

Article

Rolling Bearing Residual Useful Life Prediction Model Based on the Particle Swarm Optimization-Optimized Fusion of Convolutional Neural Network and Bidirectional Long–Short-Term Memory–Multihead Self-Attention

Jianzhong Yang ¹, Xinggong Zhang ², Song Liu ², Ximing Yang ² and Shangfang Li ^{3,*}

¹ College of Electronic and Information Engineer, Beibu Gulf University, Qinzhou 535011, China; gxyjz@bbgu.edu.cn

² College of Naval Architecture and Ocean Engineering, Beibu Gulf University, Qinzhou 535011, China; 2102011146@stu.bbgu.edu.cn (X.Z.); 2302010117@stu.bbgu.edu.cn (S.L.); 2202011143@stu.bbgu.edu.cn (X.Y.)

³ College of Mathematics and Statistics, Yulin Normal University, Yulin 537000, China

* Correspondence: monklsf@ylu.edu.cn

Abstract: In the context of predicting the remaining useful life (RUL) of rolling bearings, many models often encounter challenges in identifying the starting point of the degradation stage, and the accuracy of predictions is not high. Accordingly, this paper proposes a technique that utilizes particle swarm optimization (PSO) in combination with the fusing of a one-dimensional convolutional neural network (CNN) and a multihead self-attention (MHSA) bidirectional long short-term memory (BiLSTM) network called PSO-CNN-BiLSTM-MHSA. Initially, the original signals undergo correlation signal processing to calculate the features, such as standard deviation, variance, and kurtosis, to help identify the beginning location of the rolling bearing degradation stage. A new dataset is constructed with similar degradation trend features. Subsequently, the particle swarm optimization (PSO) algorithm is employed to find the optimal values of important hyperparameters in the model. Then, a convolutional neural network (CNN) is utilized to extract the deterioration features of rolling bearings in order to predict their remaining lifespan. The degradation features are inputted into the BiLSTM-MHSA network to facilitate the learning process and estimate the remaining lifespan of rolling bearings. Finally, the degradation features are converted to the remaining usable life (RUL) via the fully connected layer. The XJTU-SY rolling bearing accelerated life experimental dataset was used to verify the effectiveness of the proposed method by k-fold cross-validation. After comparing our model to the CNN-LSTM network model and other models, we found that our model can achieve reductions in mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) of 9.27%, 6.76%, and 2.35%, respectively. Therefore, the experimental results demonstrate the model's accuracy in forecasting remaining lifetime and support its ability to forecast breakdowns.

Keywords: rolling bearing; residual service life prediction; particle swarm optimization; bidirectional long short-term memory network; convolutional neural network; multi-head self-attention mechanism



Citation: Yang, J.; Zhang, X.; Liu, S.; Yang, X.; Li, S. Rolling Bearing Residual Useful Life Prediction Model Based on the Particle Swarm Optimization-Optimized Fusion of Convolutional Neural Network and Bidirectional Long–Short-Term Memory–Multihead Self-Attention. *Electronics* **2024**, *13*, 2120. <https://doi.org/10.3390/electronics13112120>

Academic Editors: Séjir Khojet El Khil, Chiara Boccaletti and Monia Ben Khader Bouzid

Received: 31 March 2024

Revised: 23 May 2024

Accepted: 24 May 2024

Published: 29 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In contemporary industrial production systems, several parts cooperate to accomplish a task [1,2]. As a basic component in complex industrial equipment, rolling bearings significantly impact the functionality of the whole mechanical system [3]. From automobile wheels to aircraft engines, from high-speed trains to precision machine tools, the normal operation of these bearings is directly related to the performance and safety of all equipment [4,5]. However, as a result of being subjected to complex mechanical and environmental loads over long periods of time, rolling bearings inevitably experience

wear, fatigue, and performance degradation, which ultimately leads to failure [6]. Rolling bearings usually operate in harsh and variable operating environments. This invariably increases the difficulty of predicting remaining useful life (RUL) in applications related to engineering [7]. Even while rolling bearings are utilized in the same working conditions as product components, they might have very different useful lives [8]. In addition to ensuring the effective and seamless operation of mechanical equipment, evaluating rolling bearing performance can help to identify and promptly eliminate unexpected failure events [9]. With the rise of Industry 4.0 and smart manufacturing, the instantaneous observation and predictive maintenance of the operating status of machinery and equipment have become increasingly important [10]. In this context, the application of rolling bearing residual life prediction technology is particularly important, as is research for its development [11]. An evaluation of the health of bearings and an estimate of their remaining life can be obtained by gathering and analyzing data on a variety of parameters during real-time bearing operation, such as vibration, noise, temperature, load, and rotational speed [12]. This can be done by using advanced signal processing techniques, machine learning methods, and physical models. Therefore, real-time monitoring based on sensor signals and evaluations of their performance degradation are extremely important [13].

Many techniques based on time series prediction and vibration signal feature extraction have been put forth by academics in recent years, and they have been effectively used to predict bearings' RUL. The advantage of not having to consider the structure's mechanics, operating conditions, and failure modes significantly increases the accuracy of remaining life predictions. In an effort to address the drawbacks of conventional machine learning models that necessitate extensive training and labeling, Zou et al. [14] proposed a multi-domain adversarial network-based technique that can be used to predict a rolling bearing's RUL under various operating situations. Hu et al. [15] proposed the use of wavelet analysis and an improved CNN to achieve smart and small sample gear fault detection. Zhou et al. [16] proposed a small, one-wave neural network to predict the remaining service life of bearings. This method adopts manual feature extraction, which is complicated in modeling, and, consequently, a few of the features that are manually extracted can be easily ignored. Song et al. [17] proposed the use of a multi-scale attention residual network for predicting rolling bearings' RUL. Qi et al. [18] researched anomaly detection- and multistep estimation-based techniques for the prediction of the RUL of rolling bearings, which increased the accuracy of RUL prediction to some extent.

Due to its own multi-level internal structure and the benefits of re-learning training methods, deep learning has been able to address the inadequacies of manual feature extraction more effectively in recent years. Kong et al. [19] proposed a multi-level internal structure and re-learning training methods to address the inadequacies of manual feature extraction more effectively. On the other hand, since one of the obvious features of RUL prediction is the time correlation, traditional methods cannot extract the time correlation feature of time series data. In response to this problem, LSTM, a variant of the recurrent neural network (RNN) structure, has emerged, and its effect is remarkable in various fields. To address the limitations in fully utilizing the time-dependent characteristics of the initial signal and the high cost of parameter tuning, Song et al. [20] proposed an optimized CNN with BiLSTM to further enhance the performance of fault diagnostics. Wang et al. [21] proposed the use of multi-feature fusion based on Pearson correlation-KPCA for the prediction of the RUL of rolling bearings. Zhong et al. [22] proposed a technique utilizing an autoformer and ECA-CAE to forecast rolling bearings' RUL.

Predictive accuracy can be increased by utilizing a CNN model's advantage in deep feature extraction and exploiting the efficiency of conventional handiwork feature extraction [23]. Shang et al. [24] studied bidirectional gate recurrent unit-based rolling element bearing remaining life prediction and CNNs. This led to an improvement in rolling bearings' ability to accurately anticipate their remaining lifespans. Therefore, the combination of CNN and LSTM to establish a bearing life prediction model led to superior performance. On the other hand, the advantage of LSTM in processing time series has been exploited

to increase the precision of rolling bearings' remaining life prediction. Jiang et al. [25] proposed a technique that uses attention mechanisms and multi-scale feature extraction to forecast how long rolling bearings will last. Yao et al. [26] improved 1D-CNN with simple recurrent units for the prediction of the RUL of roller bearings by utilizing the attention mechanism. For some time now, RUL prediction has relied on a single CNN or LSTM, and the effect of the addition of other elements has attracted the attention of a growing number of researchers. Obviously, these methods are unable to capture mutations, resulting in insufficient precision in forecasting. Therefore, in an effort to enhance CNNs or LSTM algorithms, several researchers have implemented an attention mechanism. For instance, Zhang et al. [27] offered a new end-to-end RUL prediction algorithm and proposed an architecture for predicting a rolling bearing's estimated remaining lifespan utilizing a convolutional recursive attention network. They used manual feature refinement, which meant they were limited to extracting basic features. But, certain deep or complex traits are rarely or easily extracted. An LSTM neural network and the Laplace technique were combined by Mohd Saufi et al. [28] to precisely predict the RUL of mechanical components. The resultant model has shown significant improvements in reducing prediction times while meeting certain prediction accuracy criteria. However, in more complex working environments and in today's rapidly evolving world, the precision and accuracy of these methods often need to be improved to meet real-world engineering requirements.

In this work, a process for estimating rolling bearings' remaining lifespan is proposed. The proposed approach involves merging the PSO optimization algorithm, 1D-CNN combined with group convolution (group convolution), BiLSTM, and the injected MHSA mechanism. For this study, preprocessing tasks such as normalization, sliding average filtering, and the precise selection of SPT points were carried out prior to feeding information into the neural network, thus suppressing the noise of the raw data to some extent. This research reconstructed the preprocessed data to obtain a better dataset. Then, the advantages of the 1D-CNN model in deep feature extraction were exploited to improve the efficacy and predictability of conventional manual feature extraction. Finally, a BiLSTM network with an MHSA mechanism was applied to forecast rolling bearings' RUL. Within acceptable prediction timescales, the RUL prediction accuracy and precision improvement values achieved in this study are more pronounced when contrasted with the cutting-edge and traditional techniques employed in previous studies. The proposed model could help to shorten computation and prediction times and improve the accuracy of RUL predictions. Thus, its practical application could give engineering projects a competitive advantage.

The following is a summary of the contributions made by this paper:

- (1) The rolling bearing residual life prediction model introduces a starting prediction point (SPT) to identify structural components during the preliminary stage. Consequently, the quantity of data needed to train the model during the rolling bearing health life data initialization phase is decreased. As a result, forecasting takes less time, meaning that projects can gain a competitive edge and productivity can increase. Through the application of the proposed model, we hope to achieve a precise understanding of the health of rolling bearings and optimize the implementation of preventive maintenance strategies in order to greatly extend the service life of equipment and increase operational effectiveness.
- (2) In this study, the PSO optimization algorithm was combined with the PSO optimization algorithm to find the global optimum of the model's important hyperparameters and obtain the most reasonable network parameters. This not only increased prediction accuracy but also decreased the time required to manually adjust the parameters.
- (3) Combining signal processing with deep learning feature extraction methods while overcoming noise allowed for degraded features to be more fully extracted. In the prediction section, rather than just adding more network layers, an MHSA structure was added to enhance the weight of key information and reduce unnecessary parts in order to increase prediction accuracy.

