Journal of Computer Science and Frontier Technologies
ISSN Print: 3104-4204
ISSN Online: 3104-4212
CD ISSN: 3104-4220
https://doi.org/10.63313/JCSFT.9061

Remaining Useful Life Prediction of Rolling Bearings Based on SSA-Optimized TCN-CBAM-BiLSTM Model

GuangSen Du. JiHeng Chen and JianWei Cheng
Chinese People's Liberation Army Naval Non commissioned Officer School, Bengbu, Anhui 233000, China

How to cite this paper: Du, G. S., Chen, J. H., & Cheng, J. W. (2026). Remaining Useful Life Prediction of Rolling Bearings Based on SSA-Optimized TCN-CBAM-BiLSTM Model. *Journal of Computer Science and Frontier Technologies*, 3(1), 65-75. ISSN Print: 3104-4204; ISSN Online: 3104-4212.
https://doi.org/10.63313/JCSFT.9061
Published: 2026-04-20

Copyright © 2026 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Abstract

In the field of remaining useful life (RUL) prediction for rolling bearings, existing deep learning models suffer from issues such as insufficient capability to capture long-term temporal dependencies, poor focus on key degradation features, reliance on manual experience for hyperparameter tuning, and weak generalization ability under complex operating conditions. To address these challenges, this paper proposes a RUL prediction method for rolling bearings based on a Sparrow Search Algorithm (SSA)-optimized Temporal Convolutional Network-Convolutional Block Attention Module-Bidirectional Long Short-Term Memory Network (TCN-CBAM-BiLSTM). Firstly, a basic TCN-CBAM-BiLSTM prediction model is constructed. The dilated convolutions and residual connections in TCN are employed to capture long-span degradation correlations throughout the bearing's lifecycle. The CBAM's dual attention mechanism dynamically suppresses noise and enhances key degradation features, while BiLSTM's bidirectional temporal modeling capability fully extracts contextual information from sequences. Building on this foundation, SSA is utilized to adaptively optimize core hyperparameters such as the model's learning rate and the number of TCN convolutional kernels, resulting in the development of a high-precision SSA-TCN-CBAM-BiLSTM prediction model that resolves the subjectivity and inefficiency associated with manual parameter tuning. Multiple comparative experiments are conducted using the XJTU-SY rolling bearing full-lifecycle degradation dataset. The results demonstrate that the proposed model significantly reduces both the root mean square error and mean absolute error, two key metrics. Moreover, its predicted curves closely align with the actual RUL curves under multiple operating conditions, showcasing superior prediction accuracy and cross-condition generalization capability.

Keywords

Rolling bearing; Remaining useful life prediction; Temporal convolutional network; Bidirectional long short-term memory network; Sparrow search algorithm

1. Introduction

In the intelligent transformation of the high-end equipment manufacturing industry, fault prediction and health management (PHM) for rotating machinery represent core technologies for ensuring the safe and efficient operation of equipment. As a critical component of rotating machinery, rolling bearing failure is the primary cause of equipment malfunctions, making accurate remaining useful life (RUL) prediction of rolling bearings of significant engineering value. Among current RUL prediction methods for rolling bearings, physics-based models exhibit poor adaptability to varying operating conditions, while data-driven deep learning approaches have become mainstream. However, basic models such as CNNs and unidirectional LSTMs suffer from deficiencies in capturing long-term temporal dependencies and fully utilizing contextual information. Existing hybrid models also face issues such as high subjectivity in manual hyperparameter tuning and weak generalization capabilities under varying operating conditions. To address these shortcomings, this paper aims to achieve high-precision RUL prediction for rolling bearings by constructing an SSA-TCN-CBAM-BiLSTM prediction model. Multiple experiments conducted on publicly available datasets systematically validate the effectiveness and superiority of the proposed method.

