Structural Durability & Health Monitoring

Tech Science Press

DOI: 10.32604/sdhm.2024.053998

ARTICLE

Data-Driven Method for Predicting Remaining Useful Life of Bearings Based on Multi-Layer Perception Neural Network and Bidirectional Long Short-Term Memory Network

Yongfeng Tai[1], Xingyu Yan[2], Xiangyi Geng[3], Lin Mu[4], Mingshun Jiang[2] and Faye Zhang[2,*]

[1]CRRC Qingdao Sifang Co., Ltd., Qingdao, 266111, China

[2]School of Control Science and Engineering, Shandong University, Jinan, 250061, China

[3]Public (Innovation) Experimental Teaching Center, Shangdong University, Qingdao, 266237, China

[4]Engineering Training Center, Shangdong University, Jinan, 250061, China

[*]Corresponding Author: Faye Zhang. Email: zhangfaye@sdu.edu.cn

Received: 15 May 2024 Accepted: 31 July 2024 Published: 15 January 2025

ABSTRACT

The remaining useful life prediction of rolling bearing is vital in safety and reliability guarantee. In engineering scenarios, only a small amount of bearing performance degradation data can be obtained through accelerated life testing. In the absence of lifetime data, the hidden long-term correlation between performance degradation data is challenging to mine effectively, which is the main factor that restricts the prediction precision and engineering application of the residual life prediction method. To address this problem, a novel method based on the multi-layer perception neural network and bidirectional long short-term memory network is proposed. Firstly, a nonlinear health indicator (HI) calculation method based on kernel principal component analysis (KPCA) and exponential weighted moving average (EWMA) is designed. Then, using the raw vibration data and HI, a multi-layer perceptron (MLP) neural network is trained to further calculate the HI of the online bearing in real time. Furthermore, The bidirectional long short-term memory model (BiLSTM) optimized by particle swarm optimization (PSO) is used to mine the time series features of HI and predict the remaining service life. Performance verification experiments and comparative experiments are carried out on the XJTU-SY bearing open dataset. The research results indicate that this method has an excellent ability to predict future HI and remaining life.

KEYWORDS

Remaining useful life prediction; rolling bearing; health indicator construction; multilayer perceptron; bidirectional long short-term memory network

1 Introduction

Remaining useful life (RUL), also known as Residual Life, refers to the left time before machinery loses its operation capability based on current age and conditions, as well as past operation conditions [1]. Remaining useful life is usually defined as a conditional random variable specified in relative or absolute time units, such as load cycles, flight hours, etc. The RUL for the current time can be calculated as:

$$l_k = \inf\{l_k : x(l_k + t_k) \geq \lambda | x_{0:k}\} \tag{1}$$

where $x(l_k + t_k)$ represents the health status at future time $l_k + t_k$, and $x_{0:k}$ represents the health status observed at the previous time $t_{0:k}$, λ is a predetermined fault threshold.

One of the most important components of rotating equipment is the rolling bearing. Life prediction is essential for early defect diagnosis and for quickly creating a maintenance schedule to prevent poor performance and lower maintenance expenses. Thus, it is important to develop an efficient rolling bearing life forecast system. The two primary life prediction approaches are data-driven and physical model-based approaches. Physical model-based methods have limited generalizability and substantially depend on the domain expertise of physical systems [2,3]. The life prediction process in data-driven approaches is based on historical data modelling [4], which typically entails three steps: data collection, HI design, and life expectancy [5]. It is acknowledged that HI design is an essential first step in evaluating rolling bearing degradation [6]. Nowadays, a popular approach for designing HI is to combine several statistical features (such as Principal Component Analysis, Local Linear Embedding, Local Preserving Projection, etc.) into one HI through dimensionality reduction. By removing unnecessary elements, these techniques often result in the collection of more thorough health data [7]. On the other hand, the deterioration index creation approach based on artificial feature extraction mostly depends on empirical knowledge [8]. Consequently, in order to increase the residual life forecast method's accuracy and offer technical assistance for its engineering application, more research must be done, and an adaptive deterioration index must be created [9].