This paper is arranged as follows: Section 1 introduces the benefits of PSO algorithms, CNNs, BiLSTM, and MHSA mechanisms. Section 2 describes the model's construction and the corresponding assessment indicators. A description of how this study obtained unprocessed data, subsequently preprocessed these data, and carried out the extraction of relevant features is presented in Section 3. In Section 4, this study explains how we performed experiments in order to verify the method's better impact by contrasting it with both traditional and cutting-edge techniques. Section 5 concludes the paper.

2. Basic Theory

2.1. CNN

The concept of CNNs is abstract, and their structure is similar to the neuronal organization of the human body, which is a kind of feed-forward neural network with a deep structure. CNNs are widely used in the processing of vibration signals and image information, as well as in text time evaluations; they have the ability to convert the original bearing vibration signals to ensure that they follow the abstract, deeper expression. This means that CNNs can continuously and accurately extract the input data's features through the convolution operation again and again, making it more representative, as a way to facilitate the subsequent diagnostic and prediction work [29]. In this study, RUL prediction was performed for one-dimensional vibration data; therefore, a convolutional neural network in one dimension, i.e., 1D-CNN, needed to be selected [30]. Convolutional neural networks have the following advantages over traditional machine learning methods:

Automatic learning of features: convolutional neural networks can automatically take characteristics out of the input data through convolutional layers, reducing the effort required to extract features manually.

Robustness: convolutional neural networks have translation invariance and partial translation invariance and are robust to small changes in the input data.

Scalability: convolutional neural networks can increase the expressive power of the model by increasing the depth of the network, expanding the convolutional kernel count, and increasing the size of the convolutional kernels to cope with more complex tasks.

The CNN convolutional block structure applied in this study is shown in Figure 1 below.

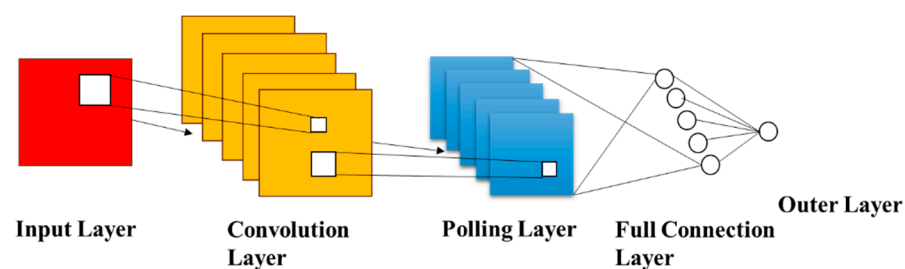


Figure 1. Structure of the CNN model.

2.2. PSO

The population-based optimization method called particle swarm optimization (PSO) finds optimal solutions by simulating the mannerisms of fish schools and flocks of birds [31]. The process of selecting the optimal parameters for applying the PSO algorithm to the CNN-BiLSTM-attention network was as follows:

- (1) First, a swarm of particles (the particle swarm), each representing a potential solution, was initialized. For the 1D-CNN-BiLSTM-MHSA network, the position and velocity of each particle could correspond to various parameters in the network, such as the kernel size of the convolutional layer, the step size of the pooling layer, and the number of hidden layer units in the BiLSTM layer.
- (2) Assessing adaptation: for each particle, the performance (fitness) of its corresponding CNN-MHSA network was calculated. This could be carried out by training the network on a training set and evaluating its performance on a test set.

- (3) Updating the particle velocity and position: each particle's position and velocity were updated based on the PSO algorithm's equation. This process simulated how individuals in a flock of birds adjust their behavior through their own experience and the experience of the group.
- (4) Iterative optimization: steps 2 and 3 were repeated until the termination conditions were met.
- (5) Deciding on the best course of action: upon completion of the iteration, the parameter configuration represented by the most adapted particle was considered the optimal solution.

Since PSO is a heuristic optimization algorithm, its results depend on elements like the nature of the issue, the initialization settings, and the number of iterations. Therefore, when using PSO to optimize the parameters of 1D-CNN-BiLSTM-MHSA networks, several trials and adjustments may be required to obtain satisfactory results.

2.3. BiLSTM

The primary method of RUL prediction is to use mechanical equipment sensor data to determine the spatial-temporal information that deteriorates with equipment performance, which can be fully utilized by an RNN for RUL prediction [32]. LSTM compares the memorized information with the current information, selects the important information, and forgets the secondary information so that the network obtains a stronger memory capability. LSTM can lessen the issue of gradient expansion and disappearance during long-sequence RNN training. Figure 2 depicts the network topology of LSTM.

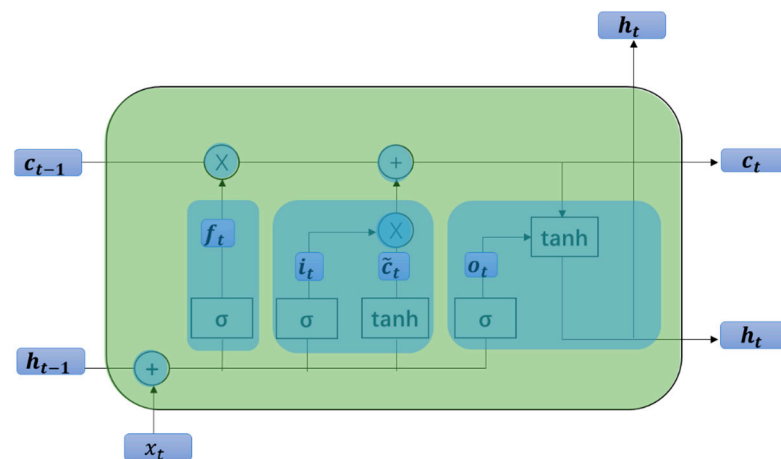


Figure 2. LSTM network structure diagram.

In the figure above, h_{t-1} and h_t are the states that are hidden at moments $t - 1$ and t , respectively; c_{t-1} and c_t are the state of the gating unit at moments $t - 1$ and t , respectively; and x_t is the input at time t . The current gating unit state c_t is jointly influenced by the gating unit state at the previous moment, while the output gate o_t of the memory unit and the gating unit state c are jointly determined. The specific calculation process is shown below:

$$\begin{cases} f_t = \sigma(W_f[x_t, h_{t-1}] + b_f) \\ i_t = \sigma(W_i[x_t, h_{t-1}] + b_i) \\ g_t = \tanh(W_g[x_t, h_{t-1}] + b_g) \\ o_t = \sigma(W_o[x_t, h_{t-1}] + b_o) \\ \tilde{c}_t = f_t \times c_{t-1} + i_t \times g_t \\ h_t = o_t \times \tanh(c_t) \end{cases} \quad (1)$$

where w_f and b_f , w_i and b_i , w_g and b_g , and w_o and b_o are the weight and bias matrices of the oblivion gate, input gate, selection gate, and output hill gate, respectively; $\sigma(\cdot)$ is

the activation function of the sigmoid that transforms the output into the $[0, 1]$ interval; and $\tanh(\cdot)$ is the hyperbolic tangent activation number that transforms the output into the interval $[-1, 1]$. Equation (2) shows the relevant mathematical formula, assuming that A reflects the hidden layer state of the forward LSTM network at a given moment. It is computed by calculating state A_t at time t from state A_{t-1} at time $t - 1$, where x_t is the input at time t .

The components are as follows:

A_t —hidden layer state of the forward LSTM network at time t ;

LSTM—LSTM unit;

x_t —input at moment t ;

A_{t-1} —hidden layer state of the state-positive LSTM network at moment $t - 1$.

In a similar vein, B_t is the inverse LSTM network's hidden layer state at t . Equation (3) shows the computational equation:

$$A_t = LSTM(x_t, A_{t-1}) \quad (2)$$

$$B_t = LSTM(x_t, B_{t-1}) \quad (3)$$

where

B_t —hidden layer state of the inverse LSTM network at time t ;

LSTM—LSTM unit;

x_t —input at moment t ;

B_{t-1} —hidden layer state of the inverse LSTM network at time t .

The BiLSTM network output is a combination of two parts of the hidden layer state A_t and B_t , thereby making up the neural network's total latent state $\overset{\leftrightarrow}{A}_t$.

Information can be coordinated and conveyed by combining certain memory units to form a forward propagation chain structure [33]. The memory unit can accurately manage the information flow based on Equation (1). Gradient descent stability is ensured during model training by combining all moments of internal state information and input information.

The memory unit is the core of the LSTM network; memory units form a forward propagation chain structure. The memory unit can realize the coordination and transmission of information. Equation (1) illustrates how, by combining the internal state information and input information at all times during the model training process, the memory unit can precisely control the flow of information and guarantee gradient descent stability to increase the degree of feature extraction from the initial time series and boost the model's output accuracy. A BiLSTM network is composed of two separate LSTMs layered in distinct orientations. The particular structure is displayed in Figure 3 [34].

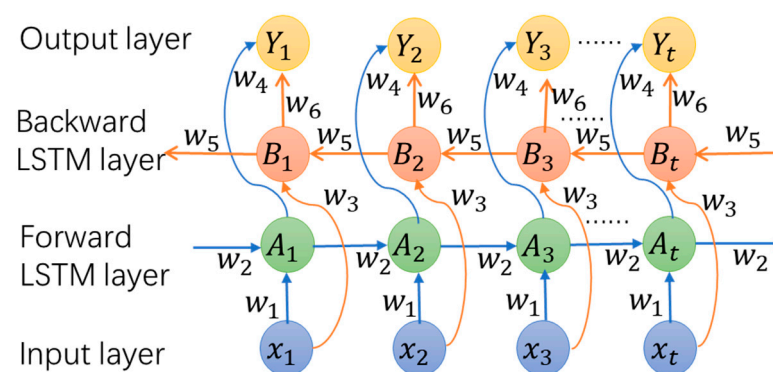


Figure 3. BiLSTM network structure diagram.