2. Temporal Convolutional Network

The Temporal Convolutional Network (TCN) is a deep learning architecture specifically optimized for time-series data processing scenarios. Its core design logic is based on a one-dimensional convolutional framework, creatively integrating multiple convolutional strategies to enhance functionality. The network's core mechanism involves a profound "temporalization" transformation of traditional one-dimensional convolutions, with targeted optimizations and upgrades built upon the architecture of classical Residual Networks (Res-Nets). By systematically integrating the end-to-end processing capabilities of one-dimensional fully convolutional networks, the temporal causality constraints of causal convolutions, the long-range dependency capture advantages of dilated convolutions, and the feature reuse mechanism of residual connections, TCN forms a comprehensive technical solution tailored for temporal modeling. The main framework of TCN consists of several standardized residual modules stacked in an orderly manner, with each residual module containing two carefully designed dilated causal convolutional layers. By introducing a direct mapping path for input features within the module, TCN significantly reduces the training difficulty of deep networks: on the one hand, it effectively mitigates the vanishing or exploding gradient problem during backpropagation, ensuring stable updates of model parameters; on the other hand, this structural design endows the network with several key advantages, including high parallel computing efficiency, flexible adjustment of receptive field size through dilation rates, more stable gradient propagation paths, and lower

memory consumption, making it outstanding in tasks such as time-series prediction and sequence classification [1].

3. Convolutional Block Attention Module (CBAM - Contextual Explanation within Temporal Modeling Framework)

In the field of temporal modeling, to enhance the deep modeling capabilities of dilated causal convolutional networks, TCN residual blocks incorporate nonlinear mappings into the foundational architecture of dilated causal convolutions while constructing inter-layer residual skip connection structures. This dual optimization design offers significant advantages: On one hand, by leveraging cross-layer information transfer through residual skip connections, TCN can achieve a large receptive field covering long temporal ranges without stacking excessive network layers. This not only streamlines the model architecture but also fully captures distant correlation features in temporal data, substantially improving the modeling efficiency of complex temporal patterns. On the other hand, the cross-layer information pathways formed by inter-layer residual connections enable direct transmission of low-level features to high-level layers, ensuring smooth gradient flow during backpropagation and effectively mitigating gradient anomalies and optimization challenges associated with increased network depth. This mechanism holds particular value in rolling bearing life prediction tasks, as it facilitates more stable and precise learning of subtle degradation features and long-term performance degradation patterns throughout the bearing's lifecycle, providing reliable feature support for subsequent life assessment [2].

The computational process of CBAM can be expressed as:

$$\begin{cases} F' = M_C(F) \times F \\ F'' = M_S(F') \times F' \end{cases} \tag{3.1}$$

Where: F represents the original feature map; F'' denotes the channel attention feature map; F'' is the output feature map of CBAM; $M_{C(F)}$ stands for the channel attention weights; and $M_{C(F')}$ represents the spatial attention weights.

4. Bidirectional Long Short-Term Memory Network

LSTM relies on a gating mechanism composed of an input gate, a forget gate, and an output gate to achieve selective memory and long-term feature storage of sequential data, effectively enhancing sequence prediction accuracy. However, its unidirectional information processing mode has inherent limitations—it can only extract forward temporal features, and early critical information tends to be diluted in long sequences, making it difficult to adapt to tasks such as equipment degradation prediction that heavily rely on full-sequence context [3]. To address this, BiLSTM achieves the fusion and extraction of both forward and backward temporal features through two parallel LSTM units. It fully retains the advantages of the LSTM gating mechanism while incorporating Dropout technology to improve

model generalization ability, thereby overcoming the performance limitations of unidirectional LSTM and demonstrating superior performance in equipment degradation prediction tasks.

The computational process of BiLSTM can be expressed as:

$$\begin{cases} \vec{h}_t = \mathrm{LSTM}(x_t, \vec{h}_{t-1}) \\ \overleftarrow{h}_t = \mathrm{LSTM}(x_t, \overleftarrow{h}_{t-1}) \\ y_t = \sigma(W_y \cdot [\vec{h}_t, \overleftarrow{h}_t] + b_y) \end{cases} \tag{4.1}$$

Among them, W_y and b_y are the weight matrix and bias term respectively; h_t and h_t are the forward and backward hidden layer states at time t.

