

Research Article

Open Access



Intelligent prediction of rail corrugation evolution trend based on self-attention bidirectional TCN and GRU

Jian-Hua Liu¹, Wei-Hao Yang¹, Jing He², Zhong-Mei Wang¹, Lin Jia¹, Chang-Fan Zhang¹, Wei-Wei Yang³

¹College of Railway Transportation, Hunan University of Technology, Zhuzhou 412007, Hunan, China.

²College of Electrical and Information Engineering, Hunan University of Technology, Zhuzhou 412007, Hunan, China.

³Zhuzhou Qingyun Electric Locomotive Accessories Factory Co., Ltd, Zhuzhou 412007, Hunan, China.

Correspondence to: Dr. Zhong-Mei Wang, College of Railway Transportation, Hunan University of Technology, No 88, Taishan West Road, Taiyuan District, Zhuzhou 412007, Hunan, China. E-mail: wangzhongmei@hut.edu.cn

How to cite this article: Liu JH, Yang WH, He J, Wang ZM, Jia L, Zhang CF, Yang WW. Intelligent prediction of rail corrugation evolution trend based on self-attention bidirectional TCN and GRU. *Intell Robot* 2024;4(4):318-38. <http://dx.doi.org/10.20517/ir.2024.20>

Received: 28 Aug 2024 **First Decision:** 14 Sep 2024 **Revised:** 21 Sep 2024 **Accepted:** 11 Oct 2024 **Published:** 18 Oct 2024

Academic Editor: Simon Yang **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Analyzing the evolution trend of rail corrugation using signal processing and deep learning is critical for railway safety, as current traditional methods struggle to capture the complex evolution of corrugation. This present study addresses the challenge of accurately capturing this trend, which relies significantly on expert judgment, by proposing an intelligent prediction method based on self-attention (SA), a bidirectional temporal convolutional network (TCN), and a bidirectional gated recurrent unit (GRU). First, multidomain feature extraction and adaptive feature screening were used to obtain the optimal feature set. These features were then combined with principal component analysis (PCA) and the Mahalanobis distance (MD) method to construct a comprehensive health indicator (CHI) that reflects the evolution of rail corrugation. A bidirectional fusion model architecture was employed to capture the temporal correlations between forward and backward information during corrugation evolution, with SA embedded in the model to enhance the focus on key information. The outcome was a rail corrugation trend prediction network that combined a bidirectional TCN, bidirectional GRU, and SA. Subsequently, a multi-strategy improved crested porcupine optimizer (CPO) algorithm was constructed to automatically obtain the optimal network hyperparameters. The proposed method was validated with on-site rail corrugation data, demonstrating superior predictive performance compared to other advanced methods. In summary, the proposed method can accurately predict the evolution trend of rail corrugation, offering a valuable tool for on-site railway maintenance.

Keywords: Mahalanobis distance, rail corrugation, evolution trend prediction, improved crested porcupine optimizer, hybrid time series network



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



1. INTRODUCTION

Long-term wheel-rail contact on railway lines can cause various types of damage, particularly in sections with small curvature radius, where corrugation damage is more prevalent^[1-3]. Rail corrugation primarily affects the inner surface of the rail in a curved section, resulting in periodic wavy wear. If left undetected and unrepaired, corrugation can cause train vibrations, significantly reducing its operating stability. In severe cases, rail breakage and major accidents, such as train derailments, can also occur^[4,5]. Therefore, in railway health management, in-depth research on the evolution of rail corrugation is critical^[6] to ensuring the safe operation of rail transit^[7].

Over the past years, scholars have conducted in-depth research on the generation and evolution process of rail corrugation, mainly using two methods: mechanism modeling and data-driven prediction. In mechanism modeling, the wheel-rail transient dynamics method is used to establish a model that reflects the evolution process of corrugation. Additionally, by using mechanical simulation software, scholars have constructed wheel-rail coupling finite element models and rail elastic-plastic analysis models to further explore the generation and evolution^[8,9] of corrugation. For example, Wang *et al.* established a vehicle-track space coupling model using multibody dynamics software and conducted a dynamic analysis of the corrugation section^[1]. Cui *et al.* established a finite element model of the wheel-rail system and a wear model for corrugation using typical rail corrugation on a curve with a small radius as the research object; they then elucidated the development mechanism of corrugation by studying the dynamic response of the wheel-rail on the rail surface^[2]. However, these methods rely on prior knowledge of factors, such as the damage mechanism, and are highly theoretical. Furthermore, achieving an optimal damage evolution process using these methods in a complex train operating environment is challenging.

In data-driven research, scholars typically use experimental or on-site data to extract damage degradation features. Machine and deep learning methods are employed for damage diagnosis or prediction tasks without requiring an in-depth understanding of the internal damage mechanisms, as these methods can indirectly consider various influencing factors^[10-13]. For example, Xiao *et al.* used machine learning to detect and assess corrugation damage in heavy haul railways; their approach, which was based on support vector machines and other technologies, could effectively detect rail corrugation damage^[14]. Deep network models such as gated recurrent units (GRUs), temporal convolutional networks (TCNs), and attention mechanisms are widely used in industrial equipment for damage diagnosis and degradation trend prediction^[15-20] because of their exceptional feature extraction and nonlinear mapping abilities. For example, Zhang *et al.* introduced a squeeze-excitation channel attention mechanism into a combined model of a convolutional neural network (CNN) and bidirectional GRU (BiGRU); this integration demonstrated that the addition of an attention mechanism improved the capability of the network to focus on excellent features^[21]. Liu *et al.* used a dynamic multiscale gated causal convolution method combined with a GRU to effectively predict the actual degradation trend of rail corrugation and address poor generalization caused by small data samples^[22]. Additionally, in the general damage evolution prediction task, degradation is a continuous change process with a front-back relationship over time^[23]. Currently, most scholars do not consider the relationship between the time series before and after the damage signal. In a complex time-series prediction task, a single model often has limitations in terms of generalization, robustness, and adaptability. The current hybrid temporal prediction networks typically rely on extensive experiments and parameter-tuning processes, thereby increasing the computational cost and making the optimality of the selected hyperparameters difficult.

To address the aforementioned shortcomings, this study constructed a self-attention (SA) bidirectional TCN and GRU (SA-BiTCN-BiGRU) hybrid network and used a new multi-strategy improved crested porcupine optimizer (MICPO) algorithm for automatic hyperparameter optimization. The proposed model integrated the advantages of each module, exhibiting robust time-series modeling capabilities, perceiving dynamic changes in a time series, and assigning more weight to important time-series features. Thus, the prediction accuracy

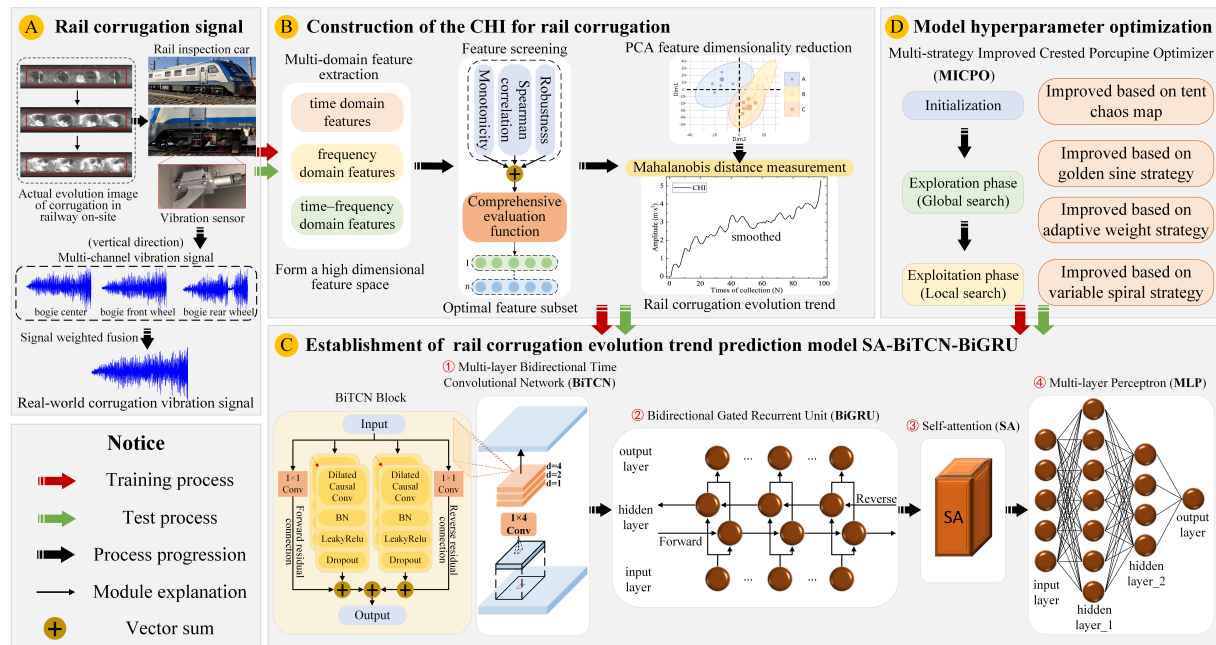


Figure 1. Frame diagram of rail corrugation trend prediction.

of the evolution trend of rail corrugation improved. The MICPO algorithm could automatically determine the optimal network hyperparameters for the proposed network using the four improvement strategies and its superior global search ability, thereby enhancing the network’s prediction accuracy and reducing the need for blind manual adjustment of hyperparameters. Finally, the efficacy and superiority of the proposed methodology were verified through experiments and compared with other advanced methods.