After the HI design is completed, a life prediction method is needed to predict future HI and RUL. In recent years, prediction methods based on artificial intelligence have been extensively studied for life prediction, including artificial neural networks (ANN) and support vector machines (SVM) [10,11]. However, some results show that the shallow architectures of ANN and SVM limit their ability to learn complex nonlinear relationships. In addition, the degradation process of rolling bearings often exhibits highly nonlinear and non-stationary characteristics. Therefore, developing a more powerful rolling bearing health prediction method is necessary. Deep learning is one of the most promising artificial intelligence approaches and has recently been successfully applied to numerous tasks [12–14]. Deep learning is built on a multi-layered structure called deep architecture, which significantly enhances its capacity to learn complicated nonlinear correlations, in contrast to classic machine learning techniques like ANN and SVM. Fault diagnosis has steadily benefited from the application of deep learning techniques, including convolutional neural networks, stacked denoising autoencoders, and deep belief networks (DBN). However, there are few deep-learning techniques available for life prediction. Among the most often used deep learning models are long short-term memory networks (LSTM) [15,16]. The LSTM model is well-suited for processing time series data. However, the LSTM model's ability to capture the temporal correlation of data is limited to a single direction, whether forward or backward. Bi-directional LSTM (BiLSTM), a variant of the LSTM model, can access sequential data in both forward and backward time directions, thus capturing hidden long-term dependencies effectively. It should improve the accuracy of residual life prediction and provide technical support for the engineering application of residual life prediction methods in engineering.

In this paper, for the problem of condition assessment and life prediction of rolling bearings, combined with the vibration signals when the bearings are running, a health indicator construction method based on the multilayer perceptron machine network and a residual life prediction method based on the multilayer BiLSTM network optimized by particle swarm algorithm is proposed. The performance of the method is verified using the XJTU-SY bearing test dataset, and the results show that the method is more effective than other methods. Our main work can be summarized as follows:

(1) An MLP network is constructed by using raw vibration data and HI to achieve real-time automatic HI calculation.

(2) A BiLSTM-based method for bearing life prediction is proposed. It captures the hidden long-term correlation between sequential signals, achieving HI prediction and lifespan prediction.

(3) The proposed method is experimentally verified on the XJTU-SY bearing dataset and compared with other prediction methods to verify its superiority.

The rest of this article is organized as follows. Section 2 introduces the basic theory of MLP and BiLSTM. Section 3 presents the proposed life prediction method in detail. Section 4 shows the experimental results, and Section 5 gives the conclusion.

2 Theoretical Introduction

2.1 Multilayer Perceptron (MLP)

Multilayer perceptron is a typical artificial neural network that can segment complex spaces and fit strongly nonlinear data by combining multiple multilayer perceptrons [17]. It is the most common structure in artificial neural networks, which has a great advantage in analyzing data sets with a large number of continuous numerical descriptors. The MLP model structure is shown in Fig. 1.

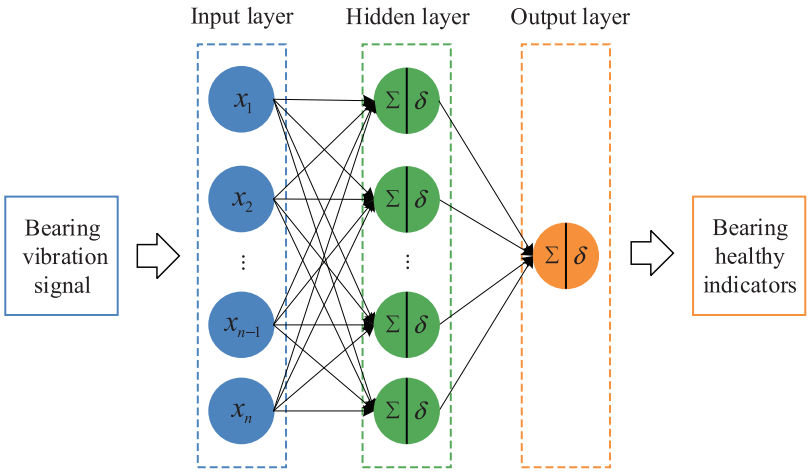

Figure 1: MLP model

In this model, the input values are connected to the hidden layer through the input layer. The hidden layer accepts all inputs and performs the appropriate actions. The resulting data is fitted as input to the next layer. These layers are completely connected. The calculation is as follows:

$$y_i = f(u_i) = f\left(\sum_{i=1}^{n} w_i x_i + b_i\right) \tag{2}$$

where x_i represents the input vector of node i or the calculated value of previous layer, w_i and b_i represent the coefficients learned at node i of the MLP model, and $f(\cdot)$ represents a non-linear activation function ReLU.

2.2 Long Short-Term Memory Networks (LSTM)

Long short-term memory networks (LSTM) are widely used in many related domains and are capable of capturing the long-term correlation of temporal signals [18]. Nevertheless, a typical LSTM can only "listen" to temporal dependencies in one direction, either backwards or forwards. In order to explore the bidirectional dependency of time series data, LSTM is combined with a bidirectional recursive network structure to construct a model.

Fig. 2 depicts a typical BiLSTM structure, including an input layer, hidden layer, and output layer. The input layer of the network receives the time series signals, which are then extracted by the hidden layer. The hidden layer also predicts future time series signals and outputs them through the output layer. Every hidden layer, as seen in Fig. 2, consists of two common hidden layers that process the input both forwards and backwards (the symbol represents the sum) in order to capture the features that exist before and after. The following is a description of the hidden layer for forward processing output:

$$\begin{cases} \overrightarrow{i}_t = \sigma\left(\overrightarrow{W}_{xi}x_t + \overrightarrow{W}_{hi}\overrightarrow{h}_{t-1} + \overrightarrow{b}_i\right) \\ \overrightarrow{f}_t = \sigma\left(\overrightarrow{W}_{xf}x_t + \overrightarrow{W}_{hf}\overrightarrow{h}_{t-1} + \overrightarrow{b}_f\right) \\ \overrightarrow{\hat{c}}_t = \varphi\left(\overrightarrow{W}_{xc}x_t + \overrightarrow{W}_{hc}\overrightarrow{h}_{t-1} + \overrightarrow{b}_c\right) \\ \overrightarrow{c}_t = \overrightarrow{f}_t \odot \overrightarrow{c}_{t-1} + \overrightarrow{i}_t \odot \overrightarrow{\hat{c}}_t \\ \overrightarrow{o}_t = \sigma\left(\overrightarrow{W}_{xo}x_t + \overrightarrow{W}_{ho}\overrightarrow{h}_{t-1} + \overrightarrow{b}_o\right) \\ \overrightarrow{h}_t = \overrightarrow{o}_t \odot \varphi(\overrightarrow{c}_t) \end{cases} \tag{3}$$