5. Sparrow Search Algorithm

The Sparrow Search Algorithm (SSA) is a swarm intelligence optimization algorithm that simulates the foraging behavior and anti-predation mechanisms of sparrow populations. In this algorithm, the population is divided into three roles: discoverers, followers, and scouts. The global optimal solution is searched for by iteratively updating the positions of the population [4]. Discoverers are responsible for providing foraging directions for the population, and their position update formula is as follows:

$$X_{i,j}^{i+1} = \begin{cases} X_{i,j}^{i} \cdot exp\left(-\frac{i}{\alpha \cdot T_{max}}\right), & R_2 < ST \\ X_{i,j}^{i} + Q \cdot L, & R_2 \geq ST \end{cases} \tag{5.1}$$

where, T_{max} is the maximum number of iterations, α is a random number within the range (0,1], R_2 is the warning value, ST is the safety threshold, Q is a random number following a normal distribution, and L is the identity matrix. Followers update their positions based on the positions of the discoverers, while scouts randomly wander at the edges of the population to evade predators. SSA is suitable for model hyperparameter optimization problems due to its advantages of fast convergence speed, high optimization accuracy, and simple parameter settings.

6. SSA-Optimized TCN-CBAM-BiLSTM Prediction

6.1. Construction of the TCN-CBAM-BiLSTM Base Model

After completing the aforementioned univariate-controlled hyperparameter optimization experiments, all core training hyperparameters and network architecture parameters for the proposed TCN-CBAM-BiLSTM model in this paper have been calibrated to their optimal values. Simultaneously, the complete cascaded architecture of the model has been finalized: the main body consists of a 4-layer TCN feature extraction network, a 3-layer BiLSTM bidirectional temporal dependency mining network, and a 3-layer fully connected network. Among them, the fully connected layers serve as the RUL prediction output layer of the model. The complete structure of the finally optimized TCN-CBAM-BiLSTM model is illustrated in Figure 4-15.

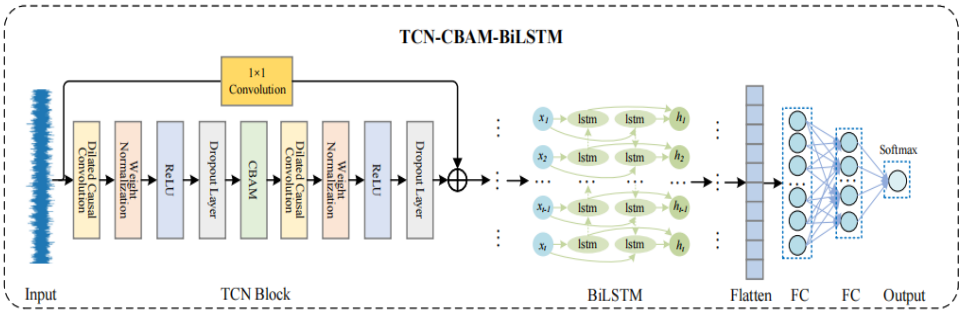


Figure 5-1. Structure of the TCN-CBAM-BiLSTM model

The TCN-CBAM-BiLSTM hybrid deep learning model proposed in this paper integrates the core mechanisms of temporal convolution, adaptive attention, and bidirectional recurrent neural networks. It is well-suited for the nonlinear, non-stationary characteristics and long-range degradation dependencies inherent in vibration time-series data throughout the entire lifecycle of rolling bearings, demonstrating significant performance advantages in complex time-series prediction tasks. The model employs a hierarchical and progressive feature processing logic, constructing an end-to-end complete prediction pipeline through the organic cascading of TCN, CBAM, and BiLSTM. This pipeline establishes a pathway for "basic feature extraction - intelligent screening of key features - global modeling of degradation patterns," providing a high-precision solution for rolling bearing RUL prediction.