The remainder of this study is organized as follows. Section 2 introduces the construction method of the rail corrugation’s comprehensive health indicator (CHI), corrugation evolution trend prediction model, and model hyperparameter optimization algorithm. Section 3 describes the experimental setup and preprocessing of the rail corrugation dataset, and subsequently analyzes the experimental results in detail. Section 4 provides a comprehensive summary of the research content and proposes current limitations and future research directions. Finally, Section 5 concludes the study.

2. METHODS

Based on the current research background, this section provides a detailed description of the process for predicting the evolutionary trend of rail corrugation. A corrugation CHI was established using the collected on-site dataset. A SA-BiTCN-BiGRU hybrid network was used to predict the evolution trend of rail corrugation, and the MICPO algorithm was constructed to adaptively adjust the hyperparameters of the network. The overall framework is illustrated in Figure 1.

As can be seen from Figure 1, first, by observing the damage changes in the corrugation image, the three vibration sensors were installed at the front wheel, rear wheel, and center position of the bogie on the track inspection car to collect corrugation vibration data in the vertical direction. After preprocessing the collected data, the rail corrugation vibration signal was obtained. Subsequently, multidomain feature extraction, feature screening, feature dimensionality reduction, and the Mahalanobis distance (MD) measurement methods were applied to this corrugation vibration signal, resulting in a CHI that effectively characterized the evolution trend of rail corrugation. The CHI was then input into the SA-BiTCN-BiGRU hybrid network to predict

the evolution trend of rail corrugation. The network integrated the advantages of BiTCN, BiGRU, and SA to address the limitations of existing models. Finally, the MICPO algorithm was used to accurately select the optimal network model hyperparameters, thereby effectively improving the prediction accuracy of the model.

2.1 Collection of rail corrugation signal and construction of corrugation CHI

In this study, three vibration sensors installed on the track inspection car were used to obtain vibration data of corrugation damage from different positions in the same direction. Compared with the data of a single sensor, the multi-channel data contains richer feature information and can more comprehensively reflect the changing characteristics of corrugation damage^[24]. Therefore, to fully explore the vibration information of the three channels, we first normalized the data of each channel to reduce the impact of the difference in signal distribution between different channels. The multi-channel signal fusion method based on kurtosis weight was then used to calculate the fusion weight of the three channels, and the signals from each channel were subjected to weighted fusion. The kurtosis value can effectively reflect the severity of rail corrugation damage. The channel with a higher kurtosis value is considered to be more sensitive to the reflection of corrugation damage, so a higher weight is assigned to ensure that more representative vibration signals have a more significant impact on the overall analysis results during the fusion process, so that the merged vibration signals can reflect the changing trend of rail corrugation damage more comprehensively and reliably^[25,26].

The CHI is an indicator used to evaluate and quantify the evolution trend of rail corrugation. The construction of a CHI is a preprocessing step for predicting the evolution of corrugation, which influences the effectiveness of subsequent prediction tasks^[27]. However, in a complex environment, various adverse factors may lead to significant deviations in the extracted rail corrugation vibration data, resulting in a lack of reliability in the constructed CHI. Therefore, this study used a custom range box line method to identify outliers in the corrugation vibration data and performed a mean correction on these outliers, thereby improving data quality. To accurately construct the CHI of corrugation and overcome the problem of relying on manual experience selection for single physical and fusion indicators, this study establishes a CHI that reflects the evolution of corrugation. The process steps are described as follows.

First, the vibration data of rail corrugation collected from the field contain numerous degradation features reflecting the evolution process of corrugation. The amplitude of these features usually deviates from the normal range with time, indicating that the corrugation damage is intensifying^[28]. Therefore, this study extracted time-domain, frequency-domain, and time-frequency domain feature indicators from the data, such as the maximum value, root mean square (RMS), standard deviation, and pulse index. These feature indicators effectively reflect corrugation degradation through the concretization of abstract real data.

Subsequently, three evaluation indicators, monotonicity (M), Spearman's correlation coefficient (S), and robustness (R), were used to quantify the damage features of rail corrugation^[29]. Concurrently, to conduct comprehensive evaluation of each rail's corrugation features, this study used normalization processing to quantify the three evaluation indicators to the same scale, eliminating the impact of dimension, and obtained comprehensive evaluation indicator (C)^[30,31] through linear weighted combination of the three evaluation indicators, using it to calculate the comprehensive score of each feature, and adaptively screen out the features sensitive to the change of rail's corrugation state, forming an optimal feature subset. The definitions of M , S , R , and C are

$$M = \frac{1}{T-1} \left| X \left(\frac{d}{df_i} > 0 \right) - Y \left(\frac{d}{df_i} < 0 \right) \right| \quad (1)$$

$$S = 1 - \frac{6 \sum d_i^2}{T(T^2 - 1)} \quad (2)$$

$$R = \frac{1}{T} \sum_{t=1}^T \exp \left(- \left| \frac{f_t - \tilde{f}_t}{f_t} \right| \right) \quad (3)$$

$$C = \frac{M + S + R}{3} \quad (4)$$

respectively, where T indicates the length of the degradation feature sequence, f_t is the extracted value of the feature at time t , $X(\cdot)$ denotes the number of positive derivatives in the degradation feature sequence, $Y(\cdot)$ denotes the number of negative derivatives in the degradation feature sequence, Σ denotes the summation symbol, d_t is the difference between the degradation feature index sequence and time series, $\exp(\cdot)$ represents an exponential function based on the natural constant e , and \tilde{f}_t is the value of the feature at time t after sliding average.

Principal component analysis (PCA) was then used to reduce the dimensionality of the optimal feature subset, and the principal components were weighted according to their contribution degrees to generate a multidimensional principal component vector that retains the important information of the original optimal feature subset.

Finally, the MD^[32] was used to calculate the difference between the initial and subsequent samples in the generated multidimensional principal component vector. The obtained results were then smoothed using the exponential weighted moving average, yielding the CHI, which reflected the evolution of corrugation. The MD is calculated as follows:

$$D_M(m, n) = \sqrt{(m - n)^T K^{-1} (m - n)} \quad (5)$$

where m and n are the sample vectors; K represents the covariance matrix of the corrugation evolution features.

2.2 Establishment of trend prediction model for rail corrugation

To accurately predict the evolution trend of rail corrugation, we constructed a SA-BiTCN-BiGRU model. Using the initial corrugation data in the established CHI as the input, the subsequent CHI values were predicted.

The structure of the model is illustrated in [Figure 2](#). First, the bidirectional local features of the initial corrugation data were effectively extracted using a three-layer BiTCN to improve the receptive field and feature extraction capability of the model. Subsequently, based on the local features extracted by BiTCN, BiGRU was used for time-series prediction, and the output results were passed through the Leaky rectified linear unit (ReLU) nonlinear activation function and dropout regularization technology. The attention weight provided by the SA was then used to enhance the interpretability of the network. Finally, the multilayer perceptron (MLP) network output continuous prediction results and the error between them and the actual value was calculated to evaluate the prediction effect of the model.

2.2.1 BiTCN

In this study, the constructed corrugation CHI is a continuous process that changes over time, and the data are closely related. To capture the features of the corrugation CHI over a wider range, a BiTCN was constructed to comprehensively consider the historical and forthcoming temporal information of the corrugation CHI. Additionally, multiple dilated causal convolution layers were stacked to improve the receptive field, effectively observe the change patterns in the rail corrugation data, and enhance the model's capacity to acquire key information. The structure of the BiTCN is shown in [Figure 3](#).

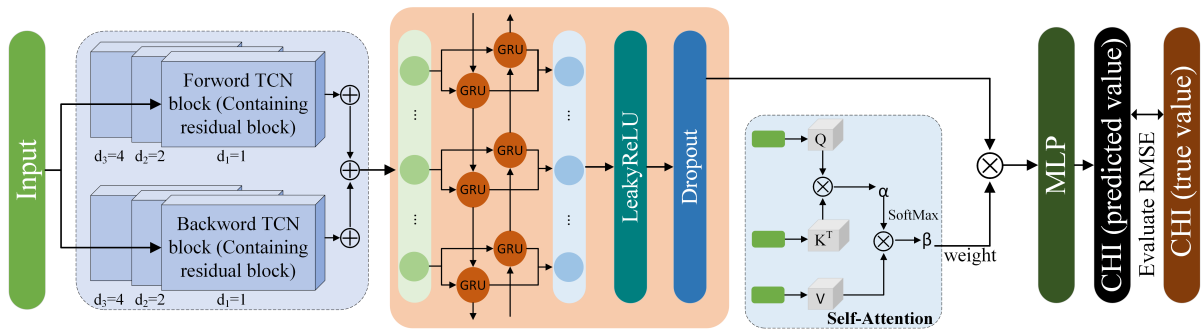


Figure 2. Architecture of prediction model for rail corrugation trend.