Output A_t of the forward implicit layer is computed from instant 0 to moment t , given the input of x_t . The output of the backward implicit layer B_t is calculated by reversing the

input to the reverse layer from instant t to moment 0. To obtain the ultimate output Y , the forward and reverse layers' outputs are fed into the fully connected layer:

$$Y = Y(A_t, B_t) \quad (4)$$

where the fully linked layer's mapping functions are represented by $Y(\cdot)$.

2.4. Attention Mechanism

For images, attention refers to the area that people see as the center of attention in an image, i.e., the part of the image that is the focus of information. For sequences, the attention mechanism is essentially designed to find the interrelationships between the different tokens in the input, finding the relationship between the front and back parts spontaneously through a weight matrix [35]. Self-attention is a type of attention [36]. Self-attention is computed for each Q with each K in turn for the attention coefficients. Entering data for each position in the sequence allows one to focus on information about other positions. From this, the attention score is used to extract features or capture the relationship between each token within the input sequence. Sequence data can therefore be processed. See Figure 4.

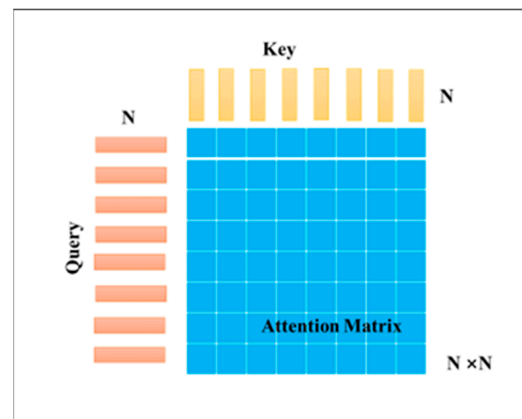


Figure 4. The fundamental concept underlying the self-attention mechanism.

The two most common types of self-attention are dot-product attention and additive attention. As shown in Figure 5, the former is more computationally efficient, so it was used in this study.

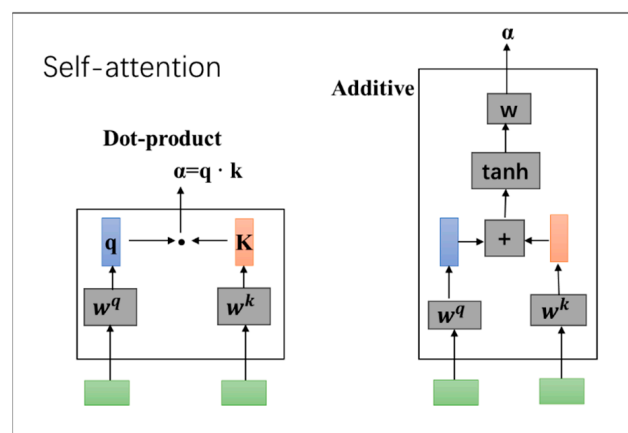


Figure 5. The fundamental structural of dot-product self-attention and self-additive attention.

The concrete implementation of scaled attention is shown in Equation (5), and the purpose of this operation is to prevent the inner product from being too large, avoid it being close to 1, and make it easy to train from the gradient point of view.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (5)$$

Similarly to how CNNs have multiple channels to extract feature information in different dimensions of an image, self-attention performs similar operations to extract the different features of the input layer for grouping to obtain multi-dimensional information. According to Figure 6, the multiple channels of the input features are separated into several groups to carry out the convolution individually and then the convolution operation is performed. This reduces the number of convolution operations required and, thus, increases computational efficiency.

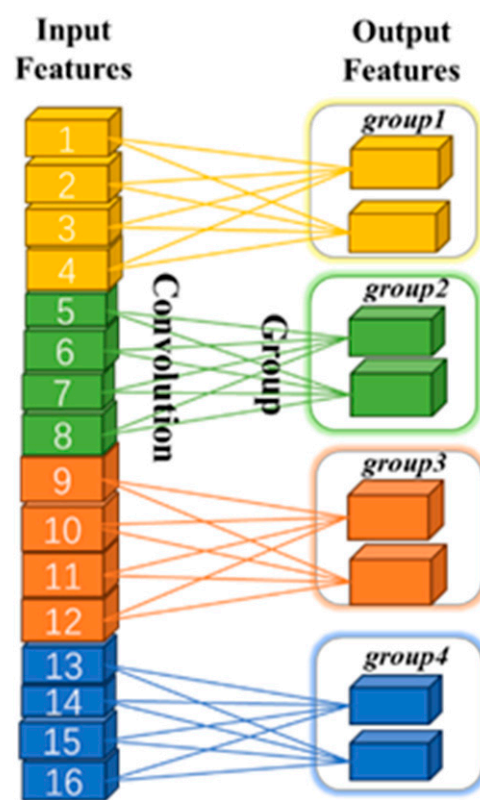


Figure 6. Group convolution.

Finally, the computed vectors are then subjected to the convolution operation, as shown in Figure 7. MHSA does this by partitioning the query, key, and value matrices of the input sequence into multiple heads, computing the attention weights independently in each head and synthesizing the results by running multiple self-attention layers in parallel. This enables the simultaneous capture and integration of many interacting features associated with bearing degradation in different subspaces. The outputs of these “heads” are then combined (often spliced and passed through a linear layer) to produce the final output representation, thus improving the expressive power of the model.

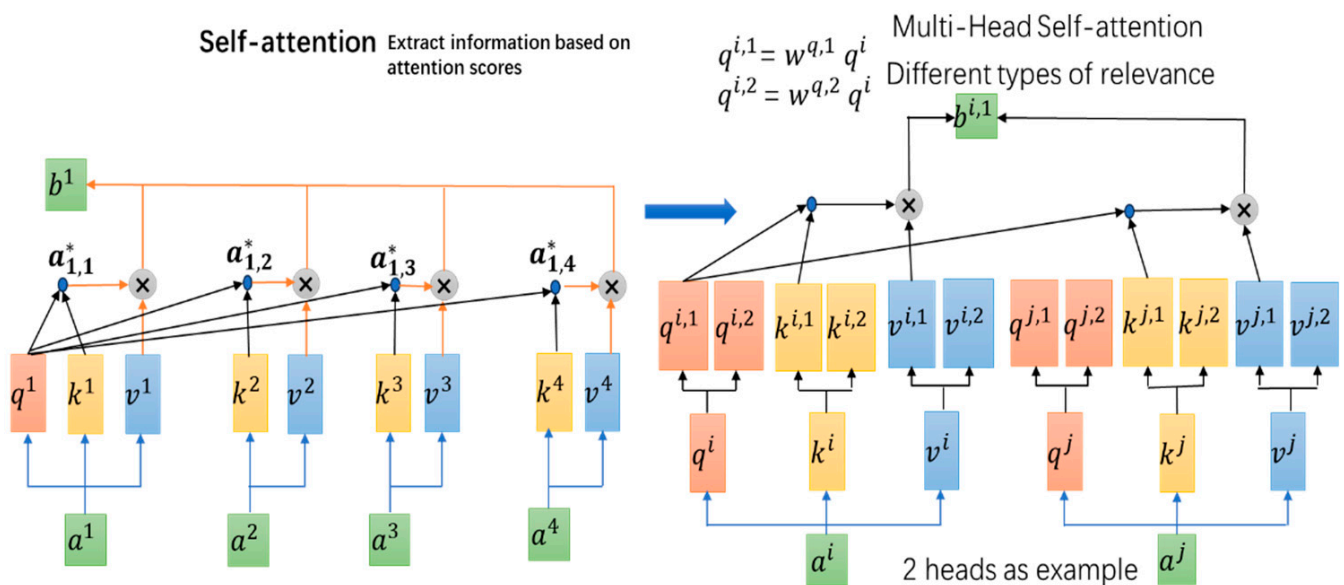


Figure 7. MHSA schematic.

To guarantee the precision and reliability of the prediction, the masked operation prevents the i th data from knowing the information after $i + 1$ data. This entails predicting the output at the t -th moment and not being able to see those inputs after the t -th moment, thus ensuring that training and prediction are consistent.

3. Modeling and Corresponding Indicators

3.1. Model Building

The proposed model fully exploits the advantages of the 1D-CNN multilayer perceptron structure and better preserves the original data's properties. Additionally, it overcomes the drawbacks of earlier bearing signal feature extraction methods that relied on human experience, better solving the rolling bearing vibration data feature extraction problem and increasing the precision of rolling bearing life prediction. A flowchart is displayed in Figure 8. The following are the specific operations:

- (1) Data preprocessing. The bandpass filtering interval was set according to the spectrum of the original signal and the noise reduction process was applied to the signal.
- (2) Determining the SPT. The starting point of remaining life prediction was determined based on the full-life trend map, with features such as root mean square and cliff, and the training and test sets were divided based on the SPT.
- (3) Deep feature extraction. The PSO algorithm was used to optimize the number of nodes in each layer of the four-layer CNN structure of 1D-CNN to extract the intrinsic characteristics of the bearing vibration data and finalize the unsupervised learning training.
- (4) Training phase. The extracted profound features from the optimized CNN were inputted into the PSO-optimized BiLSTM-MHSA network and, based on Equations (1)–(3) and the distinctive features of the memory cell structure in the BiLSTM network, a prediction model was built by training with BiLSTM-MHSA.
- (5) Testing phase. The partitioned test set was fed into the trained prediction model to determine the RUL of the bearings. A PSO-based 1D-CNN-BiLSTM-MHSA lifetime prediction model was constructed.