The TCN module is responsible for basic feature extraction from input time-series data: it strictly adheres to temporal causality constraints through causal convolutions, aligning with the irreversible nature of bearing degradation processes. By incorporating dilated convolution layers with varying dilation rates, it achieves exponential expansion of the receptive field without increasing network parameters or computational complexity, accurately capturing local detailed features and long-span temporal correlations of bearing degradation. The residual connection structure embedded within the module ensures stable gradient signal propagation during deep network backpropagation, effectively mitigating the vanishing gradient problem and overcoming inherent limitations of traditional recurrent neural networks.

The CBAM module is embedded in the model pipeline with its core function of adaptive feature screening, specifically addressing the industry pain points of weak early degradation features and strong noise interference in bearing vibration data. Through the serial collaboration of channel attention and spatial attention branches, it refines and optimizes the features output by TCN. The channel attention branch adaptively computes weight coefficients for each feature channel, enhancing sensitive feature channels highly correlated with bearing degradation while suppressing noisy and redundant ineffective channels. The spatial attention branch

precisely focuses on critical time points of bearing degradation and temporal segments with significant features. This dual attention mechanism significantly enhances the pertinence and discriminative power of feature representation without incurring substantial additional computational overhead.

The BiLSTM module is responsible for global temporal modeling of bearing degradation evolution patterns, comprising two parallel LSTM units: forward and backward. The forward LSTM learns the historical degradation trajectory of the bearing along the positive time axis, modeling the progressive evolution pattern of bearing wear. The backward LSTM explores temporal dependency correlations in the reverse direction along the time axis, identifying the intrinsic mapping relationship between early weak degradation features and later fault evolution. The output features from both branches undergo concatenation for deep fusion, forming a global feature vector covering full-temporal context information. This compensates for the deficiency of convolutional networks in capturing long-range temporal dependencies and provides a complete global temporal perspective for accurate RUL prediction throughout the bearing's entire lifecycle.

6.2. SSA Hyperparameter Adaptive Optimization Design

The implementation process of SSA optimizing the TCN-CBAM-BiLSTM model is primarily divided into four core stages: population initialization, fitness calculation, population position update, and termination condition judgment. Initially, during the population initialization phase, algorithm parameters such as population size, maximum number of iterations, and the ratio of discoverers to scouts are set. Population positions are randomly initialized within the hyperparameter boundary ranges, so that each individual corresponds to a set of hyperparameter combinations to be optimized. Subsequently, in the fitness calculation stage, each set of hyperparameters is substituted into the TCN-CBAM-BiLSTM model for training, and the MAE on the validation set is used as the fitness value. A smaller value indicates better prediction performance of the hyperparameter combination. Next, population position updates are conducted. Individuals are sorted based on fitness values, and the positions of discoverers, followers, and scouts are updated in sequence. Meanwhile, a dynamic weight factor is introduced to balance the algorithm's global search and local exploitation capabilities. Finally, termination condition judgment is performed. When the number of iterations reaches the preset maximum or the fitness value does not decrease for multiple consecutive rounds, the optimization process is terminated, and the globally optimal hyperparameter combination is output.

6.3. Evaluation Metrics

When evaluating the prediction performance of bearing RUL in this paper, two key evaluation metrics widely recognized and maturely applied in the field of RUL prediction—Mean Absolute Error (MAE) and Root Mean Square Error (RMSE)—are

selected [5].

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i}^{N}(y_i - \hat{y}_i)^2} \tag{6.1}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{6.2}$$

Where: $y_i$represents the true remaining useful life (RUL) value of the i-th sample; N is the number of samples; $\hat{y}_i$ is the predicted RUL value of the i-th sample.

MAE directly reflects the average deviation between the model's prediction results and actual operating conditions by calculating the average of the absolute differences between "predicted RUL values and true RUL values" across all samples. Its core advantage lies in its relatively low sensitivity to extreme errors, allowing it to stably represent the model's overall prediction performance. RMSE, on the other hand, first squares the differences between "predicted RUL values and true RUL values," takes the average, and then calculates the square root of the result. This calculation logic makes it more sensitive to larger deviations, thereby enabling a more precise quantification of the model's prediction accuracy during critical degradation stages of bearings.