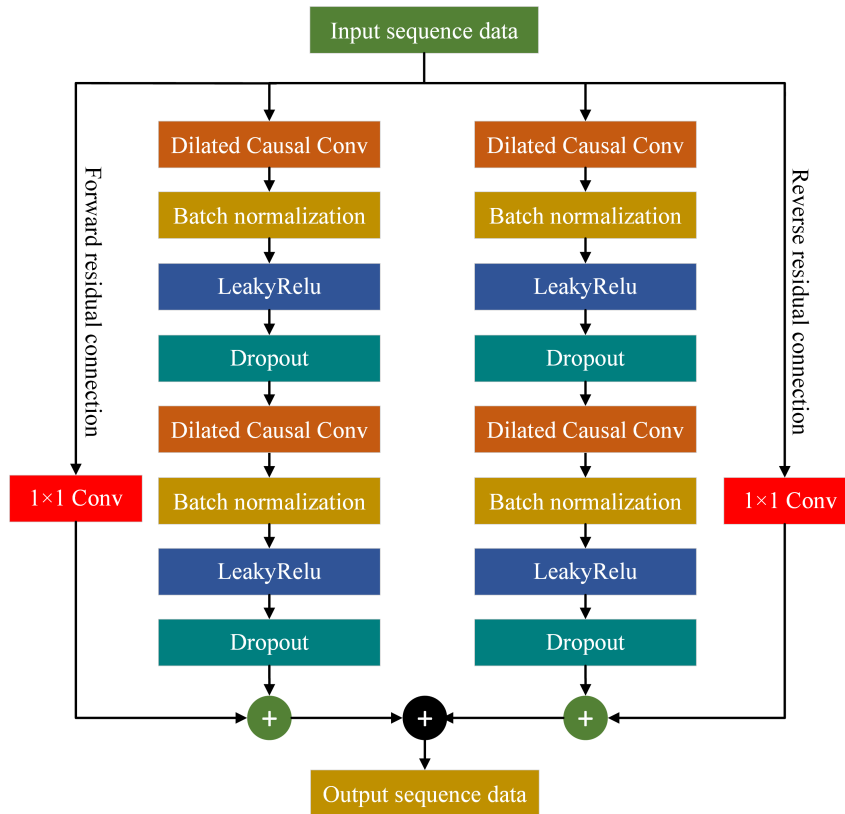


Figure 3. Schematic of BiTCN structure. BiTCN: Bidirectional temporal convolutional network.

As shown in Figure 3, BiTCN consists of a forward and a reverse TCN residual block linked together. The model's output was the combined training result of the two blocks. Each residual block contained two layers of dilated causal convolution, which enlarged the receptive field of the network. The input sequence data were derived from the one-dimensional rail corrugation CHI, and the feature information at different scales was captured using the dilated convolution operation. A batch normalization layer was employed to stabilize the model training process. The Leaky ReLU activation function enabled the BiTCN module to train a deeper network while addressing dead neurons and vanishing gradient. Additionally, dropout regularization technology was added to reduce overfitting. To accommodate possible differences in the number of input and output channels in the model, a 1×1 convolution layer was added in each training direction for the residual connection, and the number of feature channels was adjusted to suit the feature representations of different levels.

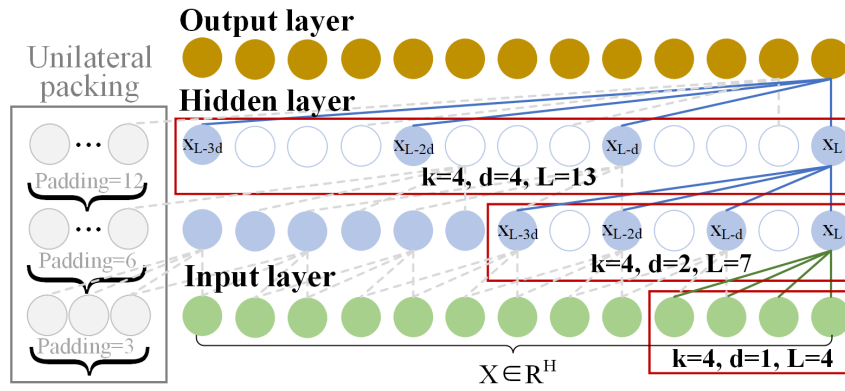


Figure 4. Visualization of dilated causal convolution.

The core concept of the dilated causal convolution involves the insertion of zero elements into the convolution kernel, which modifies the structure of the kernel and effectively expands the receptive field of the model. This enables each convolution output to encompass a broader range of time information, effectively mitigating vanishing gradient caused by numerous layers in the common convolution and enabling the model to extract more information on corrugation evolution [33]. The internal structure of the dilated causal convolution is shown in Figure 4 and defined below:

$$y = \sum_{k=0}^{K-1} \omega[k] \cdot x[L - d \cdot k] \tag{6}$$

where y is the output of the dilated causal convolution layer, $\omega[k]$ denotes the weight of the convolution kernel k , $x[L - d \cdot k]$ denotes the value of the input sequence element, L is the length of the input sequence, and d represents the expansion rate.

2.2.2 BiGRU

The GRU is a temporal prediction network proposed to alleviate the vanishing gradient problem of a recurrent neural network (RNN) [34]. The evolution of corrugation is closely related to information from past and future data, and unidirectional GRU may fail to capture this bidirectional information transmission mode. Therefore, this study constructed a BiGRU to infer the relationship between past and future corrugation characteristics and the current corrugation amplitude to improve the model's sensitivity and predictive capability regarding dynamic changes in the time series of corrugation characteristics. The BiGRU is calculated using

$$\begin{cases} \vec{h}_n = GRU(x_n, \vec{h}_{n-1}) \\ \overleftarrow{h}_n = GRU(x_n, \overleftarrow{h}_{n-1}) \\ h_n = \alpha_n \vec{h}_n + \beta_n \overleftarrow{h}_n + b_n \end{cases} \tag{7}$$

where $GRU(\cdot)$ denotes the gated cycle unit, x_n is the input, \vec{h}_n and \overleftarrow{h}_n represent the output status of the forward and reverse hidden layers, respectively, α_n and β_n are the corresponding output weights, and b_n is the corresponding bias. The structure of the BiGRU is shown in Figure 5.

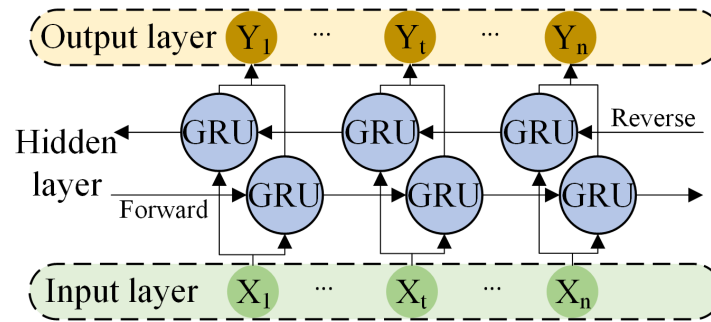


Figure 5. Schematic of BiTCN structure. BiTCN: Bidirectional temporal convolutional network.

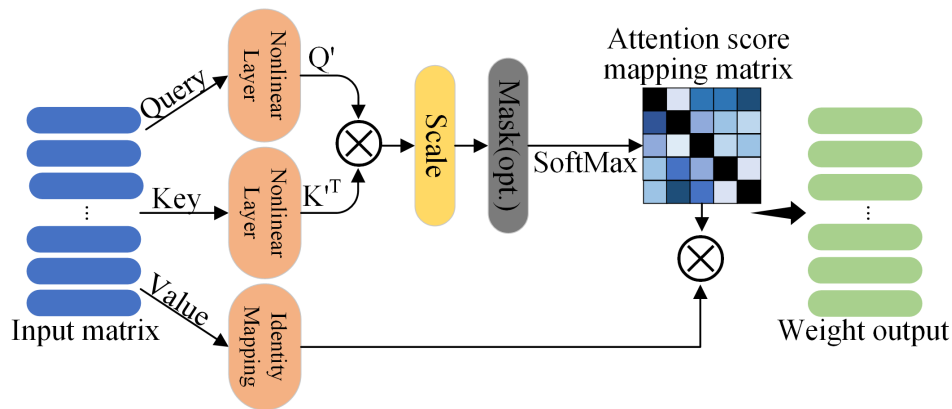


Figure 6. Illustration of SA mechanism. SA: Self-attention.

The entire network is composed of an input layer, two layers of GRUs in opposite directions, and an output layer. The input is the value after BiTCN feature extraction, and the output is determined based on the cycling training results of the BiGRU unit.

2.2.3 SA mechanism

As a variant of the attention mechanism, SA^[35] is mainly used to process serial data such as the rail corrugation time-series data used in this study. The network can calculate the attention weights of various positions at different time steps to improve its ability to obtain key information and integrate the content of all time steps. This study introduced and applied the SA mechanism to the process of model trend prediction, which was designed to improve the model's dependence on different locations in the input ripple CHI sequence. This allowed the model to better understand the internal correlations between the corrugation data at each moment, significantly improving its predictive performance. This technology can provide reliable decision-making support for the maintenance and management of railway systems. Figure 6 shows the structure of the SA.

As depicted in Figure 6, the structure initially computes and packages the query, key, and value vectors of all input matrices as matrices. The query and key vectors were used to perform a nonlinear transformation. The dot product and masking operations standardized the query and key vectors, masked invalid information, and generated an attention score. The mapping matrix of the attention score was then obtained after normalization using the softmax operation and multiplied by the value vector after identity mapping to acquire the weight output.

2.3 Model hyperparameter optimization based on MICPO algorithm

Certain hyperparameters significantly affected the predictive performance of the proposed model. For example, the convolution kernel size determined the capability of the model to capture corrugation characteristics in the time dimension. The number of BiGRU hidden layer units determines the complexity and learning ability of the network. To prevent the adverse effects of manual intervention in the selection of model hyperparameters, optimization algorithms are necessary to adaptively identify the most suitable model hyperparameters.

