Intelligence & Robotics

Research Article

Degradation trend prediction of rail stripping for heavy haul railway based on multi-strategy hybrid improved pelican algorithm

Changfan Zhang[1], Chang Jiang[1], Jianhua Liu[1], Weifeng Yang[2], Jia He[3]

[1]College of Railway Transportation, Hunan University of Technology, Zhuzhou 412007, Hunan, China.
[2]Zhuzhou CRRC Times Electric Co., Ltd, Zhuzhou 412007, Hunan, China.
[3]CHN Energy Bashan Railway Co., Ltd, Baotou 014010, Inner Mongolia, China.

Correspondence to: Prof. Jianhua Liu, College of Railway Transportation, Hunan University of Technology, No 88, Taishan West Road, Taiyuan District, Zhuzhou 412007, Hunan, China. E-mail: jhliu@hut.edu.cn

Abstract

As a key component of the heavy-haul railway system, the rail is prone to damages caused by harsh operating conditions. To secure a safe operation, it is of great essence to detect the damage status of the rail. However, current damage detection methods are mainly manual, so problems such as strong subjectivity, lag in providing results, and difficulty in quantifying the degree of damage are easily generated. Therefore, a new prediction method based on the improved pelican algorithm and channel attention mechanism is proposed to evaluate the stripping of heavy-haul railway rails. By processing the rail vibration acceleration, it predicts the stripping damage degree. Specifically, a comprehensive health index measuring the degree of rail stripping is first established by principal component analysis and correlation analysis to avoid the one-sidedness of a single evaluation index. Then, the convolutional bidirectional gated recursive network is trained and generalized, and the pelican algorithm, improved by multiple hybrid strategies, is used to optimize the hyperparameters in the network so as to find the optimal solution by constantly adjusting the search strategy. The squeeze-excitation channel attention module is then incorporated to re-calibrate the weights of valid features and to improve the accuracy of the model. Finally, the proposed method is tested on a specific rail stripping dataset and a public dataset of PHM2012 bearings, and the generalization and effectiveness performance of the proposed method is proved.

Keywords: Evolution trend of rail stripping, heavy-haul railways, improved pelican algorithm, squeeze-excitation channel attention

1. INTRODUCTION

For heavy haul railways, the steel rails are subjected to the strong impact from the heavy load through the long-term wheel-rail rolling contact, for which common defects affecting train operation will be caused, such as initial shelling, stripping, and wear, with stripping being particularly concerning. Not only does stripping strengthen the impact generated during train operation, but it also causes the damages to deteriorate. Untimely maintenance due to rapidly deteriorating defects can lead to serious failures, such as broken rails, significantly disrupting the train operation.

In recent years, scholars have carried out extensive research on the mechanisms and development processes of rail stripping. The research methods are mainly divided into two categories: mechanism modeling analysis and data-driven prediction. Mechanism modeling analysis is to study the development principle of rail stripping based on the analysis of train-rail dynamics theory, wheel-rail contact calculation theory, contact mechanics theory, and Archard law of material wear. Additionally, it involves using dynamics simulation software UM or SIMPACK to construct a heavy-haul railway train-rail coupling model and design the calculation model for predicting rail stripping development. Brunel *et al.* studied the stress-strain response of different steel grades of rails in contact with wheels by using the finite element method[1]. Pavlík *et al.* analyzed and predicted the wear of curved tracks caused by wheel-rail contact through simulations with SIMPACK software[2]. Madge *et al.* discussed the fatigue crack initiation, and wear processes were analyzed by using the Archard wear model and the critical plane method prediction model of crack initiation[3]. Liu *et al.* adopted the ANSYS finite element model of wheel-rail sliding contact to study the influence of different stripping damages and relative sliding speeds between the wheel and rail on the rail stripping area[4]. Zan *et al.* utilized the finite element method and Paris formula to calculate the crack growth rate so as to study the relationship between the fatigue crack growth and the stripping development on the rail surfaces[5]. Although the above scholars have achieved certain results in the study of rail stripping, their calculation processes based on mechanisms involve many factors, leading to complex models, large amounts of calculation, and long calculation time. However, if those models are simplified, the corresponding prediction accuracy will be reduced.

Data-driven prediction is often based on the regular characteristics of rail damage data obtained from laboratory tests or on-site measurements. It establishes rail damage prediction models by traditional machine learning, deep neural networks, and so on. By predicting rail damages, this approach is capable of indirectly considering various factors, including the operating environment factors. Ma *et al.* analyzed the development characteristics of rail side wear of curve tracks with different radii, established a prediction model for the development of side wear for curve tracks based on a nonlinear autoregressive neural network, and proposed a rail replacement strategy for curve tracks suitable for Shuozhou-Huanghua heavy-haul railway section[6]. Wang *et al.* and Li *et al.* applied Monte Carlo statistical methods in their respective studies to predict the service life of wheels based on a large number of measured wheel wear data of Guangzhou Metro in the past ten years[7,8]. Andrade and Stow determined the optimal wheel parameter values for the maintenance of 250,000 kilometers of no-turning-rolling by using linear statistical analysis and predicting railway wheel data[9]. On this basis, Han and Zhang predicted the wheel wear on the track with polynomial fitting[10]. Despite the advancements in rail damage prediction, there remain certain limitations within the current landscape. The rail damage prediction model for curved tracks developed by Ma *et al.* is restricted in its applicability to curves with varying radii, potentially lacking generalizability. Similarly, the wheel lifespan prediction models of Wang and Yang rely on a decade's worth of wheel wear data from Guangzhou Metro. However, this dataset might not adequately cover variations across different regions and conditions, potentially affecting the generalization of their predictive models. Andrade and Stow's selection of optimal parameter values for maintaining 250,000 kilometers of non-turning rolling through linear statistical analysis relies on certain assumptions and constraints. These assumptions might not hold in varied scenarios, compromising the effectiveness of the proposed maintenance. Han and Zhang's[10] use of polynomial fitting to predict wheel wear on tracks might encounter limitations in complex scenarios. This trade-off between prediction accuracy and model complexity

could affect the precision of predictions. In light of the limitations observed in the prior studies, there arises a clear need for a research methodology that not only overcomes these shortcomings but also introduces a more robust and adaptable approach to rail damage prediction.

With the development of deep learning technology, new models are constantly emerging. As classic deep learning models designed specifically for image and sequence data, Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) have been widely used in target detection, health status monitoring, fault diagnosis, computer vision, and many other fields. Ling *et al.* proposed a new method to extract multidirectional spatiotemporal features of SCADA data for wind turbine (WT) condition monitoring based on CNN and bidirectional gated recurrent unit (BIGRU) with attention mechanisms[11]. This model can effectively detect early WT faults and provide more information on fault severities. Applications of big data and artificial intelligence techniques in bridge Structural health monitoring were respectively reviewed by Sun *et al.*, offering meaningful perspectives and suggestions for employing these techniques in the field[12]. Cheng *et al.* established a fault diagnosis model. Combining WCNN with BIGRU solved the problems such as poor diagnosis effects of simple models and complex structures of hybrid models in the fault diagnosis of rolling bearing vibration signals[13]. Ye *et al.* proposed a CV method to inspect the hunting motion of high-speed trains by processing the dynamic wheel-rail contact video, i.e., the VPT method. This approach can directly characterize the dynamic wheel-rail interaction and can robustly identify the hunting motion of high-speed trains[14]. Most applications of CNN and LSTM often show better performance than traditional machine learning models. At the same time, the attention mechanism has received intense attention in recent years. Its integration enhances the ability of the model in terms of extracting important features and further improves the prediction accuracy. These findings all show that with the help of deep learning technology, it could be possible to link the research on the development law of typical rail damages with the vibration acceleration.