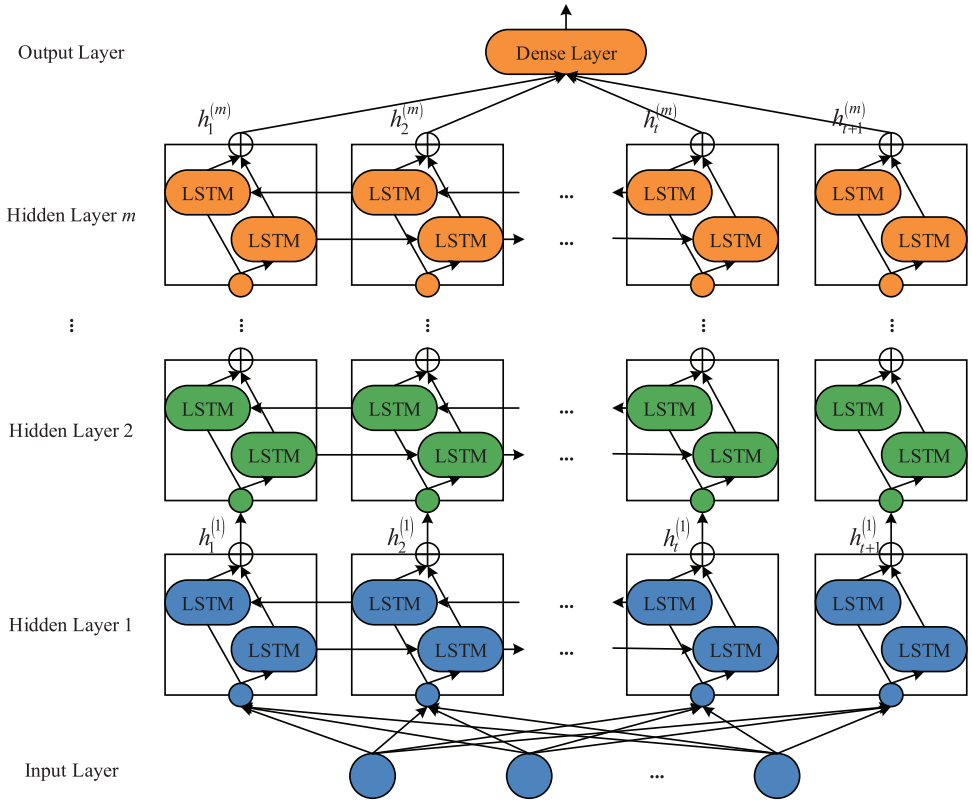

Figure 2: BiLSTM model

The output of each hidden layer of backward processing can be described as:

$$\begin{cases} \overleftarrow{i}_t = \sigma\left(\overleftarrow{W}_{xi}x_t + \overleftarrow{W}_{hi}\overleftarrow{h}_{t+1} + \overleftarrow{b}_i\right) \\ \overleftarrow{f}_t = \sigma\left(\overleftarrow{W}_{xf}x_t + \overleftarrow{W}_{hf}\overleftarrow{h}_{t+1} + \overleftarrow{b}_f\right) \\ \overleftarrow{\hat{c}}_t = \varphi\left(\overleftarrow{W}_{xc}x_t + \overleftarrow{W}_{hc}\overleftarrow{h}_{t+1} + \overleftarrow{b}_c\right) \\ \overleftarrow{c}_t = \overleftarrow{f}_t \odot \overleftarrow{c}_{t+1} + \overleftarrow{i}_t \odot \overleftarrow{\hat{c}}_t \\ \overleftarrow{o}_t = \sigma\left(\overleftarrow{W}_{xo}x_t + \overleftarrow{W}_{ho}\overleftarrow{h}_{t+1} + \overleftarrow{b}_o\right) \\ \overleftarrow{h}_t = \overleftarrow{o}_t \odot \varphi(\overleftarrow{c}_t) \end{cases} \tag{4}$$

where $\overrightarrow{h}_t$ and $\overleftarrow{h}_t$ represent the forward and backward outputs of neurons, $\overrightarrow{i}_t$, $\overrightarrow{f}_t$, $\overrightarrow{o}_t$ and $\overleftarrow{i}_t$, $\overleftarrow{f}_t$, $\overleftarrow{o}_t$ are the forward and backward processing input gate, forgetting gate, and output gate, respectively. $\overrightarrow{W}_{xi}$, $\overrightarrow{W}_{hi}$, $\overrightarrow{W}_{xf}$, $\overrightarrow{W}_{hf}$, $\overrightarrow{W}_{xc}$, $\overrightarrow{W}_{hc}$, $\overrightarrow{W}_{xo}$, $\overrightarrow{W}_{ho}$ and $\overleftarrow{W}_{xi}$, $\overleftarrow{W}_{hi}$, $\overleftarrow{W}_{xf}$, $\overleftarrow{W}_{hf}$, $\overleftarrow{W}_{xc}$, $\overleftarrow{W}_{hc}$, $\overleftarrow{W}_{xo}$, $\overleftarrow{W}_{ho}$ represent the weight matrix for input data and recursive data. $\overrightarrow{b}_i$, $\overrightarrow{b}_f$, $\overrightarrow{b}_c$, $\overrightarrow{b}_o$ and $\overleftarrow{b}_i$, $\overleftarrow{b}_f$, $\overleftarrow{b}_c$, $\overleftarrow{b}_o$ represent forward and backward deviations, and σ represent the sigmoid and tanh activation functions, $\odot$ represents Hadamard products.