Consequently, model hyperparameter optimization was performed using the crested porcupine optimizer (CPO)^[36] algorithm. This algorithm simulates four different defense strategies when a crested porcupine (CP) engages in defense against predators. The first two strategies, sight and sound, represent the exploration phase of the algorithm; the last two strategies, odor and physical-attack, represent the exploitation phase of the algorithm. Different defense strategies have distinct optimization effects on various hyperparameters, guiding the algorithm to identify the optimal hyperparameters for the model. However, the original algorithm has certain limitations, such as decreasing population diversity and the tendency to get trapped in local optimality in the later stages of a search, leading to an inaccurate selection of hyperparameters. Therefore, a multi-strategy improvement method was constructed to optimize the initialization mode and defense strategy of the CPO algorithm to acquire better model hyperparameters and enhance the prediction accuracy of the model on the evolution trend of rail corrugation. The detailed improvement strategies for the CPO algorithm are discussed in the following subsections.

2.3.1 Improved tent chaos map

In the algorithm initialization stage, an improved tent map was employed to generate chaotic sequences and address issues related to the reduction of the CP population and its tendency to converge into the local optimal solution when the CPO algorithm approached the global optimum^[37]. This method introduced random variables into a traditional tent-chaos map. Thus, the diversity of the CP individuals was increased, and the chaotic sequence was prevented from falling into unstable periodic points during the iterative process defined as follows:

$$X_{i,j+1} = \begin{cases} \frac{X_{i,j}}{tent} + rand(0, 1), & 0 \leq X_{i,j} \leq tent \\ \frac{1-X_{i,j}}{1-tent} + rand(0, 1), & tent < X_{i,j} \leq 1 \end{cases} \quad (8)$$

where i and j represent the CP population number and current dimension, respectively; $tent$ denotes the chaos coefficient, and $rand(0, 1)$ represents a random number between 0 and 1.

2.3.2 Golden sine strategy

In this study, the golden sine strategy^[38] was incorporated into the CPO algorithm to enlarge its search space and address the lack of information exchange between CP individuals in the original algorithm, thereby improving the algorithm's ability for global optimization defined as follows:

$$X_{i,j}^{t+1} = X_{i,j}^t \times |\sin(D_1)| + D_2 \times \sin(D_1) \times |x_1 \times X_{i,j}^t - x_2 \times X_{i,j}^t| \quad (9)$$

where t denotes the number of iterations. After using this formula to improve the position update strategy of the algorithm, all CP individuals exchanged information with the optimal individuals in each exploration phase. Simultaneously, the golden section coefficient gradually reduced the search space of the CP individuals.

By controlling the moving distance and direction of the CP individuals, the CPO algorithm was optimized, further coordinating the algorithm's global exploration and local exploitation abilities.

2.3.3 Adaptive weight strategy

When executing the third defense strategy, the search step of the CP individual was not set in the original algorithm, resulting in excessive freedom while running the algorithm. The adaptive weight strategy can dynamically adjust the optimal position [39], thereby effectively enhancing the convergence effect and local exploitation ability of the CPO algorithm. This adjustment ensures that individuals with CP maintain a relatively safe distance from predators while executing the third defense strategy. Therefore, this study constructed an adaptive strategy that adjusted the weight coefficient ω based on the iteration count, allowing CP individuals to utilize different weights for optimal search lengths at different stages. The ω is obtained as

$$\omega = 1 - \cosh((\exp(t/T_{\max})) / \exp(1) - 1)^2 \quad (10)$$

where $\cosh()$ denotes the hyperbolic cosine function, and T_{\max} denotes the maximum number of iterations.

2.3.4 Variable spiral search strategy

Inspired by the whale optimization algorithm (WOA) [40], the variable spiral search strategy adjusts the original spiral parameters to become variable parameters that change with each iteration. This adjustment allows the algorithm to perform extensive searches in the early phase and an elaborate exploration of a small area in the late stage [41], enhancing its local exploitation ability in the fourth defense strategy. In this study, by constructing a variable spiral search strategy, CP individuals continued to search nearby after reaching the local optimal solution. This approach compensates for the unclear convergence effect of the original CPO during local exploration, which prevents deviations in the prediction accuracy of the model in the late stages of rail corrugation development. This strategy is established as

$$Z = X_{best}(t) \times (\exp(zl) \times \cos(2\pi l)) + X_{best}(t) \quad (11)$$

$$z = \exp(k \cos(\pi t / T_{\max})) \quad (12)$$

where X_{best} denotes the best fitness value, l represents a random number between -1 and 1, and k represents a variable parameter that should be set according to the specific strategy.

Based on the above analysis, a flowchart of the MICPO algorithm is constructed [Figure 7], where N and T_{\max} represent the population size and maximum number of function evaluations, respectively. T_f indicates a constant between 0 and 1, t denotes the number of current iterations, and i is the current i th individual. In the first iteration, all CP individuals passed through the position of the initialization solution and adopted a defense strategy to obtain the current optimal candidate solution. Subsequently, the algorithm entered the next iteration. First, the defense factor and the population number N were updated. Then, the CP individuals continue to search for the best candidate solution of the model according to the selected defense strategy. This process was repeated until the iterations were complete. Consequently, the optimal solution, which represents the best parameter of the model, was obtained and substituted into the SA-BiTCN-BiGRU hybrid network to optimize the prediction performance of the model.

2.4 Algorithm validation

In this study, six benchmark functions were used to conduct the optimization experiments. The MICPO algorithm was compared with the CPO [36], WOA [40], rime optimization algorithm (RIME) [42], grey wolf opti-

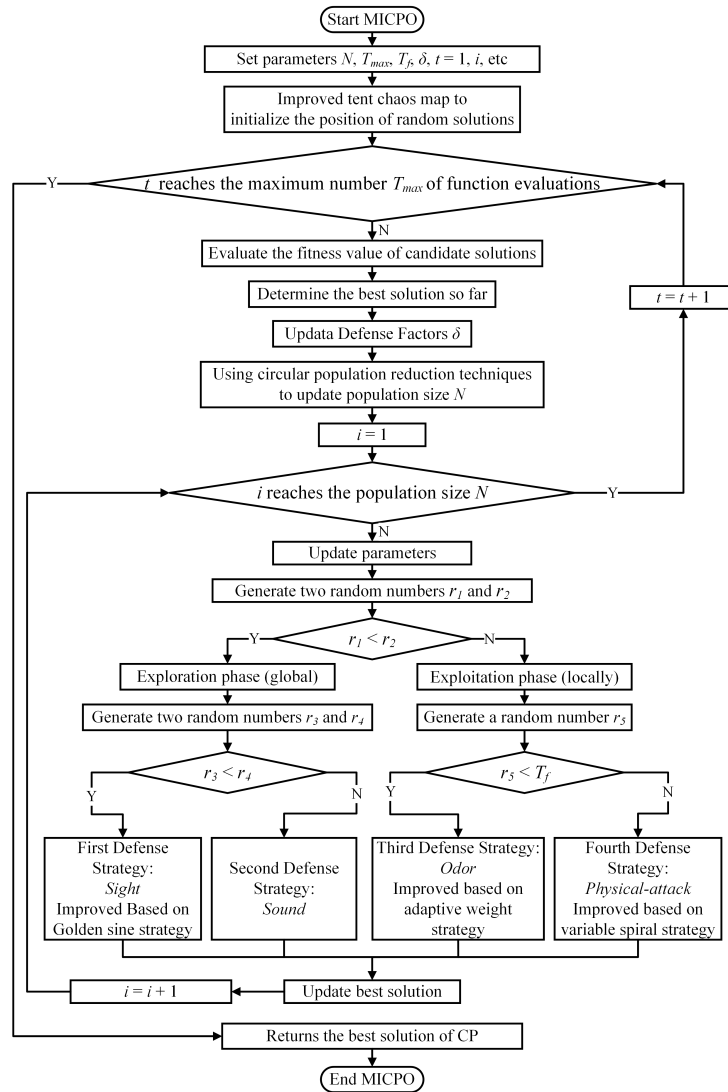


Figure 7. Flowchart of MICPO algorithm. MICPO: Multi-strategy improved crested porcupine optimizer.

mizer (GWO) [43], and dung-beetle optimizer (DBO) [44] algorithm to observe their optimal fitness values and convergence speed within a specified number of iterations, and verify the improvement effect of MICPO on the original CPO. Table 1 provides a detailed definition of the benchmark functions. F1-F3 are single-peak functions used to evaluate the local search capability of the algorithm. F4 is a multi-peak function with multiple local optimal values and requires a higher convergence performance of the algorithm. This function has important reference significance in the evaluation algorithm. F5 and F6 are the combined benchmark functions used to evaluate the global exploitation capacity of an algorithm. In this study, the population size of the experimental algorithm was set to 30, and each algorithm was optimized 100 times.

As shown in Figure 8, the convergence performance of the MICPO algorithm is effectively proven.