Therefore, in this paper, the Pelican algorithm, CNN-BIGRU network, and squeeze-excitation (SE) attention mechanism are integrated to predict the stripping development of heavy-haul railway rails. A new method of predicting the development trend of heavy-haul railway rail stripping is proposed with a new comprehensive health indicator and an improved Pelican algorithm. The main contributions of this method are as follows:

1. A new health indicator for rail stripping damage is established. It combines correlation analysis and principal component analysis (PCA) and thus empowers the model to quantitatively characterize the rail stripping;

2. A CNN-BIGRU rail stripping development trend prediction model is constructed. A SE feature compression excitation module is organically embedded in it to obtain more effective feature information. The deep neural network is trained by using a segmented step decay learning rate instead of a fixed learning rate. This is conducive to the convergence of the algorithm and makes it easier to obtain the optimal solution;

3. Based on the structural characteristics of the network model, the standard Pelican algorithm is improved by introducing an initialization through the chaos sequence of the circle map, dynamic weight factors, and an adaptive T distribution. The optimization efficiency of the algorithm is improved.

The rest of this article is organized as follows. In the"METHODS" section, the stripping evolution trend prediction network of rail and the strategy of the improvement pelican algorithm are introduced. Subsequently, the "EXPERIMENT ANALYSIS" section describes the dataset used for the experiments and discusses the experimental results. Finally, the paper concludes with the "CONCLUSIONS" section.
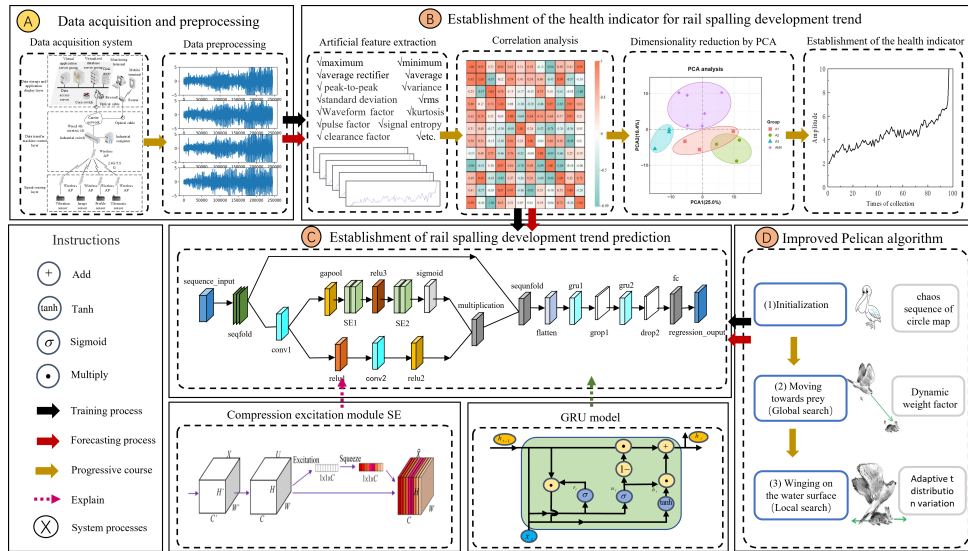
Figure 1. Prediction block diagram of rail stripping development trend.

2. METHODS

This section will elaborate on the proposed prediction process of rail stripping development trend for heavy haul railways. Figure 1 gives the framework diagram of the proposed method, which mainly includes four parts: the acquisition and preprocessing of data, the establishment of rail stripping development health indicators based on correlation analysis and PCA, the design of the prediction model rail stripping development trend based on the improved convolutional two-way gated cycle network, and the multi-strategy hybrid improvement of Pelican algorithm. The details of each part are as follows:

2.1. Establishment of the health indicator

From the quantitative point of view, the development process of rail stripping damage often manifests that the performance parameters of the rail gradually deviate from the normal range, and there will be characteristic components in the vibration signal that represent the development information. However, due to the complexity and variability of the operating environment of heavy haul railways, when only relying on single-domain characteristic indicators such as RMS, it is extremely difficult to effectively characterize the interaction between multiple influencing factors in the development of rail stripping, and it is also very hard to meet the actual engineering application requirements.

In order to obtain more comprehensive information on rail stripping development, it is necessary to comprehensively use multi-domain features to characterize its development process. Firstly, the time domain, frequency domain, and time-frequency domain characteristics of rail vibration signals (such as the maximum value, minimum value, root mean square, pulse factor, etc.) are extracted, and then the correlation analysis and PCA are performed to screen and reduce the feature dimension of the high-dimensional feature set so as to construct a comprehensive health indicator that can fully reflect the development process of rail stripping.

2.2. Establishment of the prediction model

Aiming to eliminate the problems of low prediction accuracy of a single model (CNN, gated recurrent unit (GRU)) and long running time of a combined prediction model (CNN-LSTM), a prediction model for rail stripping development trend based on CNN-BIGRU is constructed in combination with the respective advantages of CNN and GRU, into which the spatial correlation and time dependence of rail vibration sequences are introduced. In addition, considering that each segment of the rail vibration signal sequence has a different contribution to the final prediction, the SE channel attention mechanism is adopted to lock the information
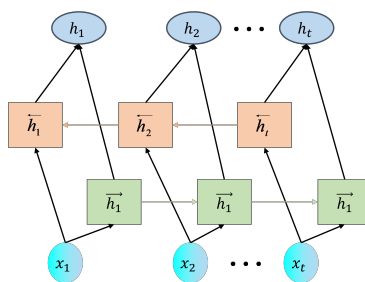
Figure 2. Diagram of BIGRU structure.

segments about different statuses in the rail data, to screen more useful feature information, and to enhance the prediction ability of the model.

2.2.1. CNN-GRU module

The CNN-GRU module consists of a CNN and a GRU part. The main structure of the basic internal neural network layer of the CNN layer is convolution - pooling - full connection. This kind of structure can reduce the number of weights, extract features of specific objects, and then identify, classify, or predict the corresponding objects according to the features.

The output of the CNN part is used as the input of the GRU part, and the GRU part performs the prediction. GRU is an effective recurrent neural network, which has been improved on the basis of LSTM so that it can better learn the time sequence characteristics of data and suppress gradient learning and gradient explosive. As shown in Figure 2, the BIGRU consists of two layers of GRUs stacked on top of each other, and the final output is jointly determined by the statuses of the two GRUs.

The final output state is the combination of the forward propagation output and the back propagation output. The calculation is carried out by:

$$\overrightarrow{h_t} = f\left(\overrightarrow{U} h_{t-1} + \overrightarrow{W} x_t + \overrightarrow{b}\right) \tag{1}$$

$$\overleftarrow{h_t} = f\left(\overleftarrow{U} h_{t-1} + \overleftarrow{W} x_t + \overleftarrow{b}\right) \tag{2}$$

$$h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}] \tag{3}$$

Where, in the forward propagation process, x_t, $\overrightarrow{U}$, $\overrightarrow{W}$, and $\overrightarrow{b}$ represent the input at time, the state, the input weight matrix, and the bias item of the hidden layer, respectively. While $\overleftarrow{U}$, $\overleftarrow{W}$, and $\overleftarrow{b}$ are the state, the input weight matrix, and the bias term of the sub-hidden layer, respectively. By using the two-layer GRU structure, the problem of gradient disappearance and gradient explosion in the LSTM network can be effectively solved, and the accuracy of prediction can be improved.

2.2.2. Compression excitation module SE

During the prediction process of stripping development, there is a variety of noise interferences. As the network is trained, these noises will occupy a large amount of memory, and negative impacts could be imposed on the network. By embedding a SE module[15], the feature weights can be adjusted adaptively to solve the feature loss problem caused by varying channel proportions in the convolution pooling process. This mainly includes Squeeze and Excitation, as shown in Figure 3.
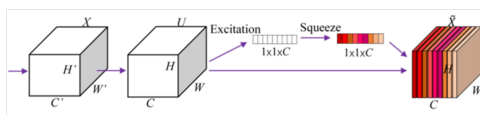
Figure 3. Diagram of SE structure.