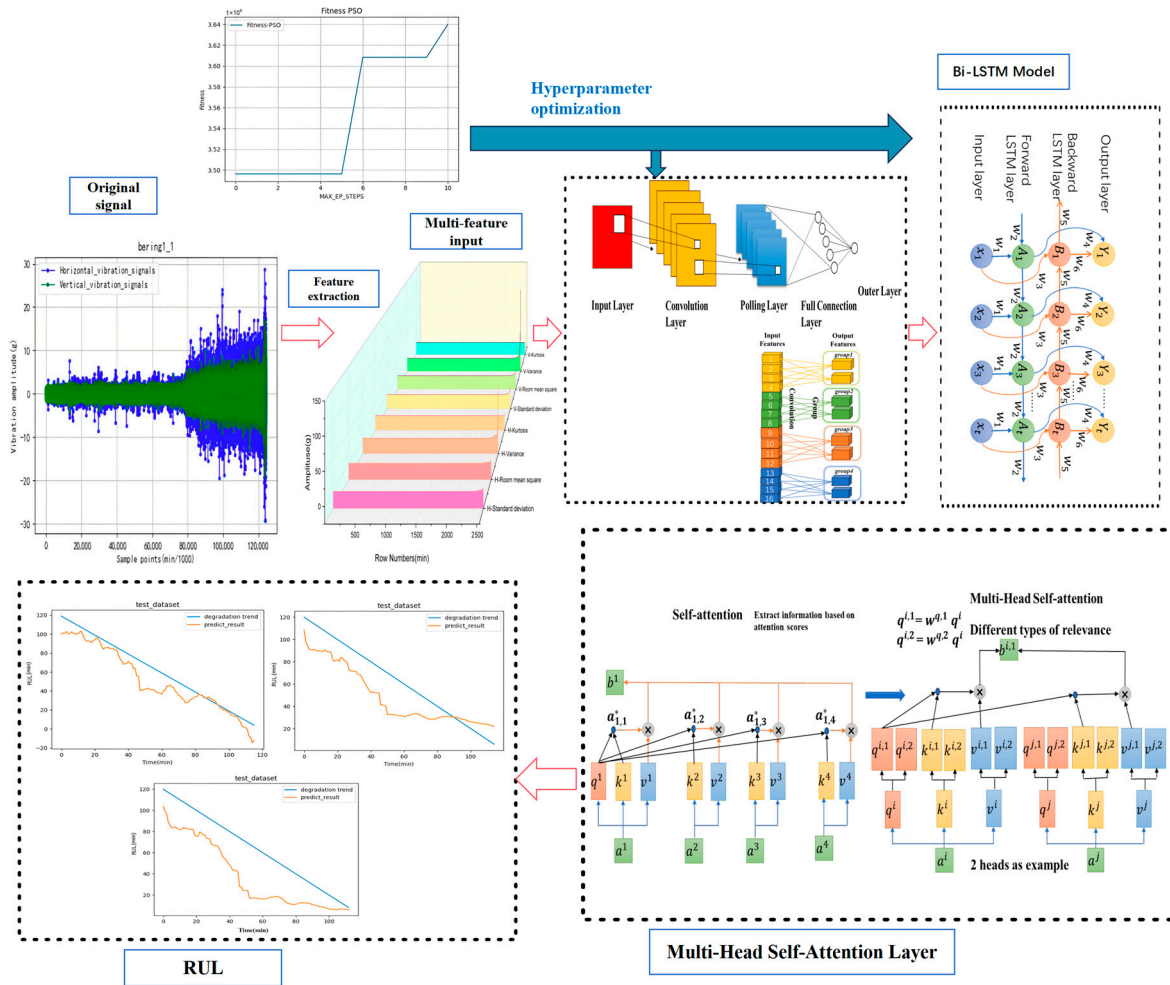


Figure 8. Flowchart of PSO-CNN-BiLSTM-MHSA.

3.2. Related Performance Indicators

The proposed prediction model's effectiveness could be confirmed by using relevant performance metrics. The mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and average accuracy were selected as the evaluation indexes for the RUL predictions. y_i represents the actual observation, y_i^2 is the average of the actual observations, and $\hat{y}_i$ represents the forecast value. The calculation formula is as follows:

$$MSE = \frac{1}{N} \sum_{i=1}^N ((y_i - \hat{y}_i))^2 \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N ((y_i - \hat{y}_i))^2} \quad (7)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (8)$$

$$R^2_score = 1 - \frac{\sum_{i=1}^N ((y_i - \hat{y}_i))^2 / N}{\sum_{i=1}^N ((y_i - y^2))^2 / N} \quad (9)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n |(\hat{y}_i - y_i) / y_i| * 100\% \quad (10)$$

$$MA = 1 - MAPE \quad (11)$$

3.3. Introduction to the Dataset

3.3.1. Description of Data Samples

The adjustable working conditions of the test platform mainly included radial force and rotational speed. Tests were designed for a total of three operating conditions, as shown in Table 1, with five bearings under each set of working conditions.

Table 1. Conducting accelerated life testing.

Condition No.	1	2	3
Angular velocity/(r/min)	2100	2250	2400
Axial force/kN	12	11	10

The XJTU-SY bearing dataset provided detailed information on each bearing tested, including each bearing's associated operating state, the total number of data samples, basic rated life L_{10} , actual life, and failure site. According to ISO 281-2007 [37], the basic rating rolling bearing lifespan is the life that can be achieved or exceeded by a group of bearings of the same type operating under the same conditions with a reliability of 90%. The calculation is as shown in Equation (12):

$$L_{10} = \frac{10^6}{60n} \left(\frac{C}{P} \right)^\varepsilon \quad (12)$$

The basic rated life is denoted by L_{10} , the rated dynamic load by C , the operating speed of the bearing by n , and the life index by ε ; the test bearing is a ball bearing and the reference standard is taken as 3; P represents the dynamic load equivalent when the bearing experiences just radial load and it may be determined using Formula (13).

$$P = f_p F_r \quad (13)$$

Here, f_p is the load factor; F_r represents the radial load. In the absence of a shock or minor shock, the value of f_p ranges from 1.0 to 1.2.

3.3.2. Selected Example Data

Vibration signals throughout the whole life cycle of bearings were acquired through the roller bearing life acceleration experimental platform. Figure 9 displays the raw signal visualization of the whole-life cycle vibration of bearing1_1 in transverse and longitudinal directions under XJTU-SY working conditions.

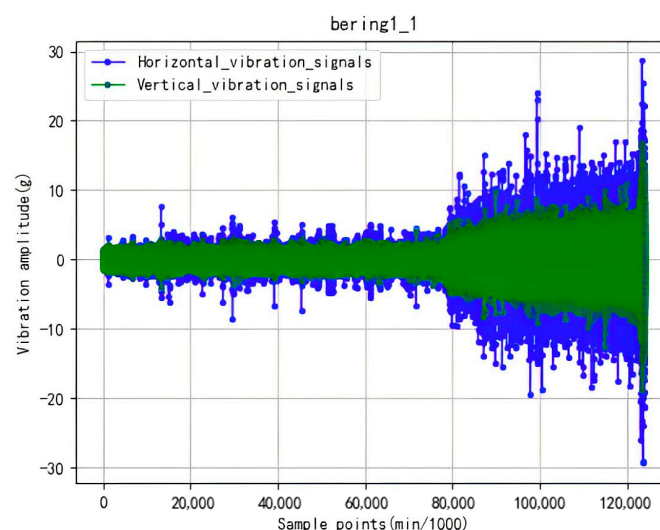


Figure 9. XJTU-SY bearing1_1 whole-life cycle transverse and longitudinal vibration signals.

4. Experimental Process and Analysis of Results

The experimental tools we used were Python 3.10 and PyCharm 2022.2.1. The environment was version 2022.9.0 based on Anaconda Navigator 3. The network framework was built on top of Keras 2.0.2 based on TensorFlow 2.10.0.

4.1. Experimental Data Preprocessing

To avoid the inconsistency of the feature metric scale exerting effects on the prediction accuracy, we carried out min–max normalization on the rolling bearing vibration signal [38]. Compared to other preprocessing methods, the normalization preprocessing method used in this study has the following advantages: (1) it is easy to implement and compute; (2) it can eliminate the effect of scaling between different time series; (3) it can scale the data to fit certain algorithms, such as neural networks. The equation for the min–max normalization course of events is as follows:

$$x'_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (14)$$

where x_i is the incipient vibration signal of the bearing, x_{min} is the minimum value in the incipient vibration signal of the bearing, x_{max} is the maximum value in the incipient vibration signal of the bearing, and x'_i is the normalized vibration signal of the bearing.

4.2. Characteristic Extraction

Since the trends of associated characteristics like standard deviation, root mean square, variance, and kurtosis have a strong correlation with the degradation trend of the remaining life of rolling bearings, these features were employed in this study to characterize the deterioration pattern effectively and reduce the number of raw data calculations, which enhanced the precision of the remaining useful life estimation.

The deterioration pattern of the rolling bearing under operating condition 1_1 can be clearly seen in Figure 10 below. During the first 77 min of the entire life cycle, the trend was relatively flat and the bearing was in good health; after 77 min, the degradation trend occurred and exceeded the value of L_{10} in Equation (14); and, at 123 min, the rolling bearing failed. The correlation multi-feature RUL deterioration trend graphs of the longitudinal and transverse vibration amplitudes for Cases 2_3 and 3_2 are displayed in Figure 11a,b, respectively.

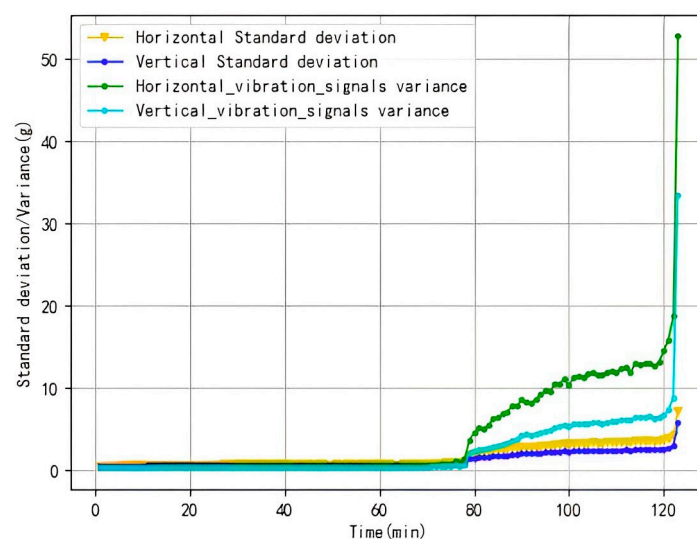


Figure 10. SD and variance of transverse longitudinal vibration amplitude for Bearing 1_1.