From the convergence curve presented in Figure 8, the CPO algorithm exhibits poor convergence performance and easily falls into the local optima, indicating that improvements in the CPO algorithm are necessary. In the unimodal function test shown in Figure 8A-C, the RIME, WOA, and the other optimization algorithms fell into local optima and slowly converged, indicating that the MICPO algorithm has certain competitive advantages

Table 1. Detailed information on benchmark function

ID	Benchmark function	Domain and dimensions	Optimal value
F1	$f_1(x) = \sum_{i=1}^n x_i^2$	$[-100, 100]^{30}$	0
F2	$f_1(x) = \sum_{i=1}^n x_i^2$	$[-100, 100]^{30}$	0
F3	$f_3(x) = \sum_{i=1}^n ix_i^4 + \text{random}[0, 1]$	$[-1.28, 1.28]^{30}$	0
F4	$f_4(x) = -20 \exp\left(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}\right) - \exp\left(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i)\right) + 20 + e$	$[-32, 32]^{30}$	0
F5	$f_5(x) = \sum_{i=1}^n \left[a_i - \frac{x_i(b_i^2 + b_1 x_1 + x_4)}{b_i^2 + b_1 x_1 + x_4} \right]$	$[-5, 5]^4$	3.075×10^{-4}
F6	$f_6(x) = -\sum_{i=1}^{10} [(x - a_i)(x - a_i)^T + c_i]^{-1}$	$[0, 10]^4$	-10

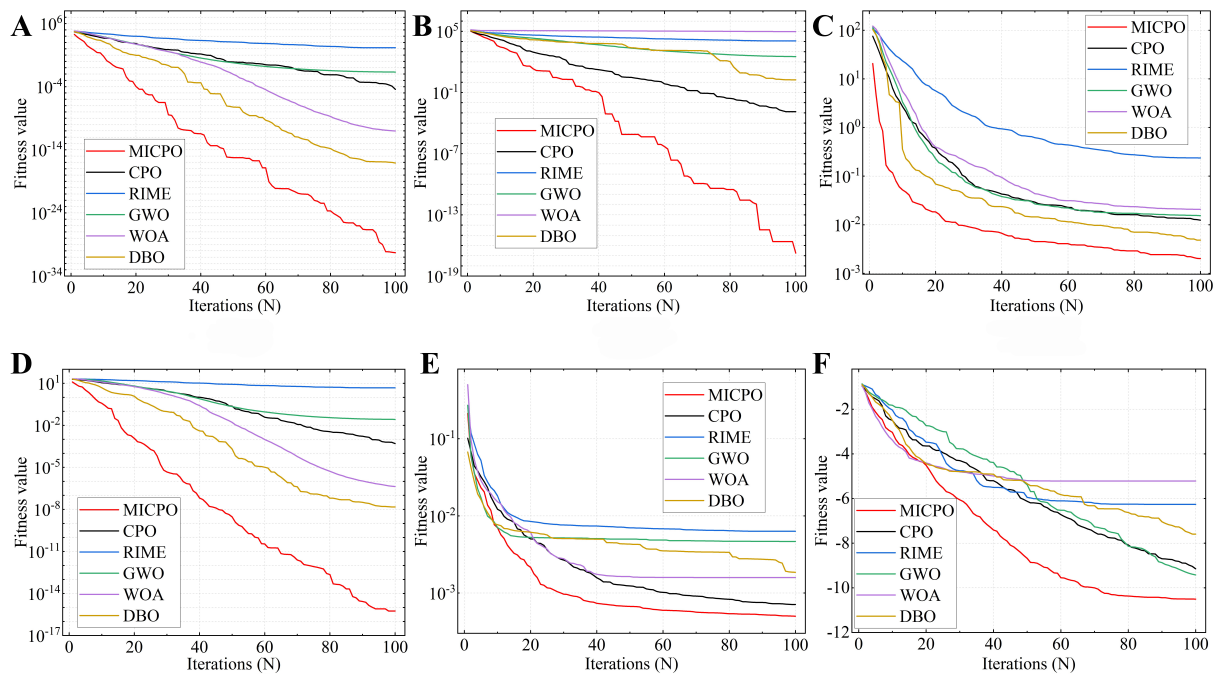


Figure 8. Convergence curves of different algorithms under different benchmark functions. (A-F) correspond to benchmark functions F1-F6, respectively.

over other optimization algorithms in solving unimodal high-dimensional functions. In the multippeak test function F4, the MICPO algorithm [Figure 8D], demonstrates an advantage by being the closest to the optimal solution within the specified number of iterations, which validates its effectiveness in improving the CPO algorithm, as well as its superiority in search accuracy and convergence speed. In the combined function test shown in Figure 8E and F, the MICPO and CPO algorithms demonstrate superior convergence performance compared to the RIME algorithm and other optimization algorithms, indicating their advancement in global optimization.

3. RESULTS

3.1 Experimental setup and rail corrugation dataset preprocessing

First, the code was written and debugged on a PyCharm platform, and the running environment consisted of a processor (Intel i7-12700H), 16 GB of random-access memory (RAM), a graphics card (RTX 3060), and a software environment with TensorFlow 2.13.0 and Python 3.9.18. The experimental data in this study were

Table 2. Evaluation results of health indicators

Indicator	<i>M</i>	<i>S</i>	<i>R</i>	<i>C</i>
RMS	0.1134	0.8336	0.9043	0.6171
PCA	0.1546	0.8706	0.9021	0.6424
LLE	0.1753	0.8983	0.9098	0.6611
Proposed indicator	0.2165	0.9322	0.928	0.6922

RMS: Root mean square; PCA: principal component analysis; LLE: linear embedding.

actual measurement data from a railway section in China. A track inspection car was used to collect vibration signals from a typical steel rail segment with corrugation, covering damage from slight to severe stages. These signals demonstrate the progression of corrugation damage^[21]. We collected 98 vibration samples on-site at the same time interval throughout the entire lifecycle of the rail after several months of continuous periodic testing, with each vibration sample containing 3,000 sample points; therefore, the original sample contained $98 \times 3,000$ data points. These data points represent the initial corrugation on the rail surface to the rail scrap. The overall vibration amplitude gradually increased with collection times, indicating that the deterioration degree of corrugation damage was worsening, reflecting the evolution of rail corrugation damage from budding to deterioration.

First, each collected sample was subjected to multidomain feature extraction to obtain 26 feature indicators that reflected the evolution of corrugation. The dimensions of the samples were 98×26 . Subsequently, the *M*, *S*, and *R* of each feature index were calculated; *C* was used to adaptively screen out the eight features with higher scores, and the corrugation optimal feature subset with a sample dimension of 98×8 was obtained. PCA was used to fuse the optimal feature subset, resulting in a two-dimensional principal component vector with a total contribution rate of 97%. The sample dimensions of the corrugation data were 98×2 . Finally, the MD was used to calculate the difference between the first column of the sample data and the subsequent 97 columns of sample data, resulting in 98×1 one-dimensional data. The corrugation CHI was obtained after smoothing to minimize the negative impact of outliers on the prediction of the evolution trend of rail corrugation.

3.2 Validation of the CHI construction method

To demonstrate the effectiveness and advantages of the method proposed for constructing the corrugation CHI, several commonly used methods for constructing health indicators were selected for comparison, including the RMS, PCA, and locally linear embedding (LLE) fusion indicators. Two fusion indicators were constructed using the optimal feature subset described in Section 3.1. The rail corrugation health indicator constructed using these four methods after smoothing is shown in Figure 9.

Figure 9 shows that these indicators are relatively sensitive to changes in the initial corrugation damage. However, the RMS indicator exhibits a larger overall fluctuation range, with the index value declining in the later stages of the corrugation evolution and deviating from the actual situation. The amplitude of the PCA indicator fluctuates significantly between the middle and late stages. The LLE indicator oscillates excessively in the early stages and becomes more stable in the middle and late stages, which is different from the actual situation. However, the CHI constructed in this study showed a better overall trend with fewer fluctuations. The indicator shows a sudden increase when the corrugation damage approached a qualitative change in the later stage, which aligns with the actual evolution law of on-site rail corrugation damage. The corrugation health indicator constructed by CHI is more consistent with the changing trend of the real-world data on rail corrugation vibration signals.

Furthermore, the *M*, *S*, *R*, and *C* [established by Equations (1)-(4)] were used to evaluate the HIs constructed using the four different methods. The results are listed in Table 2.