The Squeeze part in Figure 3 is expressed as:

$$Z = S(T) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} T_C(i, j) \tag{4}$$

The Excitation part is:

$$E(Z, W) = \sigma(W_2 \delta(W_1 Z)) \tag{5}$$

Where the parameter T indicates the feature representation, H, W, and C represent the height, width, and channel of T, respectively. δ and σ are the ReLU activation function and the Sigmoid activation function, respectively, and W_1 and W_2 are fully connected layers.

2.3. Multi-strategy hybrid improvement of Pelican algorithm

The Pelican Optimization Algorithm (POA)[16] is an intelligent optimization method that simulates the natural behavior of pelicans during hunting. It was first proposed in 2022, and its optimization process mainly consists of three stages: initialization, approaching (global search), and surface flying (local search).

However, as the number of iterations increases, the population diversity of the traditional POA decreases, and it is easy to fall into a local optimum. Therefore, an Improved POA (IPOA) implemented with multiple strategies is proposed. The specific improvements are as follows.

2.3.1. Initialization through the chaos sequence of circle map
At the initialization stage, the chaos sequence of the circle map[17] is used instead of initialization through a rand function. The initialization definition formula through a circle map is:

$$x_{i+1} = mod\left(x_i + 0.2 - (\frac{0.5}{2\pi}) \sin(2\pi x_i), 1\right) \tag{6}$$

Compared to a random distribution, the initial position distribution of the population is more uniform, and the search becomes more convenient. The search range of pelican groups in space is larger. These factors can improve the global search performance of the algorithm to a certain extent.

2.3.2. Dynamic weight factor
At the approaching stage (global search phase), a dynamic weight actor λ[18] is introduced, which is defined as:

$$\lambda = \frac{e^{2\left(1-\frac{t}{T}\right)} - e^{-2\left(1-\frac{t}{T}\right)}}{e^{2\left(1-\frac{t}{T}\right)} + e^{-2\left(1-\frac{t}{T}\right)}} \tag{7}$$

Where t indicates the current iteration number, and T is the maximum number of iterations. The updated equation of the pelican position becomes:

$$x_{i,j}^{p1} = \begin{cases} x_{i,j} + \text{rand} \cdot (p_j - I \cdot x_{i,j}) \cdot \lambda, F_p < F \\ x_{i,j} + \text{rand} \cdot (x_{i,j} - p_j) \cdot \lambda, \quad \text{else} \end{cases} \tag{8}$$

The dynamic weighting factor helps the pelicans better update their position. At the beginning of the iteration, it is larger, which facilitates a better global search. Towards the end of the iteration, it will be adaptively reduced to improve local search performance while increasing the convergence speed.

Table 1. Algorithm. Pseudo-code of IPOA

Start IPOA	
1	Input the optimization problem information.
2	Determine the IPOA population size (N) and the number of iterations (T).
3	Initialization of the position of pelicans and calculate the objective function.
4	For t = 1: T
5	Generate the position of the prey at random.
6	For i = 1: N
7	Phase 1: Moving towards prey (exploration phase).
8	For j = 1:m
9	Calculate new status of the jth dimension.
10	End.
11	Update the ith population member.
12	Phase 2: Winging on the water surface (exploitation phase).
13	For j = 1:m
14	Calculate new status of the jth dimension.
15	End.
16	Update the ith population member.
17	End.
18	Update the best candidate solution.
19	End.
20	Output the best candidate solution obtained by IPOA.
End IPOA	

2.3.3. Adaptive T distribution mutation

Adaptive T distribution mutation with the iteration number as a degree of freedom [19] is used as the operator disturbing the position of the solution, and the updated position equation is represented as follows:

$$x_i^t = x_i + x_i \cdot t(\text{iter}) \tag{9}$$

The introduction of T distribution mutation not only makes full use of the current position information but also increases the random interference information. This can further improve the ability of the algorithm to avoid local optimal solutions and speed up the convergence of the algorithm.

To sum up, the standard Pelican algorithm is improved by introducing the chaotic sequence of the circle map into the initialization, along with the dynamic weight factor and the adaptive T distribution mutation.

The improved algorithm flow chart is shown in Figure 4 and Table 1, where N and T represent the population of pelicans and the maximum number of iterations, respectively; $x_{i,j}^{p_1}, x_{i,j}^{p_2}$, and x_i represent the new status of the ith pelican in the jth dimension based on phase 1, the new status of the ith pelican in the jth dimension based on phase 2, and the new position of the pelican after the update, respectively. After all individuals are updated through three stages, namely initialization, approaching, and surface flying, the best candidate solution so far will be obtained, and the algorithm will enter the next iteration. This process will be executed repeatedly based on the position update formula until it is completely over. The obtained optimal parameters will then be outputted to the prediction network for the next step of prediction.

3. EXPERIMENT ANALYSIS

The proposed method first performs stripping evolution prediction experiments on railway rail damage vibration data. However, it is noteworthy that the dataset for rail damage vibration employed in this paper is unique, and there exists no other openly available dataset for rail vibration that could be utilized for comparative experiments with alternative algorithmic models in damage evolution and life prediction. Consequently, to assess the generalization performance of the proposed method, comprehensive generalization experiments were conducted on the publicly accessible PHM2012 rolling bearing vibration dataset.
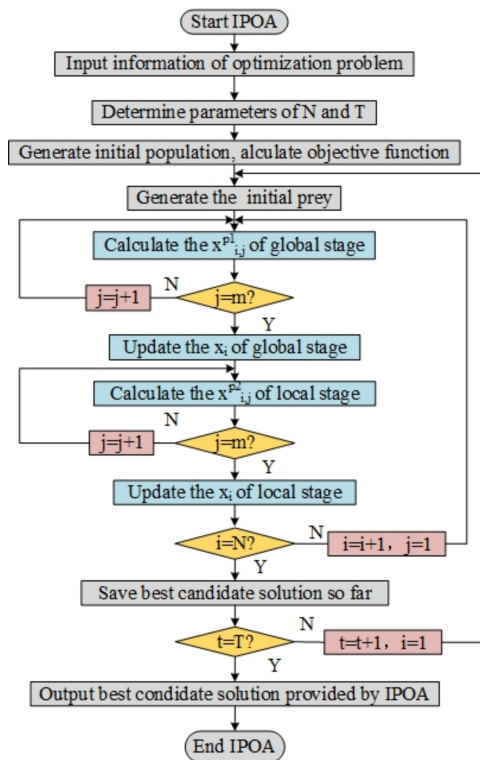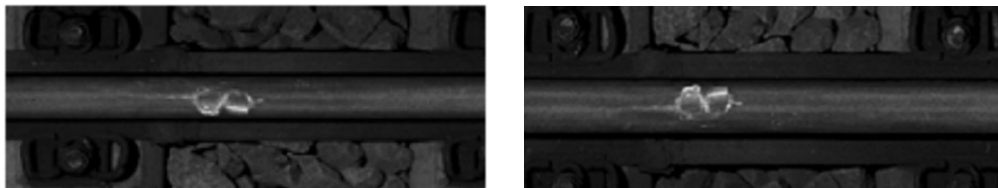
Figure 4. Flow diagram of IPOA.

Figure 5. We present the Site map of rail stripping in Figure 5A-B. A: Site map of the rail stripping 1; B: Site map of the rail stripping 2.

3.1. Experiment and analysis of rail stripping

The data of this experiment comes from a certain domestic railway section of the experimental line (as shown in Figure 5), and there are two working conditions: stripping 1 and stripping 2.