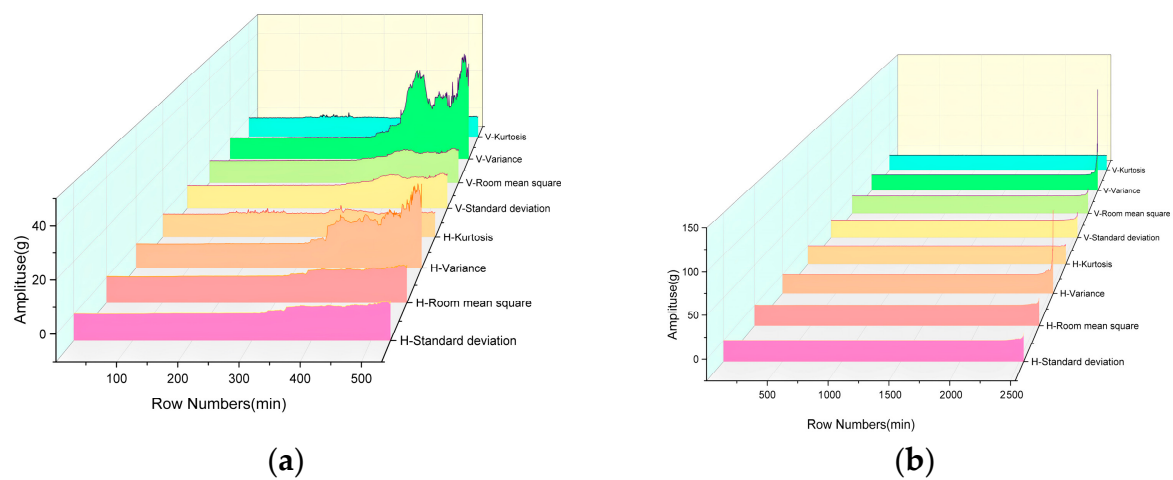


Figure 11. (a) Correlation feature maps for Bearing 2_3; (b) relevant feature maps for Bearing 3_1.

4.3. Starting Prediction Point (SPT) Optimization Selection

Having performed signal processing and correlation feature extraction on the Case 3_1 sample data as in Figure 12, it was clear that different training starting points had a significant effect on the prediction results. The predicted RMSE and MAPE values for various initial points are illustrated in Figure 13 and Table 2 below. Choosing the right starting point not only diminished the number of calculations required but also enhanced the precision of the forecast.

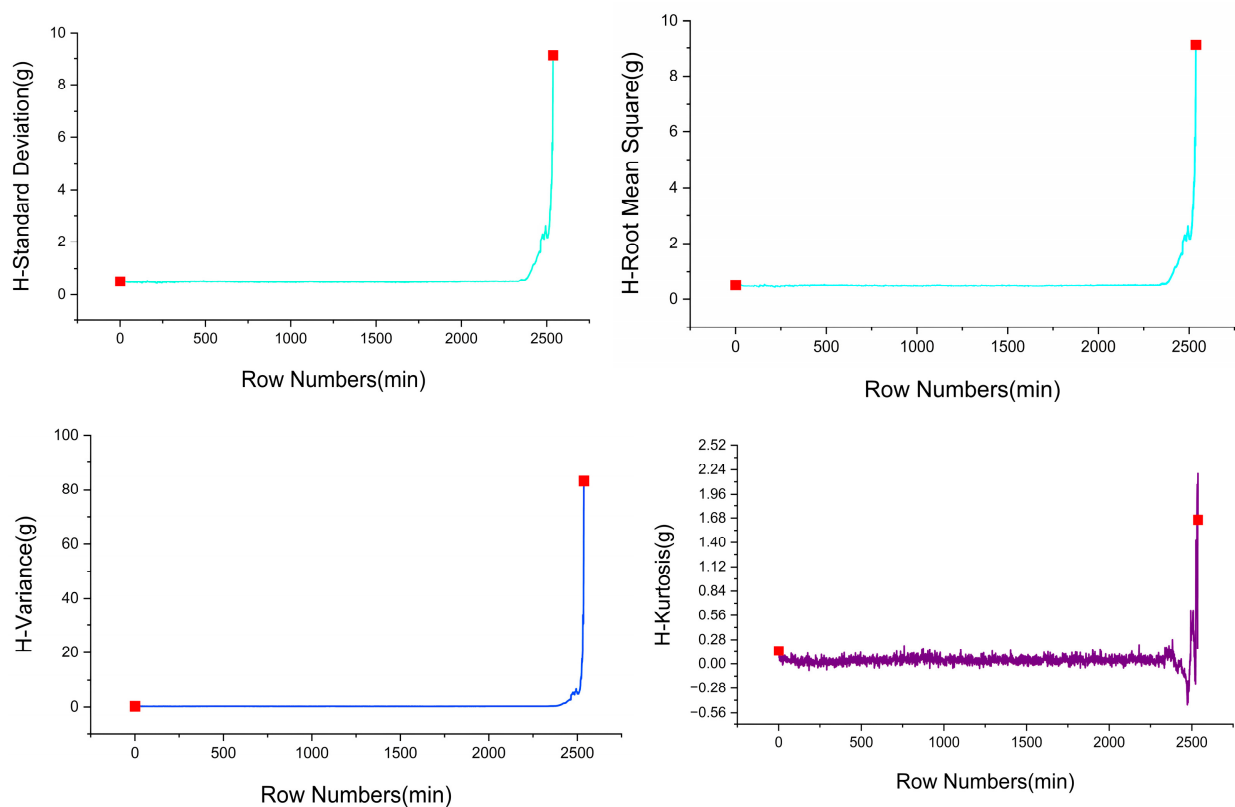


Figure 12. Trends in Bearing 3_1 related features.

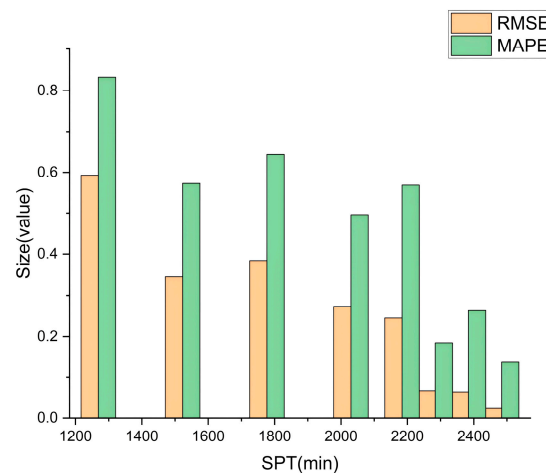


Figure 13. Comparative analysis of forecasted outcomes from various initial points.

Table 2. Comparative analysis of forecasted outcomes from various initial points.

SPT	RMSE	MAPE
2484	0.0235	0.1368
2384	0.0644	0.2633
2284	0.0673	0.1845
2184	0.2446	0.5682
2030	0.2734	0.4962
1777	0.3836	0.6435
1523	0.3442	0.5726
1269	0.5930	0.8326

Using the Rolling Bearing 3_1 dataset as an example, eight-dimensional features such as standard deviation, root mean square, variance, and crag were calculated for both horizontal and vertical directions. The prediction starting point was then determined so that a new eight-dimensional data set could be reconstructed.

To ensure the forecast's precision and the timely prediction of the bearing's lifespans, we chose to start with the 2284th data and use the 2284–2538 dataset as the dataset. The training set consisted of 80% of the data, whereas the test set comprised the remaining 20%. The process involved partitioning the dataset into K subsets, utilizing K-1 subsets as the training set for each iteration, and reserving the remaining subset as the validation set. This process was repeated K times, resulting in K model evaluation outcomes. The final result was obtained by calculating the average of these K evaluations. The value of K in this study was taken as 5.

4.4. Hyperparameter Preferences in Model-Related Partial Networks

Parameters like the number of convolutional layers and the node quantity significantly impact the efficacy of CNN models in extracting data features. Initially, the influence of the convolutional layer count on CNN feature extraction was established, and the number of convolutional layers was selected as a parameter for comparison. The average accuracy (MA) and the number of weights were selected as the judging criteria. Table 3 displays the comparative outcomes of how various CNN layers impact the prediction results. According to Figure 14 and Table 3, the optimal mean absolute (MA) error for the test sample was achieved when the convolutional model comprised four layers. Moreover, the quantity of nodes within each convolutional layer could have a great impact on the forecast's precision as well as the duration of training and testing once the count of convolutional layers was established. Optimizing the node count in the longitudinal comparison scheme became computationally demanding due to the high quantity of nodes in the convolutional layer. Thus, PSO was employed to refine the node quantity within each convolutional layer in

order to give full play to the superiority of the CNN model for the extraction of bearing fault signals. The maximum number of iterations in PSO's search for optimization of the model hyperparameters was 20, the maximum inertia weight was 0.9, the number of particles was 9, and the search dimension was 4. The optimal parameters are shown in Figure 15 after the global iteration operation. As depicted in Figure 15, the PSO optimization algorithm was optimal when the number of iterations was 6 and the optimal number of four-layer CNN nodes was 64.

Table 3. The impact of the CNN layer count on the forecast outcomes.

Layers of the CNN	MA	Weights Count
2	0.822	93,696
3	0.882	15,923
4	0.893	22,476
5	0.874	29,030
6	0.861	35,584

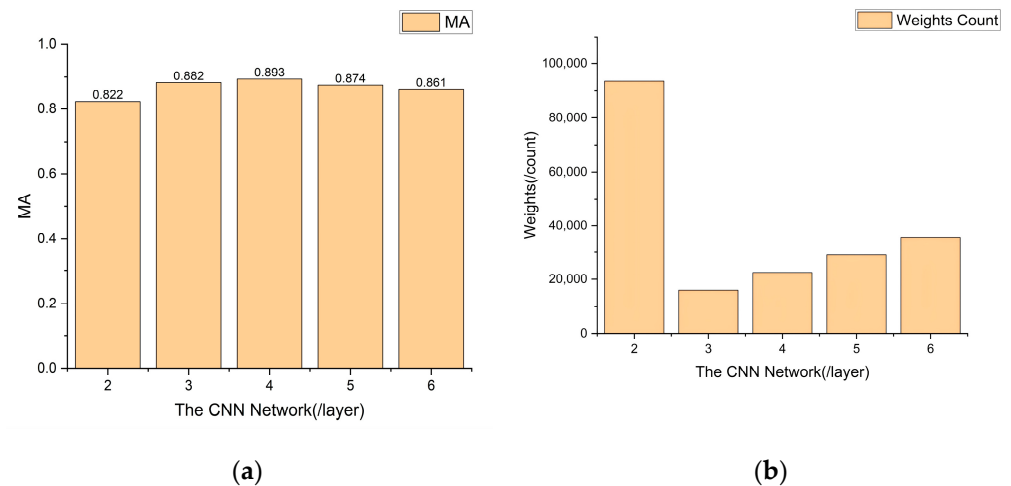


Figure 14. The impact of the CNN layer count on the forecast outcomes. (a) The effect of the number of CNN layers on MA; (b) the effect of the number of CNN layers on the number of weights.