6.4. XJTU-SY Rolling Bearing Dataset

The XJTU-SY public rolling bearing dataset used in this experiment was jointly developed by Lei Yaguo's team from Xi'an Jiaotong University and Zhejiang Changxing Shengyang Technology [6]. It covers full-lifecycle degradation data of bearings under multiple operating conditions. The experiment utilized LDKUER204-type bearings, with data collected via horizontal and vertical acceleration sensors at 1-minute intervals, a sampling frequency of 25.6 kHz, and a single sampling duration of 1.28 seconds. Details of the dataset are provided in Table 6-1.

Table 6-1. XJTU-SY Rolling Bearing Dataset

Operating conditions of the dataset	Operating condition 1	Operating condition 2	Operating condition 3
Rotational speed / (r/min)	2100	2250	2400
Radial force / (kN)	12	11	10
Bearing number	B1-1	B2-1	B3-1
	B1-2	B2-2	B3-2
	B1-3	B2-3	B3-3
	B1-4	B2-4	B3-4
	B1-5	B2-5	B3-5

6.5. Analysis of Prediction Results

To verify the cross-operating-condition generalization capability of the SSA-optimized TCN-CBAM-BiLSTM (hereinafter referred to as SSA-TCN-CBAM-BiLSTM), this chapter utilizes test set samples from Operating Condition 1, Operating Condition 2, and Operating Condition 3 for RUL prediction experiments. The changes in the loss function during the training process of the model under the three operating conditions are shown in Figure 5-2 to Figure 5-4.