3.1.1. Preprocessing of experimental data samples

A continuous test spanning several months was conducted on the rail stripping, covering the entire process from the initiation of rail stripping through its development and up to the point of rail replacement. Vibration data from the damaged rail was collected at regular intervals during the same sampling period, amounting to a total of 98 instances. The sampling frequency was set at 4 kHz, with data collection occurring at 2500 sample points each time. The data is depicted in Figure 6, and the corresponding waveforms are sequentially assembled from the vibration sequences that reflect the damage characteristics in a single sampling.

In this experiment, two sets of stripping data were utilized, each comprising 98 x 2500 sample points of stripping. Initially, a total of 26 time-frequency domain characteristics were extracted for each collected sample, including the maximum value, kurtosis, peak factor, and others. At this point, the data dimension stood at
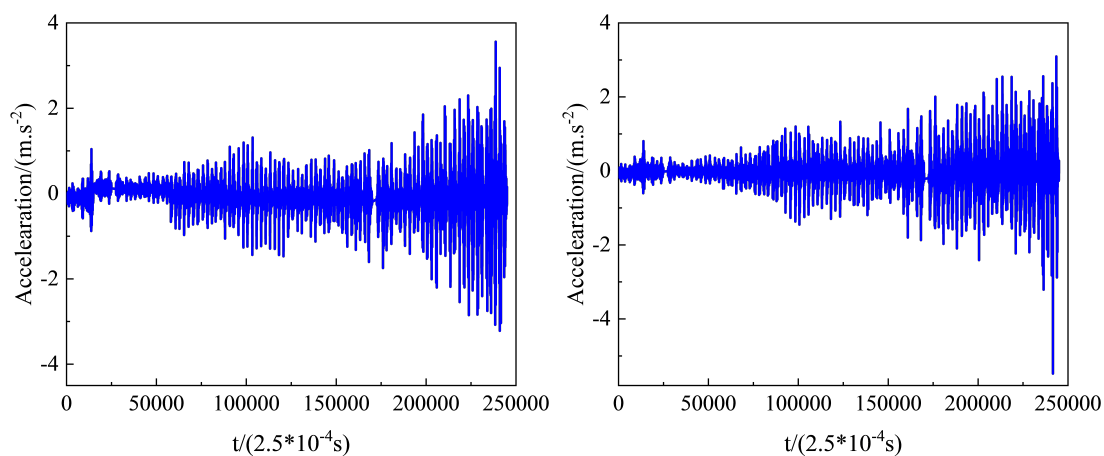
Figure 6. We present Vibration data of rail with stripping in Figure 6A-B. A: Vibration data of rail with stripping 1; B: Vibration data of rail with stripping 2.

Table 2. Date of eigenvalue and variance contribution rate

Principal component	Eigenvalue	Contribution rate	Cumulative contribution rate
1	9.802	81.686%	81.686%
2	1.586	13.219%	94.906%
3	0.359	2.990%	97.895%
4	0.124	1.033%	98.928%
5	0.063	0.527%	99.455%
6	0.048	0.399%	99.853%
7	0.010	0.081%	99.935%
8	0.006	0.049%	99.983%
9	0.001	0.011%	99.995%
10	0.000	0.004%	99.996%
11	0.000	0.001%	99.999%
12	0.000	0.000%	100.000%

98×26. Subsequently, indicators of poor performance were removed based on Spearman's correlation, resulting in 12 remaining features. The data dimension was then reduced to 98×12. Following this, the dimensionality of these 12 features was further reduced through PCA, yielding an eigenvector with a dimension of 98×1. This eigenvector serves as the health indicator capable of characterizing the development trend of rail stripping damage.

The resulting eigenvalues, contribution rates, and accumulative contribution rates corresponding to the remaining 12 principal components after correlation analysis and feature screening are listed in Table 2.

The principal components with characteristic values greater than 0.3 were selected; that is, the top three principal components were used as the final principal components, and the accumulative contribution rate reached 97.895% (far greater than 85%). Their corresponding 12×3 eigenvectors were multiplied by the standardized data to obtain the calculation formula of the top three principal components, as shown in Table 3.

3.1.2. Algorithm validity verification
To validate the accuracy and efficacy of the proposed IPOA, four distinct algorithms, Particle Swarm Optimization (PSO) Algorithm, Whale Optimization Algorithm (WOA), POA, and IPOA, were individually chosen to conduct comparative calculations across eight functions selected from a pool of 23 benchmark functions. These functions encompassed three unimodal benchmark functions, three multimodal benchmark functions, and two mixed benchmark functions, with the definitions outlined in Table 4. Each algorithm underwent

Table 3. Calculation equations of the model

Name	Interpretation	Calculation equation
F1	Score of the first principal component	$0.097x_1 + 0.102x_2 + 0.093x_3 + 0.098x_4 + 0.098x_5 + 0.078x_6 + 0.098x_7 + 0.081x_8 + 0.087x_9 + 0.083x_{10} + 0.086x_{11} + 0.086x_{12}$
F2	Score of the second principal component	$-0.051x_1 + 0.006x_2 - 0.254x_3 - 0.088x_4 - 0.145x_5 + 0.285x_6 - 0.169x_7 - 0.117x_8 - 0.304x_9 + 0.317x_{10} + 0.331x_{11} + 0.329x_{12}$
F3	Score of the third principal component	$0.6x_1 + 0.027x_2 + 0.125x_3 + 0.393x_4 + 0.008x_5 + 1.189x_6 + 0.055x_7 + 0.426x_8 - 0.194x_9 - 0.72x_{10} - 0.295x_{11} - 0.12x_{12}$
F	Total score	$\left(\frac{0.817}{0.979}\right)F_1 + \left(\frac{0.132}{0.979}\right)F_2 + \left(\frac{0.03}{0.979}\right)F_3$

Table 4. Benchmark functions

Benchmark function	Dim	Scope	Minimum
$f_1(x) = \sum_{i=1}^{n} x_i^2$	30	[-100,100]	0
$f_1(x) = \sum_{i=1}^{n} x_i^2$	30	[-10,10]	0
$f_3(x) = \sum_{i=1}^{n} ix_i^4 + \text{random}[0,1)$	30	[-1.28,1.28]	0
$f_4(x) = \sum_{i=1}^{n} [x_i^2 - 10\cos(2\pi x_i) + 10]$	30	[-5.12,5.12]	0
$f_5(x) = -20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)\right) + 20 + e$	30	[-32,32]	0
$f_6(x) = \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n}\cos(\frac{x_i}{\sqrt{i}}) + 1$	30	[-600,600]	0
$f_7(x) = \sum_{i=1}^{n}\left[a_i - \frac{x_1(b_i^2 + b_1 x_2)}{b_i^2 + b_1 x_3 + x_4}\right]$	4	[-5,5]	0.1484
$f_8(x) = -\sum_{i=1}^{10}\left[(x - a_i)(x - a_i)^T + c_i\right]^{-1}$	4	[0,10]	-1

100 iterations for every benchmark function, and the population size for all algorithms was set at 20. Each algorithm was independently executed ten times to obtain average values, and the experimental results were meticulously recorded. Convergence curves illustrating the performance are depicted in Figure 7.

In Figure 7, the lines represented by black, green, blue, and red signify the PSO Algorithm, WOA, POA, and IPOA, respectively. It is evident that the red line demonstrates the most rapid descent as the number of iterations progresses. This observation indicates that in equivalent circumstances, IPOA achieves the lowest fitness and the most rapid convergence rate. Hence, the decision was made to utilize the IPOA for optimizing the parameters of the prediction model in subsequent analyses.

3.1.3. Network parameter setting
In this experiment, the historical health indicator data of rail stripping was used to predict the health indicator of future momentum. The dimensionality of input data is 12, and 1 for the output data. Through multiple experiment runs, the optimal settings of the main parameters for this model were obtained and are shown in Table 5.