3 The Proposed Method

In this paper, a bi-directional long-term and short-term memory network based on HI is proposed for rolling bearing health prediction. As shown in Fig. 3, this method can be implemented in the following steps:

Step 1: Design experiments to obtain bearing vibration data. XJTU-SY bearing data set is selected in this paper to provide the real degradation data for the whole service life of the bearing under different working conditions.

Step 2: Health indicator construction. The rolling bearing degradation process is adequately described by fusing statistical features taken from observed signals using the KPCA-EWMA approach.

KPCA-EWMA fuses statistical features extracted from measured signals to describe the degradation process of rolling bearings effectively.

Step 3: Calculation of health indicators. By establishing the MLP network model, the common hidden features of different bearing original vibration signals are mined, and the HI of online bearing is calculated automatically.

Step 4: Life prediction. The PSO algorithm optimizes the BiLSTM network model of online bearing, and the remaining bearing service life is obtained by the iterative prediction method.

3.1 Health Indicator Design

As indicated in Table 1, many features are extracted using time-domain (TD), frequency-domain (FD), and time-frequency domain (TFD) techniques in order to create a thorough description of the rolling bearings' degrading characteristics. Time-domain features are represented by TD1 through TD16, frequency-domain features by FD1 through FD16, maximum value, root mean square error, and signal-to-noise ratio (RMS) of the second frequency band signal obtained by three-layer wavelet packet transform, respectively, and inherent energy of the first component after VMD decomposition by TFD1~TFD3. Selecting so many statistical features to provide a more comprehensive description of the bearing degradation process is beneficial in avoiding the influence of manual experience on HI construction.