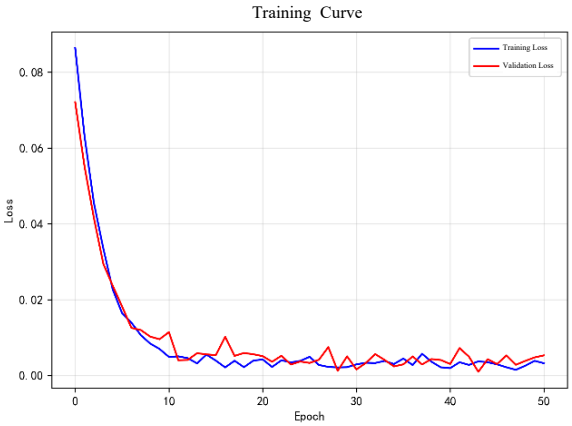


Figure 5-2. Loss Curve under Operating Condition 1

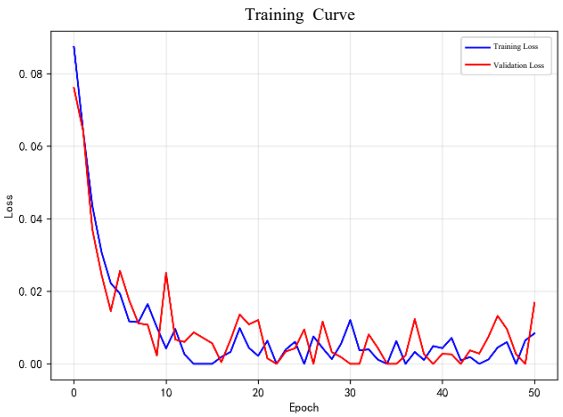


Figure 5-3. Loss Curve under Operating Condition 2

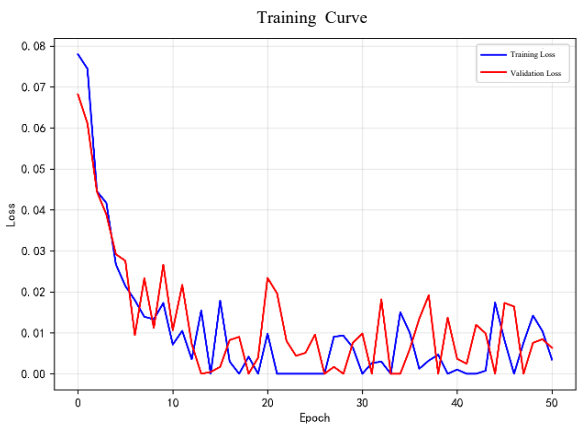


Figure 5-4. Loss Curve under Operating Condition 3

As can be seen from the loss curves, the training loss and validation loss for all three operating conditions rapidly decline within the first 10 training epochs, subsequently gradually leveling off and maintaining at relatively low levels. The loss values converge to approximately 0.002 for Operating Condition 1, around 0.010 for Operating Condition 2, and similarly around 0.010 for Operating Condition 3, indicating that the model achieves stable fitting and is in good training condition

under each operating condition.

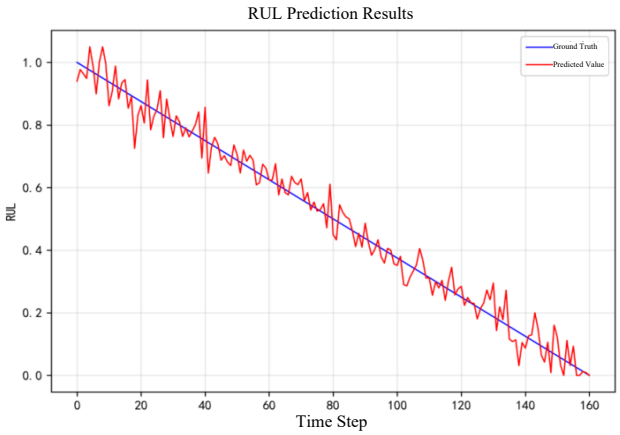


Figure 5-5. RUL Prediction Curve under Operating Condition 1

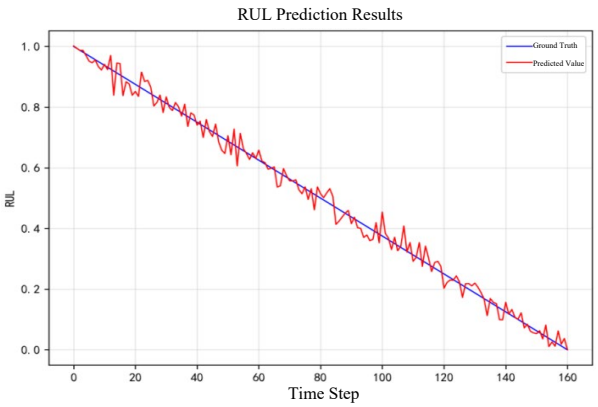


Figure 5-6. RUL Prediction Curve under Operating Condition 2

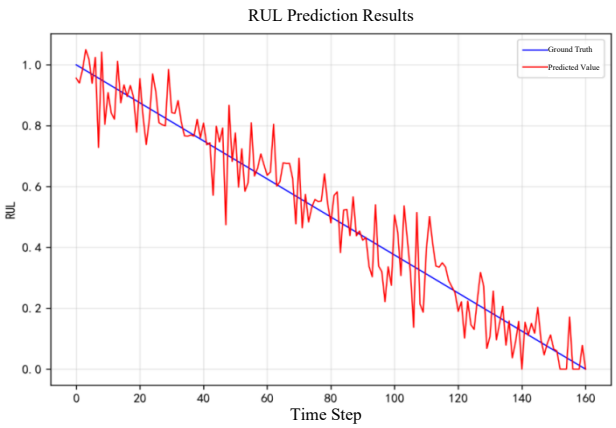


Figure 5-7. RUL Prediction Curve under Operating Condition 3

The RUL prediction results of the SSA-TCN-CBAM-BiLSTM model for the test set samples under the three operating conditions are shown in Figures 5-5 to 5-7. In these figures, the blue curve represents the true degradation curve of the bearing's remaining useful life, while the red curve represents the predicted curve output by

the model. The x-axis denotes the time step, and the y-axis represents the RUL value. Judging from the prediction results under the three operating conditions, the model effectively captures the RUL variation trend throughout the bearing's full lifecycle.