3.1.4. Performance test
In order to better assess the accuracy of the prediction performance of the proposed model, three indicators of MAE, RMSE, and R^2 were used to evaluate it, and their specific expressions are:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{10}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{12}$$
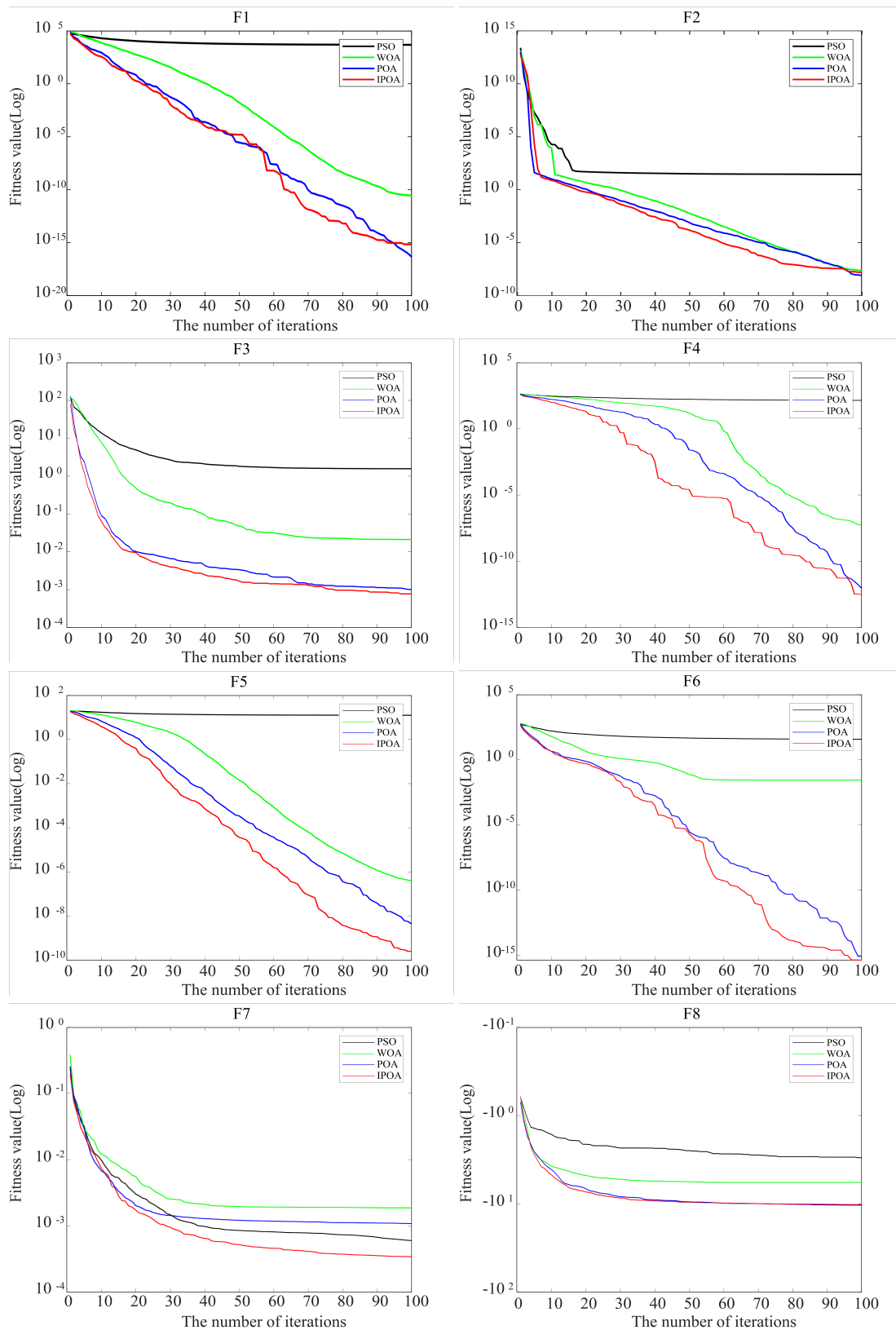
Figure 7. We present Convergence curves of benchmark functions in Figure 7A-H. A: the convergence curves of $f_1(x)$; B: the convergence curves of $f_2(x)$; C:the convergence curves of $f_3(x)$; D: the convergence curves of $f_4(x)$; E: the convergence curves of $f_5(x)$; F: the convergence curves of $f_6(x)$; G: the convergence curves of $f_7(x)$; H: the convergence curves of $f_8(x)$.

The three traditional prediction network models were selected for comparative experiments at the same conditions:(1) BP, (2) LSTM, and (3) GRU. The results are demonstrated in Figure 8, Table 6, and Table 7.

Out of the whole dataset, the first 75% of data points were used as the training set, and the last 25% as the test set. The prediction accuracy of the model was determined by comparing the stripping health indicator value of the test set with the actual health indicator. It can be seen in Figure 8, Table 6, and Table 7 that the prediction accuracy of the proposed method is significantly better than that of the other five traditional models. For the prediction of stripping 1, MAE and RMSE of the proposed model decreased by 32.5% and 56.8%, respectively, and R^2 increased by 34.3%, compared with those of GRU, the model with the best performance indicators out of the five traditional models. For the prediction of stripping 2, MAE and RMSE of the proposed model decreased by 50% and 50.7%, respectively, and R^2 increased by 35.3%, compared with those of GRU, the model with the best performance indicators out of the three traditional models.

3.1.5. Ablation experiment
In order to make the research results more objective, compare with the latest related research work, and evaluate the role and effect of several key modules in the proposed model, the key modules were removed or replaced in turn for ablation research, while other parameters were kept unchanged. The following four different prediction models were designed for comparative analysis.

(1) POA-CNN-SE-BIGRU. While keeping the other structures of the network unchanged, the multi-strategy hybrid IPOA was replaced with the standard POA. This model was mainly used to verify the role of the improved part of the algorithm.

(2) CNN-SE-BIGRU. While keeping the other structures of the network unchanged, the IPOA was removed. This model was used to evaluate the effect of the optimization algorithm.

(3) CNN-BIGRU. While keeping the other structures of the network unchanged, the SE module was removed. This model was used mainly to evaluate the role of the SE attention module.

(4) BIGRU. The regular BIGRU.

The comparisons of results are shown in Figure 9, Table 8, and Table 9.

Clearly, the convergence curves of stripping 1 and stripping 2 are shown in Figure 9: the black line represents the IPOA, and the red line represents the POA. This shows that IPOA has a low fitness value and fast convergence speed.