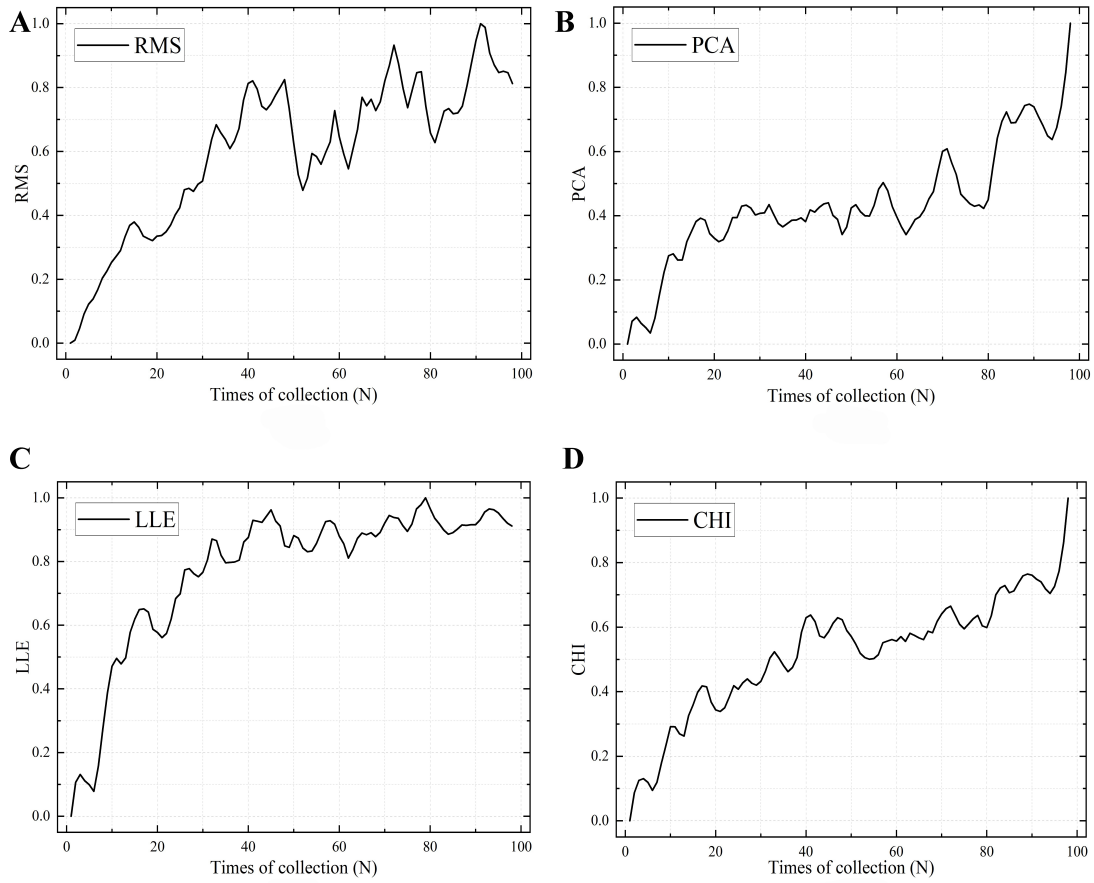


Figure 9. Health indicators of rail corrugation constructed by different methods. (A-D) correspond to RMS, PCA, LLE and CHI methods, respectively. RMS: Root mean square; PCA: principal component analysis; LLE: locally linear embedding; CHI: comprehensive health indicator.

Through the comparison of various indicators in Table 2, the constructed CHI achieved optimal performance in all cases, with the highest comprehensive evaluation function C . Therefore, this indicator is considered suitable for reflecting the evolution trend of rail corrugation.

3.3 Performance evaluation indicators

The root mean square error (RMSE), mean square error (MSE), and mean absolute error (MAE) were used to evaluate the performance of the model. These indices reflect the prediction effect by calculating the error between the predicted and true CHI values. Simultaneously, to address inconsistencies among the different indicator dimensions, R^2 was added as an evaluation criterion. The indices are estimated using

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (13)$$

$$MSE = \frac{\sum_{i=1}^N (x_i - y_i)^2}{N} \quad (14)$$

$$MAE = \frac{\sum_{i=1}^N |x_i - y_i|}{N} \quad (15)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (x_i - y_i)^2}{\sum_{i=1}^N (x_i - \bar{x}_i)^2} \quad (16)$$

Table 3. Main parameter settings of proposed network

Parameter	Value
Epochs	1000
Batch_size	128
Optimizer	Adam
Leaky rate	0.01
Learning rate	$[1 \times 10^{-4}, 1 \times 10^{-2}]$
Dropout rate	$[1 \times 10^{-3}, 1 \times 10^{-2}]$
Kernel size	[2, 7]
Number of filters	[8, 128]
Number of BiGRU hidden unit	[8, 128]

BiGRU: Bidirectional gated recurrent unit.

Table 4. Ablation experiment prediction errors

Prediction model	RMSE	MSE	MAE	R^2
TCN	0.415	0.172	0.309	0.82
TCN-GRU	0.341	0.117	0.227	0.878
BiTCN-BiGRU	0.315	0.099	0.194	0.896
SA-BiTCN-BiGRU	0.239	0.057	0.13	0.94
CPO-SA-BiTCN-BiGRU	0.171	0.029	0.119	0.969
Proposed model	0.119	0.014	0.095	0.985

RMSE: Root mean square error; MSE: mean square error; MAE: mean absolute error; TCN: temporal convolutional network; GRU: gated recurrent unit; BiTCN: bidirectional temporal convolutional network; BiGRU: bidirectional gated recurrent unit; SA: self-attention; CPO: crested porcupine optimizer.

where x_i represents the true corrugation CHI value, y_i represents the predicted value by the model, N is the data length of the corrugation CHI, and \bar{x}_i denotes the average of the true value.

3.4 Predictive experimental analysis of rail corrugation

After obtaining the CHI of the corrugation damage according to Section 3.1, the corrugation CHI data with a length of 98 can be expressed as $\{x_1, x_2, \dots, x_{98}\}$. Our study used 75% of the data as the training set and the remainder as the test set; thus, the training set was $\{x_1, x_2, \dots, x_n\}$ and the test set was $\{x_{n+1}, x_{n+2}, \dots, x_{98}\}$. The SA-BiTCN-BiGRU model was trained using the training set, whereas the test set was used to verify the effect of the model in predicting the evolution trend of rail corrugation. Subsequently, according to the input step length t of the prediction model, the single-step sliding window approach was used for forecasting, using the corrugation initial data from the CHI to predict the subsequent evolution of the corrugation. For example, the input of the first sample was $\{x_1, x_2, \dots, x_t\}$, yielding the prediction result of y_{t+1} . Then, the prediction was gradually conducted to obtain the prediction result y_{98} of the last sample. The error between the predicted value of the model and the input CHI was calculated to evaluate the prediction performance of the model. Other variables that may have affected the experimental results were controlled to ensure that the observed changes were caused by the proposed method. Table 3 presents the network model parameters used for predicting the evolution trend of the rail corrugation.

3.4.1 Ablation experiment

A comprehensive quantitative analysis of the structure and function of the proposed network was conducted to highlight the effects of each module on the MICPO-SA-BiTCN-BiGRU network. The results are summarized in Table 4.

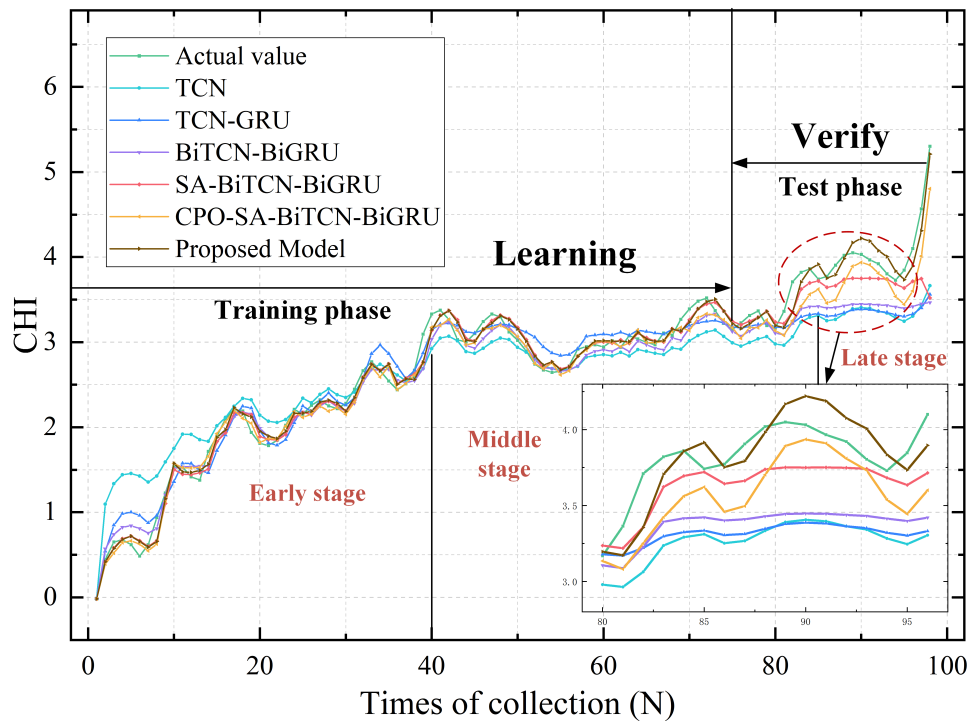


Figure 10. Model ablation experiment: prediction results of evolution trend of rail corrugation.

From Table 4, the RMSE, MSE, and MAE decreased by 17.8%, 32%, and 26.5%, respectively, from TCN to TCN-GRU, whereas R^2 increased by 7.1%. If a bidirectional network structure (BiTCN-BiGRU model) was added, the RMSE, MSE and MAE further decreased by 7.6%, 15.4%, and 14.5%, respectively, whereas R^2 increased by 2.1%. This indicates that the structure further improved its prediction by considering the information on the forward and backward evolution of rail corrugation. When SA was introduced into the BiTCN-BiGRU model, the RMSE, MSE, and MAE decreased by 24.1%, 42.4%, and 33%, respectively, and R^2 increased by 4.91%. This indicates that the introduction of SA improved the feature expression capability of the network and reduced its dependence on irrelevant information. To reduce the impact of the artificial selection of network hyperparameters on the prediction results, the CPO algorithm was added for model optimization. Consequently, the RMSE, MSE, and MAE decreased by 28.5%, 49.1%, and 8.5%, respectively, and R^2 increased by 3.1%. Subsequently, the search strategy of the original CPO algorithm was improved to effectively alleviate the problem of local convergence. Consequently, the RMSE, MSE, and MAE decreased by 30.4%, 51.7%, and 20.2%, respectively, and R^2 increased by 1.7%, indicating that the optimization algorithm and its improved strategy were effective for model prediction.