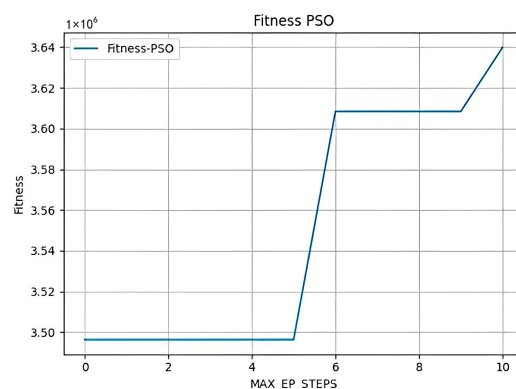


Figure 15. Evolutionary curve of PSO fitness.

In order to determine when the tested bearing reached the expected life, based on the selection of the starting prediction point (Section 3.2), the Bearing 3_1 sample data were utilized as the processed dataset within the training set for model training. As observed in Figure 16, the training error approached zero as the number of iterations increased, signifying exceptional model training. Table 4 shows the structural parameters of the trained BiLSTM network.

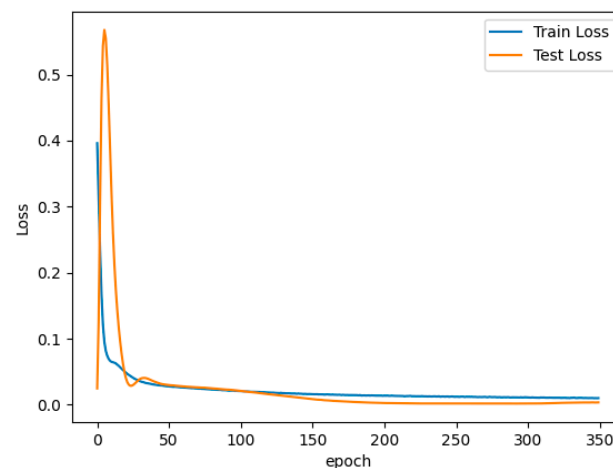


Figure 16. Bearing life prediction training and testing loss function MSE.

Table 4. BiLSTM–MHSA parameters.

Network Configuration	Parameter Configuration
Input step size	3
Number of output step sizes	3
Number of hidden layers	8, 8
Network layer	8
Learning rate	0.01

In this study, the test set comprised 250 datasets selected from points following the SPT markers within the full life cycle dataset of Bearing 3_1. These were fed into the previously trained BiLSTM–MHSA network to generate the output from this network, which is depicted in Figure 16. The specific parameters related to the BiLSTM–MHSA network framework are detailed in Table 4.

Following a series of experiments, optimal dimensional parameters were determined within a reasonable time period. These included a window size of 5, an epoch counts of 350, 64 filters for the Cov1D layer, and 64 units for the BiLSTM layer.

4.5. Training of the Model

Through the processing of the data, described in Sections 4.1 and 4.2, and the production of the dataset, this study obtained the correct prediction process. The model's training procedure is outlined in this section. The training sequence and the calculation of the relevant MSE evaluation parameters are shown in Figure 16. As shown in Figure 16, the model's performance dropped sharply at the 10th epoch value, leveled off slowly between the 10th and 60th epochs, and remained stable after 60 epochs. This shows that the experiment achieved better training results. In order to verify the validity of the experimental validation method using the XJTU-SY dataset, we further conducted a K-fold cross-validation experiment. K-fold cross-validation is a widely employed technique for assessing the performance of a model. The process involves partitioning a dataset into K subsets, utilizing K-1 subsets as the training set for each iteration, and reserving the remaining subset as the validation set. This process is repeated K times, resulting in K model evaluation outcomes. The final result is obtained by calculating the average of these K evaluations. The value of K in this study was taken as 5. This approach allowed for a more objective evaluation of the model's performance and helped mitigate bias resulting from arbitrary data partitioning.

4.6. RUL Prediction Based on CNN–BiLSTM–MultiHead–Self Attention

The 2150th to 2418th datasets of Bearing 3_1 were input into the trained network model as a training set and the predicted output values from 2149 to 2538 were obtained. A comparison of different input and output scales was plotted, and the outcomes are displayed in Figure 17. RUL prediction trends for the 3-, 5-, and 7-scale output results are shown in subfigures a, b, and c, respectively. It was evident from the experiment that the failure point for scale 3 was advanced by around 15 min, the failure point for scale 5 was advanced by approximately 20 min, and, for scale 7, the projected failure point was extremely close to the actual failure point timing.

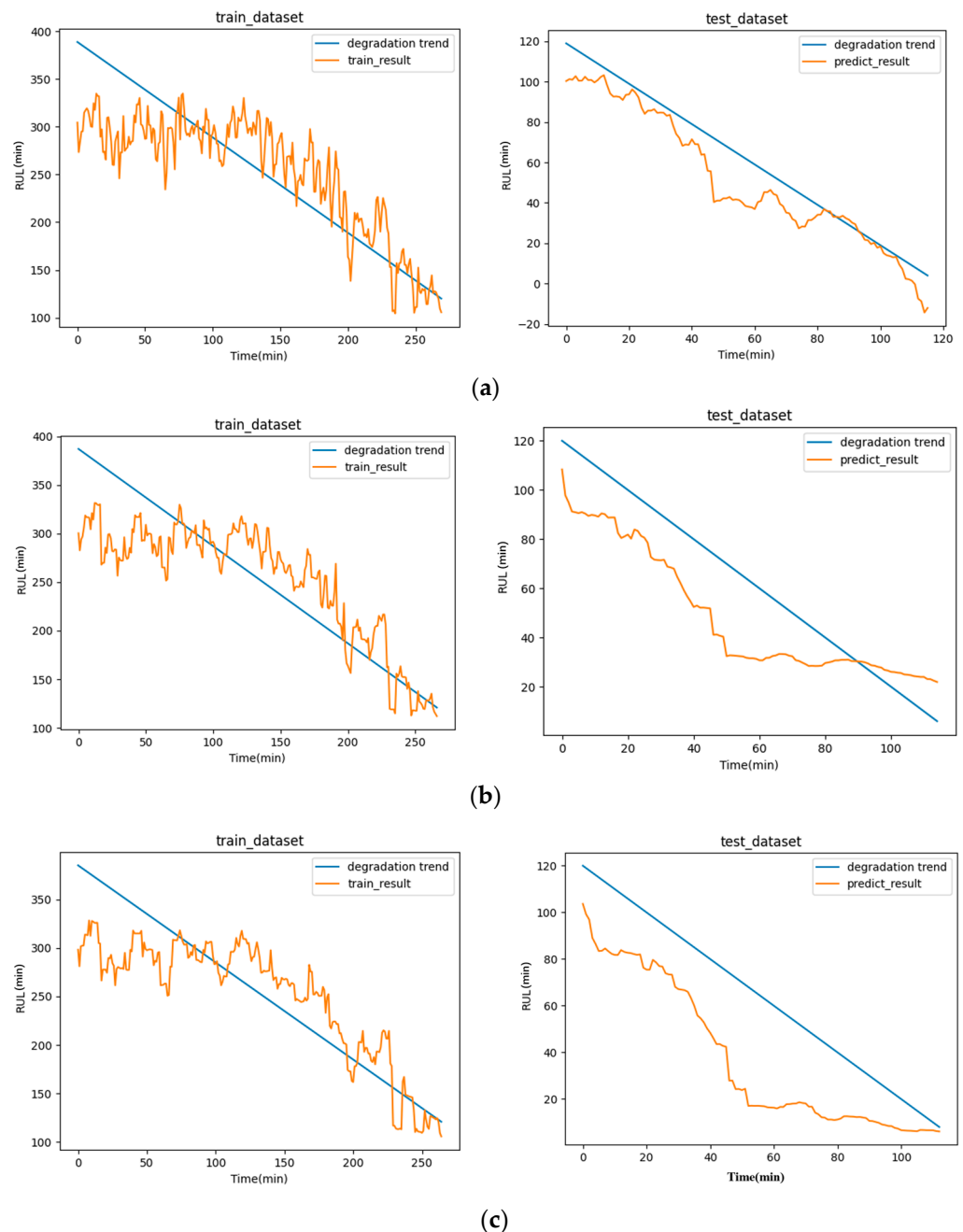


Figure 17. Trends in test degradation at different scales. (a) Output scale 3; (b) output scale 5; (c) output scale 7.

Based on Figure 18, the model presented in this paper estimated that the predicted RUL of the bearing had the same trend as the real data. The curve fit between the predicted

and actual data was high and the discrepancy between them was minimal. In the advanced stages of rolling bearing performance, the predicted values exhibited minimal fluctuation. This finding demonstrates that the eigenvalues selected in this research are capable of accurately representing the wear condition of rolling bearings. Furthermore, it validates the effectiveness of the proposed methodology for predicting the remaining operational lifespan of bearings. This study presents a more precise method for forecasting the lifespan of rolling bearings, which will aid in decreasing the occurrence of accidents and help to reduce operational and maintenance expenses in industrial manufacturing.