It can be seen in Table 8 and Table 9 that the predictive ability of the proposed model is better than that of other comparison models. The improvement of prediction results is attributed to the IPOA and the introduction of SE attention mechanisms. Taking the prediction results of stripping 1 as an example, it is shown in the ablation experiment that from BIGRU to CNN-BIGRU, MAE and RMSE decreased by 22% and 15.7%, respectively, and R^2 increased by 9.5%; When SE was introduced, as in CNN-SE-BIGRU, compared with CNN-BIGRU, MAE decreased by 3%, RMSE decreased by 6.7%, and R^2 increased by 2.5%. This means that the introduction of the SE attention mechanism has enhanced the ability of the network to capture effective features and reduce the interference of useless information to a certain degree. After POA was added, as in POA-CNN-SE-BIGRU, compared with CNN-SE-BIGRU, MAE and RMSE decreased by 3% and 17.1%, respectively, and R^2 increased by 7.2%. This means POA is effective in optimizing model parameters. Furthermore, compared with those of POA-CNN-SE-BIGRU, MAE and RMSE of the proposed model decreased by 15.6% and 29.3%, respectively, and R^2 increased by 5.6%. This is achieved from the multi-strategy hybrid improvement of IPOA.
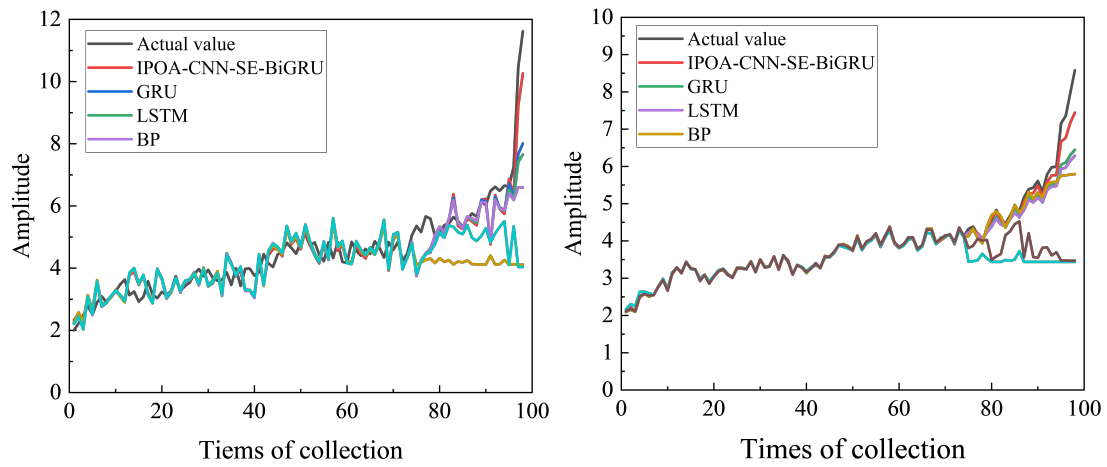
Figure 8. We present Prediction results of stripping development in Figure 8A-B. A: results of stripping 1; B: results of stripping 2.
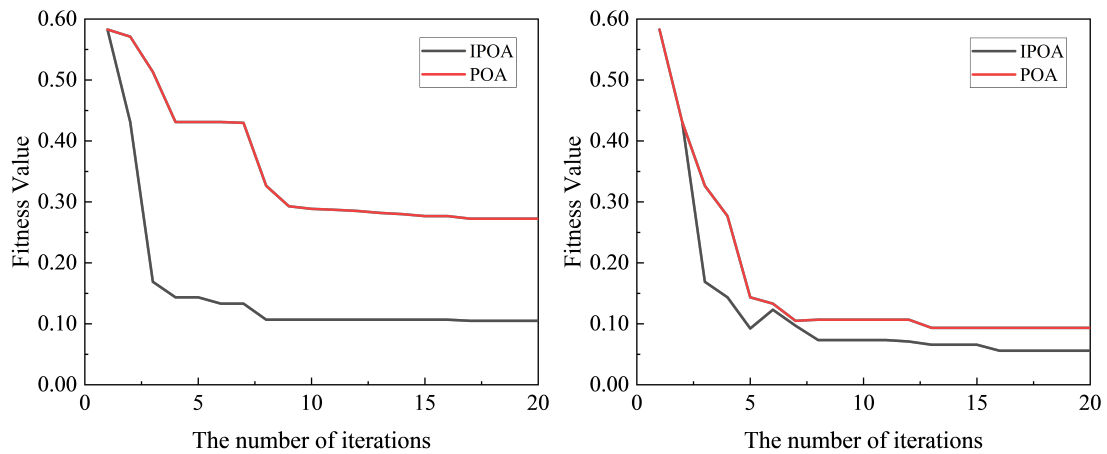

Figure 9. We present Convergence results of IPOA and POA in Figure 13A-B. A: convergence results of IPOA and POA of bearing 1; B: convergence results of IPOA and POA of bearing 2.
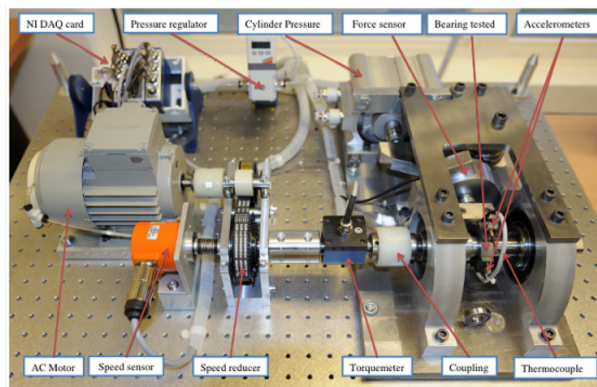

Figure 10. PRONOSTIA Bearing Data Collection Platform.

Table 5. Settings of the main parameters for the proposed model

Parameter	Setting value
Conv1	32 3×1×1 Stride [1 1] Padding [0 0 0 0]
Conv2	64 3×1×32 Stride [1 1] Padding [0 0 0 0]
Drop1	0.3
Drop2	0.3
Max_Epoch	500
Batch_size	128
Optimizers	Adam
Learning Rate Schedule	Piecewise
Learning Rate Drop Factor	0.1
D-Learning Rate Drop Period	200
Search Agents_no	8
Max_iteration	20
Dim	3
Learning Rate	[1e-4,1e-2]
BIGRU_hidden	[10,100]
L2	[1e-4,1e-1]

Table 6. Results of stripping 1

Prediction model	MAE	RMSE	R^2
BP [20]	0.53	1.32	0.42
LSTM [21]	0.42	0.96	0.69
GRU [22]	0.40	0.95	0.70
Proposed Model	0.27	0.41	0.94

Table 7. Results of stripping 2

Prediction model	MAE	RMSE	R^2
BP [20]	0.46	0.86	0.49
LSTM [21]	0.47	0.75	0.61
GRU [22]	0.42	0.69	0.68
Proposed Model	0.21	0.34	0.92

Table 8. Results of stripping 1

Prediction model	MAE	RMSE	R^2
BIGRU [23]	0.41	0.89	0.74
CNN-BIGRU [24]	0.32	0.75	0.81
CNN-SE-BIGRU	0.33	0.70	0.83
POA-CNN-SE-BIGRU	0.32	0.58	0.89
Proposed Model	0.27	0.41	0.94

Table 9. Results of stripping 2

Prediction model	MAE	RMSE	R^2
BIGRU [23]	0.38	0.60	0.75
CNN-BIGRU [24]	0.37	0.56	0.78
CNN-SE-BIGRU	0.32	0.48	0.84
POA-CNN-SE-BIGRU	0.22	0.39	0.89
Proposed Model	0.21	0.34	0.92

3.2. Experiment and analysis of rolling bearing

From the perspective of utilizing the vibration data information caused by rolling bearings throughout their entire life cycle, a mapping model is established between the monitoring vibration data and the degradation degree of the rolling bearings. The method proposed in this chapter was applied as an example on the experimental dataset of PHM 2012 rolling bearing accelerated life. The model structure and parameter settings of the rolling bearing experiment are consistent with section 3.1 of the rail experiment.

Table 10. Results of bearing 1

Prediction model	MAE	RMSE	R^2
BP [20]	0.94	1.35	0.25
LSTM [21]	0.84	1.11	0.49
GRU [22]	0.82	1.06	0.53
Proposed Model	0.44	0.55	0.88

Table 11. Results of bearing 2

Prediction model	MAE	RMSE	R^2
BP [20]	1.31	2.08	0.19
LSTM [21]	1.22	1.77	0.42
GRU [22]	1.15	1.65	0.49
Proposed Model	0.78	1.08	0.79

3.2.1. Preprocessing of experimental data samples
The data was collected on the accelerated aging platform PRONOSTIA, as shown in Figure 10. The goal of PRONOTIA is to provide real experimental data to describe the degradation of ball bearings throughout their entire service life (until complete failure). The PRONOSITIA platform provides almost all types of defects (balls, rings, and cages) for every degraded bearing.

Two accelerometers were installed on the bearing box, and the accelerometer was used to collect raw vibration signal data in the horizontal and vertical directions of the bearing. The data is shown in Figure 11. Record 2560 data points every ten seconds, or 0.1 seconds of data, with a sampling rate of 25.6 kHz. Two datasets were selected here: bearing 1 and bearing 2.