The visualization results of the model ablation experiment presented in Figure 10 show that the proposed method (brown line) closely matches the true value (green line), particularly during the model testing phase, thus effectively predicting the evolution process of corrugation in the later stages of development. Additionally, the proposed method shows a higher local prediction accuracy compared to the other models.

3.4.2 Comparison experiment

To verify the timeliness of the MICPO-SA-BiTCN-BiGRU network model in predicting the evolution trend of rail corrugation, we used the network models from recently published studies to predict the evolution trend of rail corrugation and quantitatively analyze and compare the predicted results with those of the proposed model. The results are listed in Table 5.

Table 5. Comparison experiment prediction errors

Prediction model	RMSE	MSE	MAE	R ²
TCN-GRU-attention ^[20]	0.351	0.123	0.223	0.872
CNN-BiGRU-attention ^[23]	0.36	0.13	0.276	0.864
CNN-GRU ^[45]	0.448	0.201	0.351	0.79
CNN-LSTM-attention ^[46]	0.377	0.142	0.29	0.852
SA-TCN-LSTM ^[47]	0.295	0.087	0.196	0.909
Proposed model	0.119	0.014	0.095	0.985

RMSE: Root mean square error; MSE: mean square error; MAE: mean absolute error; TCN: temporal convolutional network; GRU: gated recurrent unit; CNN: convolutional neural network; Bi-GRU: bidirectional gated recurrent unit; LSTM: long short-term memory; SA: self-attention.

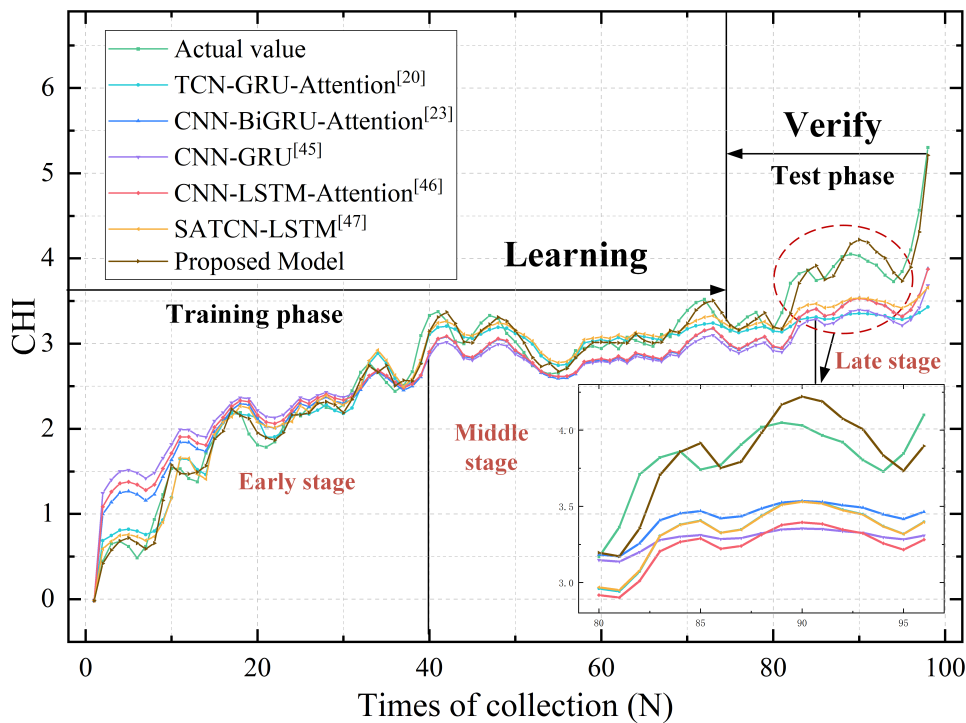


Figure 11. Model comparison experiment: prediction results of evolution trend of rail corrugation.

From [Table 5](#), the proposed model has a lower prediction error than the other models. Consequently, the RMSE decreased by 66.1% to 73.4%, the MSE by 88.6% to 93%, and the MAE by 57.4% to 72.9%. Conversely, the R² increased by 8.36% to 24.7%. This indicates that the MICPO-SA-BiTCN-BiGRU model has a suitable architecture and accurately predicts the evolutionary trend of corrugation.

To discover the evolutionary trend of corrugation more intuitively, a visualization from the comparative experiment is shown in [Figure 11](#).

[Figure 11](#) shows that the development of corrugation damage on the measured road section exists in relatively evident stages. Therefore, we divided the data collected from measurements 1 to 40 into the early stage of rail corrugation evolution, during which the CHI value increased by approximately 3.4, indicating rapid development. The data from measurements 41 to 85 were categorized as the middle stage of rail corrugation development. During this period, the CHI value increased by approximately 0.6, with the development of corrugation leveling off and fluctuations rising slowly. This indicates that the damage caused by corrugation

to the rail began to intensify, and the rail was approaching a critical state. The data from measurements 86 to 98, categorized as the late stage of corrugation development, showed that the CHI value increased by approximately 1.6. During this period, the degree of corrugation damage deterioration showed a sudden increase, indicating a sharp decline in the health of the rail within a short period, thus necessitating prompt measures to curb its development.

Overall, the prediction trends of the models were similar; however, the proposed model was the most accurate for local prediction. In particular, during the early and late phases of corrugation damage, the model effectively captured the evolution trend of rail corrugation damage, with its predicted value closely aligned with the real value.

4. DISCUSSION

Predicting the evolutionary trend of rail corrugation is critical for the safe operation and maintenance of railways. To address the difficulties involved in accurately evaluating the evolution state of corrugation, a method was proposed to predict the evolution trend of corrugation. By analyzing the existing on-site data on rail corrugation, the CHI and SA-BiTCN-BiGRU hybrid network models were constructed to predict the evolution process of corrugation in the time dimension. The results were better than those of existing studies.

However, constructing the corrugated CHI partly relies on manual experience, which is highly subjective and results in limited accuracy and standardization. In future studies, we will attempt to combine multi-source data such as on-site rail corrugation images, vibrations, and profile data to predict the evolution trend of rail corrugation. The proposed method improves the generality and reliability of our study by combining more comprehensive corrugation damage information, ensuring the safe operation of the corresponding railway line.

Further, we recognize the importance of predicting the location and duration of rail corrugations. Yet, the proposed method was not effective in predicting the location of rail corrugation, and the collected dataset made the prediction of duration challenging. In fault prediction and health management, most existing research focuses on predicting the development and evolution of rail corrugation, with significantly few studies addressing its location. Nevertheless, numerous scholars have studied the detection of rail corrugation positions. For instance, Yang *et al.* proposed an intelligent real-time detection method for rail corrugation using machine vision and CNN; Li *et al.* proposed an intelligent detection method for rail corrugation using signal decomposition and the entropy theory^[48,49]. In our future work, we will aim to combine spatial data to predict the location of rail corrugation and detect rail damage promptly. Additionally, we collected annotated data on the timing and duration of rail corrugation, which can assist in predicting its duration. These efforts will significantly improve the depth of our research and represent an important direction for future studies.

5. CONCLUSIONS

In this study, we proposed an intelligent prediction method for the evolutionary trend of rail corrugation based on SA, BiTCN, and BiGRU.

First, a health indicator reflecting the evolution state of the corrugation was obtained using the defined method for constructing corrugation CHI. The experimental results validated the effectiveness of the CHI. Second, we effectively demonstrated the interpretability and predictive ability of the proposed bidirectional hybrid network, SA-BiTCN-BiGRU, through an ablation experiment. Third, by using the MICPO algorithm, the optimal values of the key hyperparameters of the SA-BiTCN-BiGRU model were determined, thereby improving the prediction accuracy of the corrugation evolution trend. The findings demonstrated the high convergence ca-

pabilities of the MICPO algorithm compared to other swarm intelligent optimization algorithms. The ablation experiment strongly verified the positive role of the MICPO algorithm in improving model prediction results. Finally, the results of the model comparison confirmed that the MICPO-SA-BiTCN-BiGRU model is efficient. The proposed method is significant for railway maintenance, as it effectively predicts the future development trend of rail corrugation and provides a scientific basis for railway maintenance decisions.

DECLARATIONS

Acknowledgments

We thank the Editor-in-Chief and all reviewers for their comments.

Authors' contributions

Conducted experimental analysis and manuscript writing: Yang WH

Guided on the overall framework and implementation steps of this research, and proposed a train of thought for the general research objectives: Liu JH, Zhang CF

Guided English writing: He J

Provided technical support: Wang ZM, Jia L

Provided dataset support: Yang WW

Availability of data and materials

The data are available upon request. If needed, please contact the corresponding author by email.

Financial support and sponsorship

This research was funded by the National Key Research and Development Program (Grant No. 2021YFF0501101), the National Natural Science Foundation of China (Grant Nos. 52272347, 62303178), Key Scientific Research Project of the Hunan Provincial Department of Education (Grant No. 22A0391), the Natural Science Foundation of the Hunan Province (Grant No. 2024JJ7132).

Conflicts of interest

Yang WW is affiliated with Zhuzhou Qingyun Electric Locomotive Accessories Factory Co., Ltd., while the other authors have declared that they have no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2024.