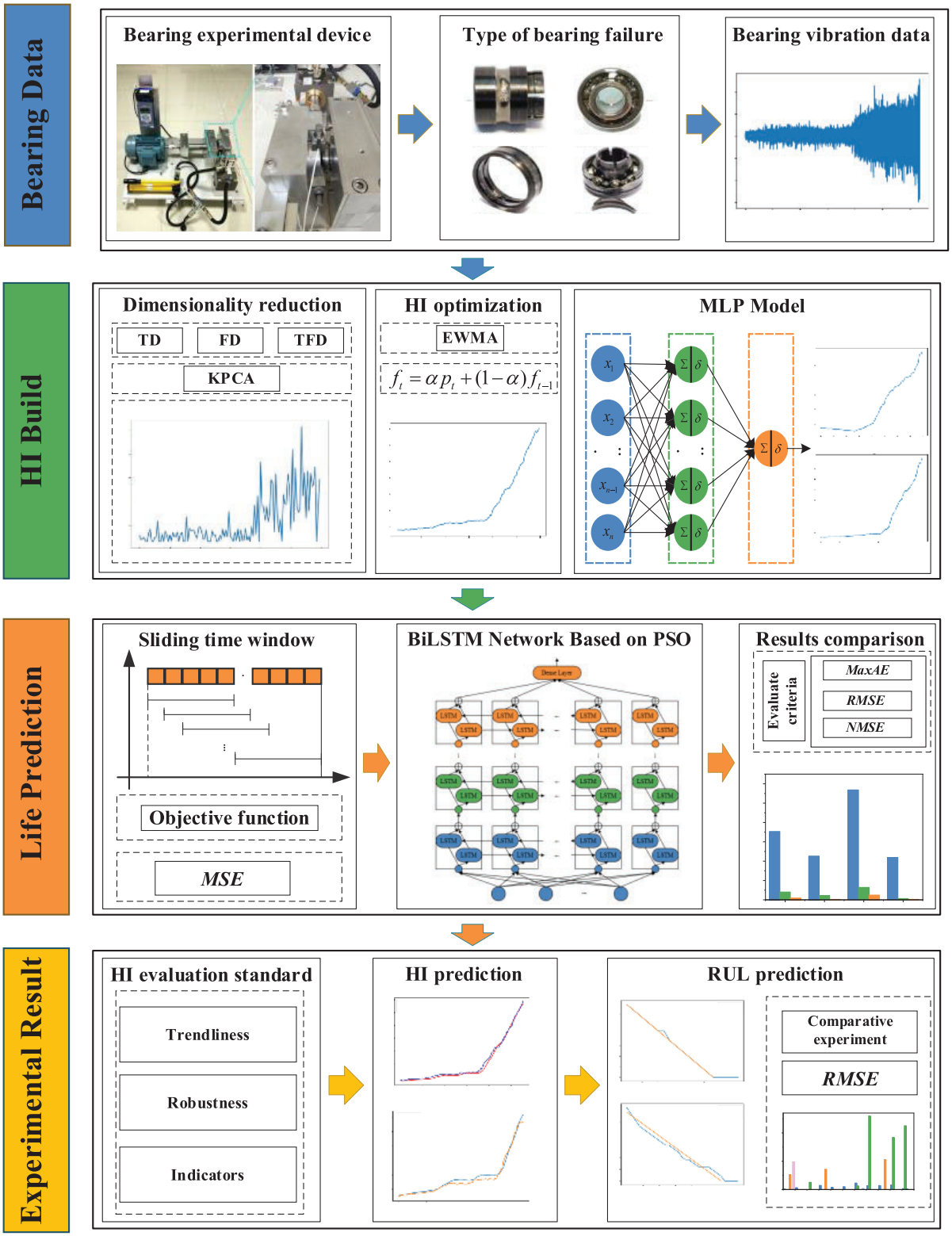

Figure 3: Prognosis procedures of the proposed method

Table 1: Feature indexes and expressions

TD	$\bar{x}=\frac{1}{N}\sum_{i=1}^{N}x_i$	$\sigma=\sqrt{\frac{\sum_{i=1}^{N}[x_i-\bar{x}]^2}{N-1}}$	$x_{SRA}=\left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{\lvert x_i\rvert}\right)^2$	$\delta=\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N-1}$
	$x_{\max}=\max(x_n)$	$x_{rms}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2}$	$x_p=\max(\lvert x_n\rvert)$	$x_{pp}=x_{\max}-x_{\min}$
	$x_{AE}=\frac{1}{N}\sum_{i=1}^{N}\lvert x(i)\rvert$	$x_{SA}=\sqrt{\frac{1}{N}\sum_{n=1}^{N}\sqrt{\lvert x(n)\rvert}}$	$S=E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$	$K=E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$
	$F_P=\frac{x_{\max}}{x_{rms}}$	$F_C=\frac{x_{\max}}{x_{SRA}}$	$F_W=\frac{x_{rms}}{x_{AE}}$	$F_I=\frac{x_{\max}}{x_{AE}}$
FD	$\bar{s}=\frac{1}{N}\sum_{i=1}^{N}s_i$	$\delta=\frac{\sum_{i=1}^{N}(s_i-\bar{s})^2}{N-1}$	$s_{rms}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}s_i^2}$	$\sigma=\sqrt{\frac{\sum_{i=1}^{N}(s_i-\bar{s})^2}{N-1}}$
	$F_1=\frac{\sum_{i=1}^{N}(s_i-\bar{s})^3}{N\sigma^3}$	$F_2=\frac{\sum_{i=1}^{N}(s_i-\bar{s})^4}{N\sigma^4}$	$F_3=\frac{\sum_{i=1}^{N}f_is_i}{\sum_{i=1}^{N}s_i}$	$F_4=\sqrt{\frac{\sum_{i=1}^{N}(f_i-F_3)^2s_i}{N}}$
	$F_5=\sqrt{\frac{\sum_{i=1}^{N}f_i^2s_i}{\sum_{i=1}^{N}s_i}}$	$F_6=\sqrt{\frac{\sum_{i=1}^{N}f_i^4s_i}{\sum_{i=1}^{N}f_i^2s_i}}$	$F_7=\frac{\sum_{i=1}^{N}f_i^2s_i}{\sqrt{\sum_{i=1}^{N}s_i\sum_{i=1}^{N}f_i^4s_i}}$	$F_8=\frac{F_4}{F_3}$
	$F_9=\frac{\sum_{i=1}^{N}s_i(f_i-F_3)^3}{NF_4^3}$	$F_{10}=\frac{\sum_{i=1}^{N}s_i(f_i-F_3)^4}{NF_4^4}$	$F_{11}=\frac{\sum_{i=1}^{N}f_i\sqrt{s_i-F_3}}{N\sqrt{F_4}}$	$F_{12}=\frac{\sqrt{\sum_{i=1}^{N}\sqrt{f_i}}\Big/N}{\sum_{i=1}^{N}\lvert f_i\rvert}$
TFD	$TFD_1=\max(\omega c_i)$	$TFD_2=\sqrt{\sum_{i=1}^{N}\omega c_i^2}\Big/N$	$TFD_3=\left(\sum_{i=1}^{N}\sqrt{\lvert\omega c_i\rvert}\Big/N\right)^2$	$TFD_4=\sum_{i=1}^{N}IMF_1(i)^2\Big/N$

