Z., Bai, D., & Si, F. (2021). A novel multi-mode Bayesian method for the process monitoring and fault diagnosis of coal mills. IEEE Access, 9, 22914-22926.

[11] Saluja, G., Mallick, S. S., & Karmakar, S. (2024). Predicting pneumatic conveyability and flowability of fly ash using bulk property characterization. Particulate Science and Technology, 42(3), 482-494.

[12] Sharma, J., Malviya, G., & Shah, V. (2025). Controllability Analysis of Coal Mill Pulverizer Model. fuel, 32(2s).

[13] Gao, Y., Zeng, D., Liu, J., & Jian, Y. (2017). Optimization control of a pulverizing system on the basis of the estimation of the outlet coal powder flow of a coal mill. Control Engineering Practice, 63, 69-80.

[14] Li, X., Wang, H., & Wang, S. (2023, March). Study on optimization and adjustment of high sodium coal pulverization system for 660MW boiler. In International Conference on Mechatronics Engineering and Artificial Intelligence (MEAI 2022) (Vol. 12596, pp. 333-339). SPIE.

[15] Su Pan, Xingyang Nie, Xiaoyu Zhai, Cheng He & Zhenping Ding. (2025). Classification of power quality disturbances using residual networks with channel attention mechanism. Engineering Applications of Artificial Intelligence,151,110641-110641.

[16] Ji Zhongping, Zhou Chengqin, Zhang Qiankan, Zhang YuWei & Wang Wenping. (2020). A Deep Residual Network for Geometric Decontouring. Computer Graphics Forum,39(7),27-41.

[17] Song Lepeng, Liang Qin, Chen Hui, Hu Hao, Luo Yu & Luo Yanling. (2022). A New Approach to Optimize SVM for Insulator State Identification Based on Improved PSO Algorithm. Sensors,23(1),272-272.

[18] Ben Regaya Chiheb, Hamdi Hichem, Farhani Fethi, Marai Afef, Zaafouri Abderrahmen & Chaari Abdelkader. (2023). Real-time implementation of a novel MPPT control based on the improved PSO algorithm using an adaptive factor selection strategy for photovoltaic systems. ISA transactions,146,496-510.