3.2.2. Performance test
The following three traditional prediction network models were selected for comparative experiments at the same conditions: (1) BP, (2) LSTM, and (3) GRU. The prediction results are demonstrated in Figure 12, Table 10, and Table 11.

Out of the whole dataset, the first 75% of data points were used as the training set, and the last 25% as the test set. The prediction accuracy of the model was determined by comparing the bearing health indicator value of the test set with the actual health indicator. It can be seen in Figure 12, Table 10, and Table 11 that the prediction accuracy of the proposed method is significantly better than that of the other five traditional models. For the prediction of bearing 1, MAE and RMSE of the proposed model decreased by 46.3% and 48.1%, respectively, and R^2 increased by 66.0%, compared with those of GRU, the model with the best performance indicators out of the five traditional models. For the prediction of bearing 2, MAE and RMSE of the proposed model decreased by 32.2% and 34.5%, respectively, and R^2 increased by 61.2%, compared with those of GRU, the model with the best performance indicators out of the three traditional models.

3.2.3. Ablation experiment
In the same way, in order to make the research results more objective, compare with the latest related research work, and evaluate the role and effect of several key modules in the proposed model, the key modules in the proposed model were removed or replaced in turn for ablation research, while other parameters were kept unchanged. The following four different prediction models were designed for comparative analysis: (1) POA-CNN-SE-BIGRU, (2) CNN-SE-BIGRU, (3) CNN-BIGRU, and (4) BIGRU. The results are shown in Figure 13, Table 12, and Table 13.

Clearly, the convergence curves of bearing 1 and bearing 2 are shown in Figure 13; the black line represents the

Table 12. Results of bearing 1

Prediction model	MAE	RMSE	R^2
BIGRU [23]	0.74	0.94	0.64
CNN-BIGRU [24]	0.68	0.88	0.68
CNN-SE-BIGRU	0.62	0.77	0.76
POA-CNN-SE-BIGRU	0.55	0.67	0.82
Proposed Model	0.44	0.55	0.88

Table 13. Results of bearing 2

Prediction model	MAE	RMSE	R^2
BIGRU [23]	1.19	1.57	0.54
CNN-BIGRU [24]	1.08	1.50	0.58
CNN-SE-BIGRU	1.12	1.45	0.61
POA-CNN-SE-BIGRU	0.96	1.21	0.73
Proposed Model	0.78	1.08	0.79

IPOA, and the red line represents the POA. This shows that IPOA has a low fitness value and fast convergence speed.

It can be seen in Table 12 and Table 13 that the predictive ability of the proposed model is better than that of other comparison models. The improvement of prediction results is attributed to the IPOA and the introduction of the SE attention mechanism. Taking the prediction results of bearing 1 as an example, it is shown in the ablation experiment that from BIGRU to CNN-BIGRU, MAE and RMSE decreased by 8.1% and 6.4%, respectively, and R^2 increased by 6.3%; When SE was introduced, as in CNN-SE-BIGRU, compared with CNN-BIGRU, MAE decreased by 8.8%, RMSE decreased by 12.5%, and R^2 increased by 11.8%. This means that the introduction of the SE attention mechanism can enhance the ability of the network to capture effective features and reduce the interference of useless information to a certain degree. After POA was added, as in POA-CNN-SE-BIGRU, compared with CNN-SE-BIGRU, MAE and RMSE decreased by 11.3% and 13.0%, respectively, and R^2 increased by 7.9%. This means POA is effective in optimizing model parameters. Compared with those of POA-CNN-SE-BIGRU, of the proposed model, MAE and RMSE decreased by 20% and 10.7%, respectively, and R^2 increased by 8.2%. This is achieved from the multi-strategy hybrid improvement of IPOA.

4. CONCLUSIONS

Aiming to eliminate the difficulty in determining the health indicator of rail stripping damage development of heavy-haul railway lines and to solve the problem of low prediction accuracy of the conventional CNN-BIGRU model, a new, improved Pelican algorithm was proposed for predicting the rail stripping development trend by integrating the SE channel attention mechanism and CNN-BIGRU neural network. Multi-domain manually extracted features of rail stripping were selected. Compared with single-domain features, they can reflect more comprehensively the degradation information of rail vibration signals. By introducing the Spearman's correlation and PCA method for feature selection and dimensionality reduction, the influence of the correlation between features on the development of stripping damage was reduced so that the comprehensive health indicator after fusion is more in line with the stripping development trend. The parameters of the proposed model were optimized by using the IPOA improved through a variety of strategies. Finally, by comparing with multiple experiments, it was verified that the proposed method is reasonable and effective.

However, manually extracting features consumes a lot of time, requires substantial manpower, and incurs high material costs. In the future, we will consider directly using the original vibration signals as the input of the deep network to construct a health indicator that describes the development process of rail stripping damage so as to improve the rapidity of the model while ensuring the accuracy of prediction.
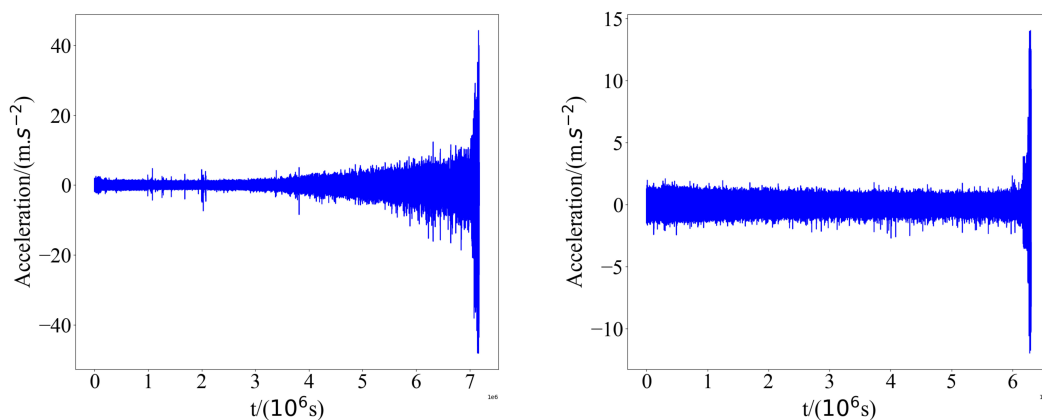
Figure 11. We present Vibration data of bearings in Figure 11A-B. A: vibration data of bearing 1; B: vibration data of bearing 2.
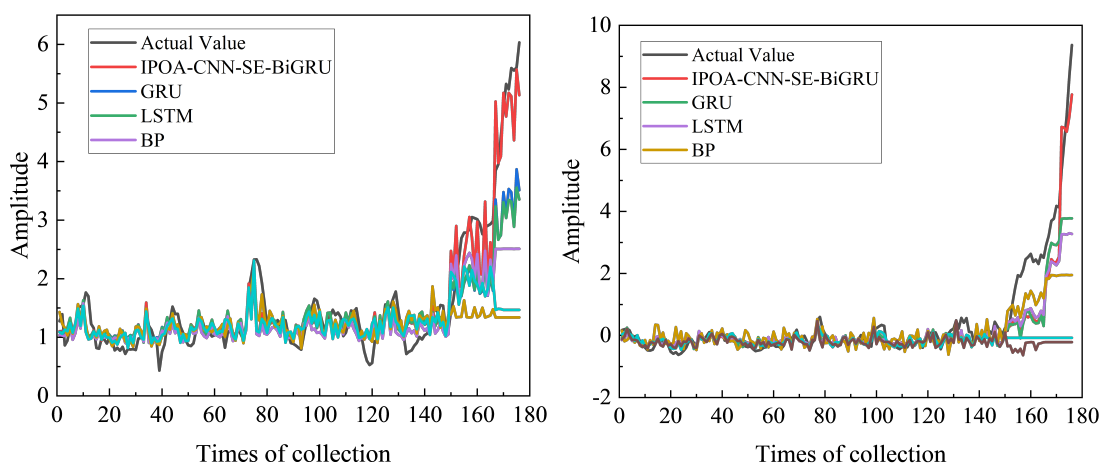