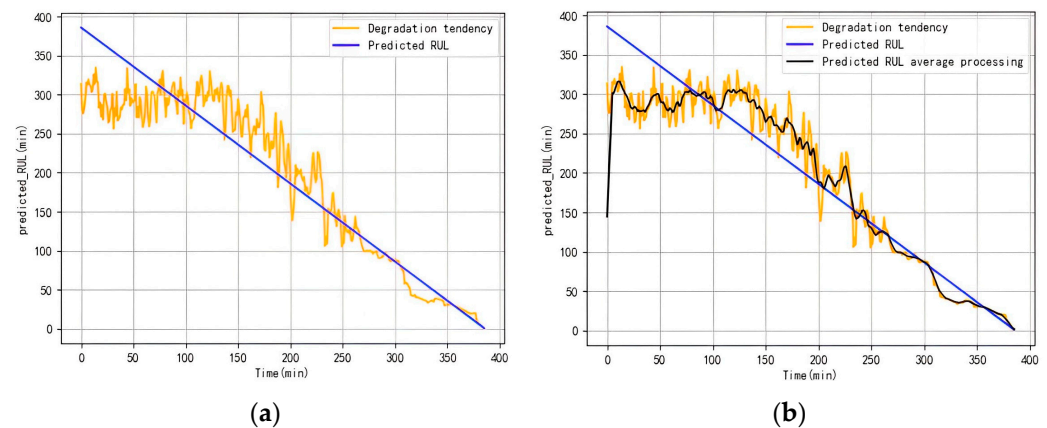


Figure 18. (a) Residual life degradation trend and RUL prediction chart; (b) RUL after sliding smoothing filtering.

4.7. Comparative Analysis of Experimental Results

To verify the effectiveness of deep feature extraction relative to manual feature extraction, the methodology of this paper considered Bi-LSTM networks capable of leveraging the advantages of historical degradation data that were temporally correlated. Specifically, experiments with articles compared a baseline model, the traditional LSTM neural network, CNN-GRU, the traditional CNN-LSTM structure, CNN-BiLSTM, and the model presented in this paper with the predictive outcomes of each approach depicted in Figure 17. Figure 19 and Table 5 compare the predictions of the six methods.

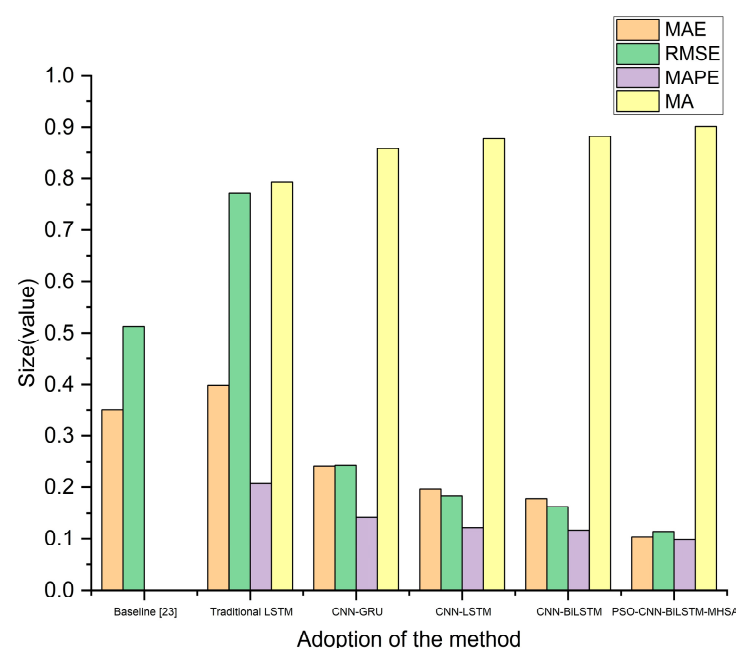


Figure 19. Prediction accuracy plot of different methods.

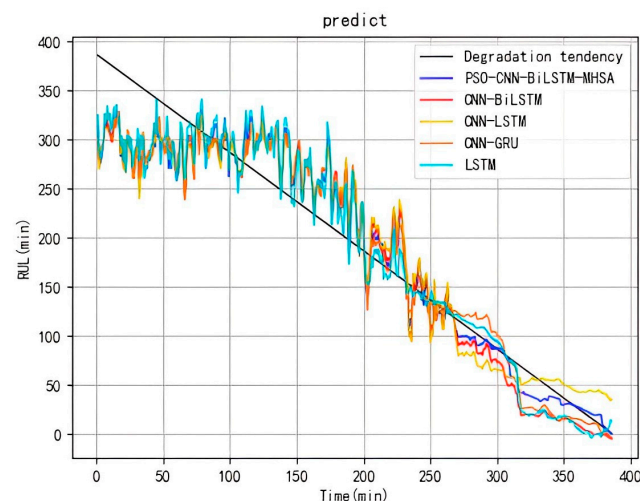
Table 5. Forecasting inaccuracies across six techniques.

Adoption of the Method	MAE	RMSE	MAPE	MA
Baseline [23]	0.3506	0.5136	-	-
Traditional LSTM	0.3985	0.7722	0.2065	0.7935
CNN-GRU	0.2423	0.2438	0.1426	0.8574
CNN-LSTM	0.1963	0.1824	0.1223	0.8777
CNN-BiLSTM	0.1772	0.1624	0.1173	0.8827
PSO-CNN-BiLSTM-MHSA	0.1036	0.1148	0.0988	0.9012

Note: “-” means the data are not provided.

The outcomes of the conducted ablation studies, illustrated in Table 5, reveal that the CNN-LSTM model incorporating a CNN demonstrated its superiority in extracting the deep characteristics of the data compared to the traditional LSTM. While the MAE and RMSE were reduced by 0.2022 and 0.5898, respectively, the MAPE declined by 0.0842. After further combining the PSO optimization algorithm and incorporating the MHSA, there was an even more significant performance improvement. Compared to the CNN-LSTM model, the PSO-CNN-BiLSTM-MHSA model introduced in this study reduced the MAE and RMSE by 0.0927 and 0.0676, respectively, while the MAPE decreased by 0.0235. Evidently, the composite CNN-LSTM technique outperformed the standalone LSTM model. The model presented in this research article outperformed CNN-LSTM in terms of optimizing model hyperparameters. Additionally, the MHSA weight assignment in this model was superior, leading to a more significant impact on the projected results. This demonstrates that the model developed in this study displays exceptional predictive capabilities.

Figure 20 illustrates that the CNN-GRU, CNN-LSTM neural network, and conventional LSTM model exhibited significant deviations from the actual values during the later stages of bearing operation. These three approaches failed to fully harness the inherent connections within the time series data. While the CNN-BiLSTM network model demonstrated improved alignment compared to the first three models, it displayed some level of discretization towards the end of the bearing’s operation. This was due to the influence of the relevant hyperparameters on the network, which resulted in feature extraction that did not extract the deeper features more fully. In terms of accuracy and fit, the PSO-optimized CNN-BiLSTM-MHSA model presented in this study outperformed the other four conventional models and the method outlined in [23]. Additionally, the methodology’s predicted values were essentially in line with the true values, demonstrating the viability of the approach presented in this paper. Table 5 reveals that the life prediction based on PSO-CNN-BiLSTM-MHSA, with MAE = 0.1036, RMSE = 0.1148, and MAPE = 0.0988, was smaller than that of the other methods, proving that this technique exhibited the utmost precision in our forecasts, accurately determining the RUL of rolling bearings.

**Figure 20.** Comparison of results of four methods.

5. Conclusions

A model for predicting the residual life of rolling bearings utilizing PSO-optimized CNN-BiLSTM-MHSA parameters was proposed. The precision in forecasting the RUL of rolling bearings was improved by merging CNN's deep feature extraction capabilities with BiLSTM's proficiency in handling time series data. Our experimental study yielded the following conclusions.

- (1) Through graphs depicting the trends in the SD, RMS, variance, and kurtosis characteristics over the whole life cycle of the bearings, SPT points were identified to enhance the accuracy of bearing life prediction.
- (2) The PSO-optimized 1D-CNN could fully extract the characteristic deep features, enhancing the ability to detect the deterioration of bearings using high-dimensional complex data. The proposed approach leverages the autonomous learning capability of the BiLSTM network to enhance prediction accuracy for time series data. It utilizes historical data on deterioration over time to determine the degradation state and successfully applies nonlinear function mapping.
- (3) Upon conducting a comparative analysis with various prediction methods, the method for predicting bearing degradation was discovered based on PSO-optimized CNN-BiLSTM-MHSA, surpassing the other three approaches in terms of forecast accuracy and enhancing the precision of predicting the RUL of rolling bearings. Compared to the CNN-LSTM model, the PSO-CNN-BiLSTM-MHSA model introduced in this study reduced the MAE and RMSE by 0.0927 and 0.0676, respectively, while the MAPE decreased by 0.0235.

In future studies, it will be necessary to further enhance the model structure, minimize the training time, and decrease the model's storage requirements. Implementing these changes will enhance the dependability and punctuality of rolling bearing RUL forecasts, resulting in cost savings due to decreased expenditure on storage resources and application expenses. Furthermore, we may enhance the information integration capabilities of the model and better optimize the front-end preprocessing and back-end prediction of the data. By employing this approach, it is possible to develop a more efficient, cost-effective, and intelligent technique for predicting the remaining lifespans of bearings.

Author Contributions: Conceptualization, J.Y. and X.Z.; methodology, J.Y., X.Z. and S.L. (Shangfang Li); software, X.Z. and S.L. (Song Liu); validation, S.L. (Song Liu) and X.Y.; writing—original draft preparation, S.L. (Shangfang Li) and J.Y.; writing—review and editing, J.Y. and S.L. (Shangfang Li); visualization, X.Y.; supervision, S.L. (Shangfang Li). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data involved in the making of this paper are presented in the article.