For the prediction results under Operating Condition 1, the red predicted curve and the blue true curve maintain a high degree of consistency throughout the entire lifecycle, with the two curves fitting closely together and only exhibiting slight differences at minor local fluctuations. This indicates that the model possesses excellent fitting accuracy under this operating condition. Regarding the prediction results for Operating Condition 2, the predicted curve experiences minor fluctuations in the mid-stage, resulting in a temporary deviation from the true curve. However, the overall trend remains consistent, and it swiftly returns to near the true values in the later stage, demonstrating that the model has certain adaptive capabilities to changes in operating conditions. For the prediction results under Operating Condition 3, the red predicted curve and the blue true curve remain highly consistent from the initial stage to the failure stage, with the two curves almost overlapping. The prediction effect is the most outstanding among the three, with the maximum deviation controlled within an extremely small range. The aforementioned results fully demonstrate that the SSA-optimized model can maintain stable and high-precision prediction performance under three different operating conditions, exhibiting excellent cross-operating-condition generalization capabilities.

7. Summary

This paper addresses the issues of manual experience dependence for hyperparameter tuning and insufficient cross-operating-condition generalization capability in the basic TCN-CBAM-BiLSTM model. It introduces the Sparrow Search Algorithm (SSA) for adaptive optimization of the model's core hyperparameters, constructing an SSA-TCN-CBAM-BiLSTM model for rolling bearing Remaining Useful Life (RUL) prediction. The model's validation is completed based on the XJTU-SY multi-operating-condition bearing dataset. The study first introduces the test rig, sampling parameters, and data selection scheme of the dataset, integrating bidirectional vibration signals to extract degradation features. It then elaborates on the fundamental principles of SSA, establishing the minimization of the Mean Absolute Error (MAE) on the validation set as the optimization objective. Core hyperparameters such as the learning rate and the number of TCN convolutional kernels are selected for optimization, and the implementation process of the SSA-optimized model is designed, along with the specification of experimental hardware, software, and basic parameters. Experimental results demonstrate that the optimized model exhibits rapid loss convergence and a high degree of alignment between the predicted and true curves across multiple operating conditions. The "fixed core + fine-tuned adaptation" parameter combination obtained through

optimization adapts to the operational characteristics of different conditions. In horizontal comparison experiments, the model significantly outperforms other comparative models in terms of Root Mean Square Error (RMSE) and MAE metrics, achieving an RMSE of 0.0268 and an MAE of 0.025 on the B1 sample, indicating the highest prediction accuracy. This study achieves adaptive optimization of model hyperparameters through SSA, resolving the subjectivity associated with manual parameter setting and enhancing the model's cross-operating-condition generalization capability. It provides practical technical support for the engineering application of rolling bearing RUL prediction.

References

[1] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015: 448-456.

[2] Tibshirani R. Regression Shrinkage and Selection via the Lasso[J]. Journal of the Royal Statistical Society: Series B (Methodological), 1996, 58(1): 267-288.

[3] Kingma D P, Ba J. Adam: A Method for Stochastic Optimization[C]//Proceedings of the International Conference on Learning Representations (ICLR). San Diego, CA, USA: ICLR, 2015: 1-15.

[4] NECTOUX P, GOUIVEAU R, MEDJAHER K, et al. PRONOSTIA: An experimental platform for bearings accelerated degradation tests[C]//Proceedings of the IEEE International Conference on Prognostics and Health Management (PHM 2012). Denver, USA: IEEE, 2012: 1-8.

[5] HAO X L, LIU Y, PEI L L, et al. Atmospheric temperature prediction based on a BiLSTM-attention model[J]. Symmetry, 2022, 14(11): 2470.

[6] LEA C, VIDAL R, REITER A, et al. Temporal convolutional networks: a unified approach to action segmentation[C]//Proceedings of Computer Vision - ECCV 2016 Workshops. Amsterdam: Springer International Publishing, 2016: 47-54.