Figure 12. We present Prediction results of bearing damage development in Figure 12A-B. A: prediction results of bearing 1; B: prediction results of bearing 2.
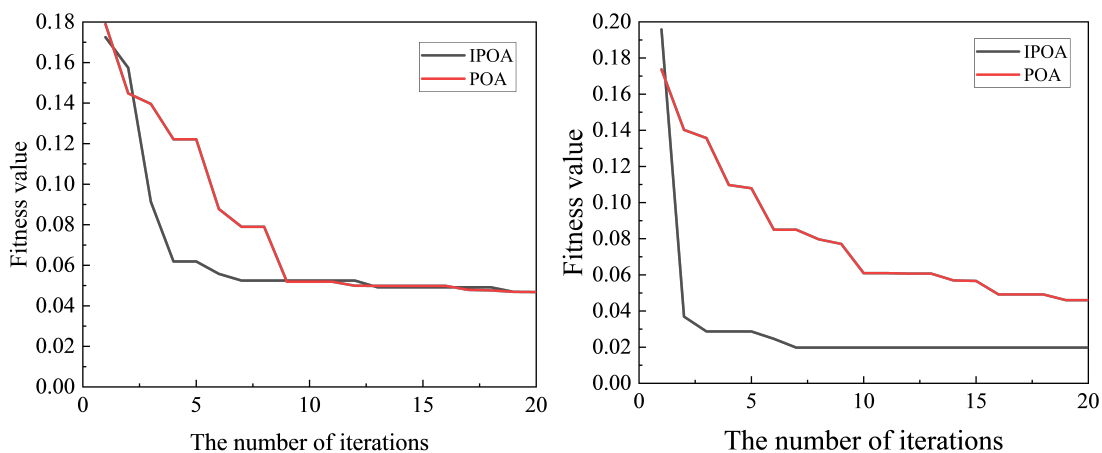

Figure 13. We present Convergence results of IPOA and POA in Figure 9A-B. A: convergence results of IPOA and POA of stripping 1; B: convergence results of IPOA and POA of stripping 2.

DECLARATIONS

Acknowledgments

Authors' contributions
Implemented the methodologies and presented and wrote the paper: Jiang C
Performed oversight and leadership responsibility for the research activity planning and execution and developed ideas and evolution of overarching research aims: Zhang C, Liu J
Performed providing administrative and Dataset support: Yang W, He J
All authors have revised the text and agreed to the published version of the manuscript.

Availability of data and materials
Not applicable.

Financial support and sponsorship

Conflicts of interest
All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Copyright

REFERENCES

1. Brunel JF, Charkaluk E, Dufrénoy P, Demilly F. Rolling contact fatigue of railways wheels: influence of steel grade and sliding conditions. *Proc Eng* 2010;2:2161–9. DOI
2. Pavlík A, Gerlici J, Lack T, Hauser V, Šťastniak P. Prediction of the rail-wheel contact wear of an innovative bogie by simulation analysis. *Trans Res Proc* 2019;40:855–60. DOI
3. Madge JJ, Leen SB, McColl IR, Shipway PH. Contact-evolution based prediction of fretting fatigue life: effect of slip amplitude. *Wear* 2007;262:1159–70. DOI
4. Liu Y, Jiang S, Wu YP, Duan ZD, Wang LB. Effects of spallation on rail thermo-elasto-plasticity in wheel-rail sliding contact. *Journal of Traffic and Transportation Engineering* 2016;16:46–55. (in Chinese) DOI
5. Zan XD, Li XT, Xing SB, Zhang YK, Jiang XY. Analysis of rail surface shelling resulting from fatigue crack propagation. *Journal of Railway Science and Engineering* 2018;15:3082–8. (in Chinese) DOI
6. Ma S, Liu X, Ren S, Chen Z, Liu Y. Research on side wear prediction of curve rail in shuohuang heavy haul railway. *J Mech Eng* 2021;57:118–25. (in Chinese) DOI
7. Wang L, Yuan H, Na WB, et al. Optimization of the re-profiling strategy and remaining useful life prediction of wheels based on a datadriven wear model. *Systems Engineering-Theory & Practice* 2011;31:1143–1152. (in Chinese). Available from: https://en.cnki.com.cn/Article_en/CJFDTOTAL-XTLL201106020.htm [Last accessed on 20 Nov 2023]
8. Li B, Yang Z, Xing Z, Gao X. Optimization of wheel re-profiling strategy based on a statistical wear model. In: Jia L, Qin Y, Suo J, Feng J, Diao L, An M, editors. Proceedings of the 3rd International Conference on Electrical and Information Technologies for Rail Transportation (EITRT) 2017; Lecture Notes in Electrical Engineering, vol 483. Springer, Singapore. DOI
9. Andrade AR, Stow J. Statistical modelling of wear and damage trajectories of railway wheel sets. *Qual Reliab Eng Int* 2016;32:2909–23. DOI

10. Han P, Zhang WH. A new binary wheel wear prediction model based on statistical method and the demonstration. *Wear* 2015;324-325:90–9. DOI

11. Xiang L, Yang X, Hu AJ, Su H, Wang PH. Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks. *Appl Energy* 2022;305:117925. DOI

12. Sun LM, Shang ZQ, Xia Y, Bhowmick S, Nagarajaiah S. Review of bridge structural health monitoring aided by big data and artificial intelligence: from condition assessment to damage detection. *J Struct Eng* 2020;146:04020073. DOI

13. Cheng QL, Peng B, Li Q, Liu SP. A rolling bearing fault diagnosis model based on WCNN-BiGRU. In: 2021 China Automation Congress(CAC) 2021, pp. 3368-72. DOI

14. Ye YG, Gao H, Huang CH, et al. Computer vision for hunting stability inspection of high-speed trains. *Measurement* 2023;220:113361. DOI

15. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell* 2023;220:113361. DOI

16. Trojovský P, Dehghani M. Pelican optimization algorithm: a novel nature-inspired algorithm for engineering applications. *Sensors* 2022:1424-8220. DOI

17. Hu SS, Liu H, Feng YF. Tool wear prediction in flass fiber reinforced polymer small hole drilling based on an improved circle chaotic mapping grey wolf algorithm for BP neural network. *Appl Sci* 2023;ISBN No. 2076-3417. DOI

18. Liu JS, Yuan MM, Zuo F. Global search-oriented adaptive leader salp swarm algorithm. *Contr Dec Contr Dec* 2021;36:2152–60. DOI

19. Wu WS. A study on Rényi entropy and Shannon entropy of image segmentation based on finite multivariate skew t distribution mixture model. *Math Methods in App Sciences* 2021. DOI

20. Deng Y, Qiao L, Zhu JC, Yang B. Mechanical performance and microstructure prediction of hypereutectoid rail steels based on BP neural networks. *IEEE Access* 2020;8:41905–12. DOI

21. De Simone L, Caputo E, Cinque M, et al. LSTM based failure prediction for railway rollin-g stock equipment. *Exp Syst Appl* 2023;222:119767. DOI

22. Li YJ, Chen CJ, Sun Q. Application of GRU in Prediction of Subway Wheel Wear. In: 2022 International Conference on Machine Learning and Knowledge Engineering (MLKE) 2022. pp. 212–6. DOI

23. Fan MG, Zheng W. Feature selection for prediction of railway disruption length. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC); 2019. pp. 351–6. DOI

24. Zhang N, Shen SL, Zhou A. A new index for cutter life evaluation and ensemble model for prediction of cutter wear. *Tunnelling and Undergroun-d Space Technology* 2023. DOI