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

References

1. Ding, X.; Wang, H.; Cao, Z.; Liu, X.Z.; Liu, Y.B.; Huang, Z.F. An Edge Intelligent Method for Bearing Fault Diagnosis Based on a Parameter Transplantation Convolutional Neural Network. *Electronics* **2023**, *12*, 1816. [\[CrossRef\]](#)
2. Fan, H.W.; Xue, C.Y.; Ma, J.T.; Cao, X.G.; Zhang, X.H. A novel intelligent diagnosis method of rolling bearing and rotor composite faults based on vibration signal-to-image mapping and CNN-SVM. *Meas. Sci. Technol.* **2023**, *34*, 044008. [\[CrossRef\]](#)
3. Du, J.F.; Li, X.Y.; Gao, Y.P.; Gao, L. Integrated gradient-based continuous wavelet transform for bearing fault diagnosis. *Sensors* **2022**, *22*, 8760. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Lin, C.-L.; Pozzebon, M.; Sokolowski, K.A.; Meehan, P.A. Experimental investigation on rolling contact wear in grease lubricated spherical roller bearings using microcomputed tomography (μ CT). *Wear* **2023**, *534*, 205121. [\[CrossRef\]](#)
5. Bertocco, M.; Fort, A.; Landi, E.; Mugnaini, M.; Parri, L.; Peruzzi, G. Roller bearing failures classification with low computational cost embedded machine learning. In Proceedings of the 2022 IEEE International Workshop on Metrology for Automotive (MetroAutomotive), Modena, Italy, 4–6 July 2022; pp. 12–17. [\[CrossRef\]](#)
6. Xia, Z.F.; Wu, D.; Zhang, X.C.; Wang, J.Q.; Han, E.-H. Rolling contact fatigue failure mechanism of bearing steel on different surface roughness levels under heavy load. *Int. J. Fatigue* **2024**, *179*, 108042. [\[CrossRef\]](#)

7. Wang, B.; Guo, Y.B.; Zhang, Z.; Wang, D.G.; Wang, J.Q.; Zhang, Y.S. Developing and applying OEGOA-VMD algorithm for feature extraction for early fault detection in cryogenic rolling bearing. *Measurement* **2023**, *216*, 112908. [\[CrossRef\]](#)
8. Wang, B.; Lei, Y.; Li, N.; Li, N. A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings. *IEEE Trans. Reliab.* **2020**, *69*, 401–412. [\[CrossRef\]](#)
9. Hakim, M.; Omran, A.A.B.; Ahmed, A.N.; Al-Waily, M.; Abdellatif, A. A systematic review of rolling bearing fault diagnoses based on deep learning and transfer learning: Taxonomy, overview, application, open challenges, weaknesses and recommendations. *Ain Shams Eng. J.* **2023**, *14*, 101945. [\[CrossRef\]](#)
10. Chen, C.C.; Liu, Z.; Yang, G.; Wu, C.C.; Ye, Q. An Improved Fault Diagnosis Using 1D-Convolutional Neural Network Model. *Electronics* **2020**, *10*, 59. [\[CrossRef\]](#)
11. Li, X.C.; Elasha, F.; Shanbr, S.; Mba, D. Remaining Useful Life Prediction of Rolling Element Bearings Using Supervised Machine Learning. *Energies* **2019**, *12*, 2705. [\[CrossRef\]](#)
12. Gu, K.; Zhang, Y.; Liu, X.; Li, H.; Ren, M. DWT-LSTM-Based Fault Diagnosis of Rolling Bearings with Multi-Sensors. *Electronics* **2021**, *10*, 2076. [\[CrossRef\]](#)
13. Xie, Z.; Du, S.; Lv, J.; Deng, Y.; Jia, S. A Hybrid Prognostics Deep Learning Model for Remaining Useful Life Prediction. *Electronics* **2020**, *10*, 39. [\[CrossRef\]](#)
14. Zou, Y.; Li, Z.; Liu, Y.; Zhao, S.; Liu, Y.; Ding, G. A method for predicting the remaining useful life of rolling bearings under different working conditions based on multi-domain adversarial networks. *Measurement* **2022**, *188*, 110393. [\[CrossRef\]](#)
15. Hu, P.; Zhao, C.S.; Huang, J.C.; Song, T.G. Intelligent and Small Samples Gear Fault Detection Based on Wavelet Analysis and Improved CNN. *Processes* **2023**, *11*, 2969. [\[CrossRef\]](#)
16. Zhou, K.; Tang, J. A wavelet neural network informed by time-domain signal preprocessing for bearing remaining useful life prediction. *Appl. Math. Model.* **2023**, *122*, 220–241. [\[CrossRef\]](#)
17. Song, L.; Wu, J.; Wang, L.P.; Chen, G.; Shi, Y.L.; Liu, Z.G. Remaining Useful Life Prediction of Rolling Bearings Based on Multi-Scale Attention Residual Network. *Entropy* **2023**, *25*, 798. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Qi, J.Y.; Zhu, R.; Liu, C.Y.; Mauricio, A.; Gryllias, K. Anomaly detection and multi-step estimation based remaining useful life prediction for rolling element bearings. *Mech. Syst. Signal Process.* **2024**, *206*, 110910. [\[CrossRef\]](#)
19. Kong, W.L.; Li, H. Remaining useful life prediction of rolling bearing under limited data based on adaptive time-series feature window and multi-step ahead strategy. *Appl. Soft Comput.* **2022**, *129*, 109630. [\[CrossRef\]](#)
20. Song, B.Y.; Liu, Y.Y.; Fang, J.Z.; Liu, W.B.; Zhong, M.Y.; Liu, X.H. An optimized CNN-BiLSTM network for bearing fault diagnosis under multiple working conditions with limited training samples. *Neurocomputing* **2024**, *574*, 127284. [\[CrossRef\]](#)
21. Wang, Y.P.; Zhao, J.J.; Yang, C.N.; Xu, D.; Ge, J.H. Remaining useful life prediction of rolling bearings based on Pearson correlation-KPCA multi-feature fusion. *Measurement* **2022**, *201*, 111572. [\[CrossRef\]](#)
22. Zhong, J.; Li, H.; Chen, Y.; Huang, C.; Zhong, S.; Geng, H. Remaining Useful Life Prediction of Rolling Bearings Based on ECA-CAE and Autoformer. *Biomimetics* **2024**, *9*, 40. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Jun, H.; Jung, I.Y. Enhancement of Product-Inspection Accuracy Using Convolutional Neural Network and Laplacian Filter to Automate Industrial Manufacturing Processes. *Electronics* **2023**, *12*, 3795. [\[CrossRef\]](#)
24. Shang, Y.J.; Tang, X.L.; Zhao, G.Q.; Jiang, P.G.; Lin, T.R. A remaining life prediction of rolling element bearings based on a bidirectional gate recurrent unit and convolution neural network. *Measurement* **2022**, *202*, 111893. [\[CrossRef\]](#)
25. Jiang, C.H.; Liu, X.Y.; Liu, Y.Z.; Xie, M.J.; Liang, C.; Wang, Q.M. A Method for Predicting the Remaining Life of Rolling Bearings Based on Multi-Scale Feature Extraction and Attention Mechanism. *Electronics* **2022**, *11*, 3616. [\[CrossRef\]](#)
26. Yao, D.C.; Li, B.Y.; Liu, H.C.; Yang, J.W.; Jia, L.M. Remaining useful life prediction of roller bearings based on improved 1D-CNN and simple recurrent unit. *Measurement* **2021**, *175*, 109166. [\[CrossRef\]](#)
27. Zhang, Q.; Ye, Z.J.; Shao, S.Y.; Niu, T.L.; Zhao, Y.W. Remaining useful life prediction of rolling bearings based on convolutional recurrent attention network. *Assem. Autom.* **2022**, *42*, 372–387. [\[CrossRef\]](#)
28. Saufi, M.S.R.M.; Hassan, K.A. Remaining useful life prediction using an integrated Laplacian-LSTM network on machinery components. *Appl. Soft Comput.* **2021**, *112*, 107817. [\[CrossRef\]](#)
29. Zhang, X.G.; Yang, J.Z.; Yang, X.M. Residual Life Prediction of Rolling Bearings Based on a CEEMDAN Algorithm Fused with CNN-Attention-Based Bidirectional LSTM Modeling. *Processes* **2024**, *12*, 8. [\[CrossRef\]](#)
30. Ahmad, Z.; Nguyen, T.K.; Kim, J.M. Leak detection and size identification in fluid pipelines using a novel vulnerability index and 1-D convolutional neural network. *Eng. Appl. Comput. Fluid. Mech.* **2023**, *17*, 2165159. [\[CrossRef\]](#)
31. Li, Y.X.; Mu, L.X.; Gao, P.Y. Particle Swarm Optimization Fractional Slope Entropy: A New Time Series Complexity Indicator for Bearing Fault Diagnosis. *Fractal Fract.* **2022**, *6*, 345. [\[CrossRef\]](#)
32. Wan, S.; Li, X.; Zhang, Y.; Liu, S.; Hong, J.; Wang, D. Bearing remaining useful life prediction with convolutional long short-term memory fusion networks. *Reliab. Eng. Syst. Saf.* **2022**, *224*, 108528. [\[CrossRef\]](#)
33. Yu, K.; Kong, C.Y.; Zhong, L.M.; Fu, J.F.; Shao, J. Delay prediction with spatial-temporal bi-directional LSTM in railway network. *ICT Express* **2023**, *9*, 921–926. [\[CrossRef\]](#)
34. Sun, H.B.; Cui, Q.; Wen, J.Y.; Kou, L.; Ke, W.D. Short-term wind power prediction method based on CEEMDAN-GWO-Bi-LSTM. *Energy Rep.* **2024**, *11*, 1487–1502. [\[CrossRef\]](#)
35. Dong, S.; Xiao, J.; Hu, X.; Fang, N.; Liu, L.; Yao, J. Deep transfer learning based on Bi-LSTM and attention for remaining useful life prediction of rolling bearing. *Reliab. Eng. Syst. Saf.* **2023**, *230*, 108914. [\[CrossRef\]](#)

36. Zhuang, J.; Liu, Y.; Xu, N.; Zhu, Y.; Xiao, J.; Gu, J.; Mao, T. Fast Self-Attention Deep Detection Network Based on Weakly Differentiated Plant Nematodes. *Electronics* **2022**, *11*, 3497. [[CrossRef](#)]
37. *BS ISO 281:2007*; Rolling Bearings: Dynamic Load Ratings and Rating Life. 2nd ed. BS ISO: London, UK, 2007.
38. Hernández-Cámara, P.; Vila-Tomás, J.; Laparra, V.; Malo, J. Neural networks with divisive normalization for image segmentation. *Pattern Recognit. Lett.* **2023**, *173*, 64–71. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.