REFERENCES

1. Wang Z, Lei Z. Analysis of influence factors of rail corrugation in small radius curve track. *Mech Sci* 2021;12:31-40. DOI
2. Cui X, Li J, Bao P, Yang Z, Ren Z, Xu X. Investigation into the abnormal phenomenon of rail corrugation superposition in small-radius curve section of intercity railway. *Transport Res Rec* 2023;2677:540-55. DOI
3. Wang Z, Lei Z, Zhao Y, Xu Y. Rail corrugation characteristics of cologne egg fastener section in small radius curve. *Shock Vib* 2020;2020:1-12. DOI
4. Wang QA, Huang XY, Wang JF, et al. Concise historic overview of rail corrugation studies: from formation mechanisms to detection methods. *Buildings* 2024;14:968. DOI
5. Jin F, Xiao H, Nadakatti MM, Yue H, Liu W. Field investigation and rapid deterioration analysis of heavy haul corrugation. *Appl Sci* 2021;11:6317. DOI

6. Bai T, Xu J, Wang K, et al. Investigation on the transient rolling contact behaviour of corrugated rail considering material work hardening. *Eng Fail Anal* 2023;153:107575. DOI
7. Wang W, Sun Q, Zhao Z, et al. Novel coil transducer induced thermoacoustic detection of rail internal defects towards intelligent processing. *IEEE Trans Ind Electron* 2024;71:2100-11. DOI
8. Andrade AR, Stow J. Statistical modelling of wear and damage trajectories of railway wheelsets. *Qual Reliab Eng Int* 2016;32:2909-23. DOI
9. Cui XL, Chen GX, Yang HG, Zhang Q, Ouyang H, Zhu MH. Study on rail corrugation of a metro tangential track with Cologne-egg type fasteners. *Int J Veh Mech Mobil* 2016;54:353-69. DOI
10. Hu J, Weng L, Gao Z, Yang B. State of health estimation and remaining useful life prediction of electric vehicles based on real-world driving and charging data. *IEEE Trans Veh Technol* 2023;72:382-94. DOI
11. Zhang Y, Xin Y, Liu Z, Chi M, Ma G. Health status assessment and remaining useful life prediction of aero-engine based on BiGRU and MMoE. *Reliab Eng Syst Safe* 2022;220:108263. DOI
12. Wang Q, Song Y, Zhang X, et al. Evolution of corrosion prediction models for oil and gas pipelines: from empirical-driven to data-driven. *Eng Fail Anal* 2023;146:107097. DOI
13. Ji A, Woo WL, Wong EWL, Quek YT. Rail track condition monitoring: a review on deep learning approaches. *Intell Robot* 2021;1:151-75. DOI
14. Xiao B, Liu J, Zhang Z. A heavy-haul railway corrugation diagnosis method based on WPD-ASTFT and SVM. *Shock Vib* 2022;2022:1-14. DOI
15. He J, Xiao Z, Zhang C. Predicting the remaining useful life of rails based on improved deep spiking residual neural network. *Proc Saf Environ Prot* 2024;188:1106-17. DOI
16. Yang H, He J, Liu Z, Zhang C. LLD-MFCOS: a multiscale anchor-free detector based on label localization distillation for wheelset tread defect detection. *IEEE Trans Instrum Meas* 2024;73:1-15. DOI
17. Wu JY, Wu M, Chen Z, Li XL, Yan R. Degradation-aware remaining useful life prediction with LSTM autoencoder. *IEEE Trans Instrum Meas* 2021;70:1-10. DOI
18. Zheng X, Zhao Y, Peng B, Ge M, Kong Y, Zheng S. Information filtering unit-based long short-term memory network for industrial soft sensor modeling. *IEEE Sens J* 2024;24:13530-44. DOI
19. Galassi A, Lippi M, Torrioni P. Attention in natural language processing. *IEEE Trans Neur Net Learn Syst* 2021;32:4291-308. DOI
20. He Y, Wang W, Li M, Wang Q. A short-term wind power prediction approach based on an improved dung beetle optimizer algorithm, variational modal decomposition, and deep learning. *Comput Electr Eng* 2024;116:109182. DOI
21. Zhang C, Jiang C, Liu J, Yang W, He J. Degradation trend prediction of rail stripping for heavy haul railway based on multi-strategy hybrid improved pelican algorithm. *Intell Robot* 2023;3:647-65. DOI
22. Liu J, Du D, He J, Zhang C. Prediction of remaining useful life of railway tracks based on DMGDCC-GRU hybrid model and transfer learning. *IEEE Trans Veh Technol* 2024;73:7561-75. DOI
23. Xiang L, Yang X, Hu A, Su H, Wang P. Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks. *Appl Energ* 2022;305:117925. DOI
24. Liang H, Cao J, Zhao X. Multi-sensor data fusion and bidirectional-temporal attention convolutional network for remaining useful life prediction of rolling bearing. *Meas Sci Technol* 2023;34:105126. DOI
25. Ye Z, Yu J. Feature extraction of gearbox vibration signals based on multi-channels weighted convolutional neural network. *J Mech Eng* 2021;57:110-20. DOI
26. Qiao FJ, Li B, Gao MQ, Li JJ. ECG signal classification based on adaptive multi-channel weighted neural network. *Neural Netw World* 2022;32:55-72. DOI
27. Li S, Zhang C, Zhang X. A novel spatiotemporal enhanced convolutional autoencoder network for unsupervised health indicator construction. *IEEE Trans Instrum Meas* 2024;73:1-10. DOI
28. Chen L, Xu G, Zhang S, Yan W, Wu Q. Health indicator construction of machinery based on end-to-end trainable convolution recurrent neural networks. *J Manuf Syst* 2020;54:1-11. DOI
29. Lei Y, Li N, Guo L, Li N, Yan T, Lin J. Machinery health prognostics: a systematic review from data acquisition to RUL prediction. *Mech Syst Signal Proc* 2018;104:799-834. DOI
30. Yu X, Deng L, Tang B, Xia Y, Li Q. Gear degradation trend prediction by meta-learning gated recurrent unit networks under few samples. *J Mech Eng* 2022;58:149-56. DOI
31. Jiao L, Chen J, Liu L. Degradation trend prediction of rolling bearings based on CAE and AGRU. *Shock Vib* 2023;42:109-17. Available from: <https://jvs.sjtu.edu.cn/EN/Y2023/V42/I12/109>. [Last accessed on 14 Oct 2024]
32. Sarmadi H, Entezami A, Saeedi Razavi B, Yuen KV. Ensemble learning-based structural health monitoring by Mahalanobis distance metrics. *Struct Control Health Monit* 2020;28:e2663. DOI
33. Wang Y, Deng L, Zheng L, Gao RX. Temporal convolutional network with soft thresholding and attention mechanism for machinery prognostics. *J Manuf Syst* 2021;60:512-26. DOI
34. Li X, Ma X, Xiao F, Xiao C, Wang F, Zhang S. Time-series production forecasting method based on the integration of bidirectional gated recurrent unit (Bi-GRU) network and sparrow search algorithm (SSA). *J Petrol Sci Eng* 2022;208:109309. DOI
35. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. arXiv. [Preprint.] Aug 2, 2023 [accessed on 2024 Oct 14]. Available from: <https://doi.org/10.48550/arXiv.1706.03762>.
36. Abdel-Basset M, Mohamed R, Abouhawwash M. Crested porcupine optimizer: a new nature-inspired metaheuristic. *Knowl Based Syst*

- 2024;284:111257. DOI
37. Ge Z, Feng S, Ma C, Dai X, Wang Y, Ye Z. Urban river ammonia nitrogen prediction model based on improved whale optimization support vector regression mixed synchronous compression wavelet transform. *Chemometr Intell Lab Syst* 2023;240:104930. DOI
 38. Li M, Liu Z, Song H. An improved algorithm optimization algorithm based on RungeKutta and golden sine strategy. *Expert Syst Appl* 2024;247:123262. DOI
 39. Zhai X, Tian J, Li J. A real-time path planning algorithm for mobile robots based on safety distance matrix and adaptive weight adjustment strategy. *Int J Control Autom Syst* 2024;22:1385-99. DOI
 40. Mirjalili S, Lewis A. The whale optimization algorithm. *Adv Eng Softw* 2016;95:51-67. DOI
 41. Ouyang C, Qiu Y, Zhu D. Adaptive spiral flying sparrow search algorithm. *Sci Program* 2021;2021:1-16. DOI
 42. Su H, Zhao D, Heidari AA, et al. RIME: a physics-based optimization. *Neurocomp* 2023;532:183-214. DOI
 43. Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. *Adv Eng Softw* 2014;69:46-61. DOI
 44. Xue J, Shen B. Dung beetle optimizer: a new meta-heuristic algorithm for global optimization. *J Supercomput* 2023;79:7305-36. DOI
 45. Zhao Z, Yun S, Jia L, et al. Hybrid VMD-CNN-GRU-based model for short-term forecasting of wind power considering spatio-temporal features. *Eng Appl Artif Intel* 2023;121:105982. DOI
 46. Xiong B, Lou L, Meng X, Wang X, Ma H, Wang Z. Short-term wind power forecasting based on attention mechanism and deep learning. *Electr Pow Syst Res* 2022;206:107776. DOI
 47. Xiang L, Liu J, Yang X, Hu A, Su H. Ultra-short term wind power prediction applying a novel model named SATCN-LSTM. *Energ Convers Managem* 2022;252:115036. DOI
 48. Yang H, Liu J, Mei G, Yang D, Deng X, Duan C. Research on real-time detection method of rail corrugation based on improved ShuffleNet V2. *Eng Appl Artif Intel* 2023;126:106825. DOI
 49. Li S, Mao X, Shang P, Xu X, Liu J, Qiao P. Intelligent detection of rail corrugation using ACMP-based energy entropy and LSSVM. *Nonlin Dynam* 2023;111:8419-38. DOI