However, these features have characteristics such as different ranges, high dimensionality, and high nonlinearity. As a result, a suitable technique must be created to extract the most important data from these original traits. KPCA is a simple and efficient technique for reducing nonlinear dimensionality. It involves two steps to complete:

(1) Projecting the initial features onto a new high-dimensional feature space using nonlinear kernel functions.

(2) Principal components are extracted from a high-dimensional feature space using linear PCA techniques.

The first principal component was selected as HI to characterise the degradation process. Although HI maintains global monotonicity, it typically contains a large number of local random fluctuations, which may interfere with the life prediction of rolling bearings. As a result, using certain technologies to improve HI is meaningful.

For smoothing online bearing HI, the EWMA approach based on historical data time series smoothing is adopted. The following is how its mechanism is explained:

$$f_t = \alpha\left(p_t + \beta p_{t-1} + \beta^2 p_{t-2} + \cdots + \beta^{t-1} p_1\right) \tag{5}$$

where α represents the smoothing parameter ranging from 0 and 1, and the value of β is $(1-\alpha)$; p_t stands for the current observation, f_t stands for its estimated value, which is the value of modified HI at time t; p_{t-1}, p_{t-2}, $\cdots$, p_1 are the historical observations. EWMA can eliminate short-term fluctuations, maintain long-term trends accurately, and effectively capture abrupt changes. As a result, EWMA is an appropriate method to change HI. In addition, the simplified calculation formula for reducing the computational demand of EWMA can be expressed as follows:

$$f_t = \alpha p_t + (1-\alpha) f_{t-1} \tag{6}$$

3.2 Calculation of Health Indicators

On this basis, an MLP model is constructed to calculate HI. The MLP model can automatically extract representative features from the original vibration signal without the need for manual feature extraction. the framework of the HI calculation method based on the MLP model is shown in Fig. 4. The proposed method consists of three stages: (1) Extract HI for training rolling bearings, as mentioned in the previous section. (2) Train the MLP model. Train the MLP model using the original vibration signal as input and HI value as the target output. Typical characteristics are captured from the input original vibration signal through two full connection layers and the ReLU activation function. Then, the fully connected layer is utilized as the regression layer to generate predictive output (HI label). (3) Calculate and test the HI value of rolling bearings. After the training process is completed, the online vibration signal of the test bearing can be input into the learned MLP model. The MLP model determines the current HI value by directly capturing typical festures from the original vibration signal.

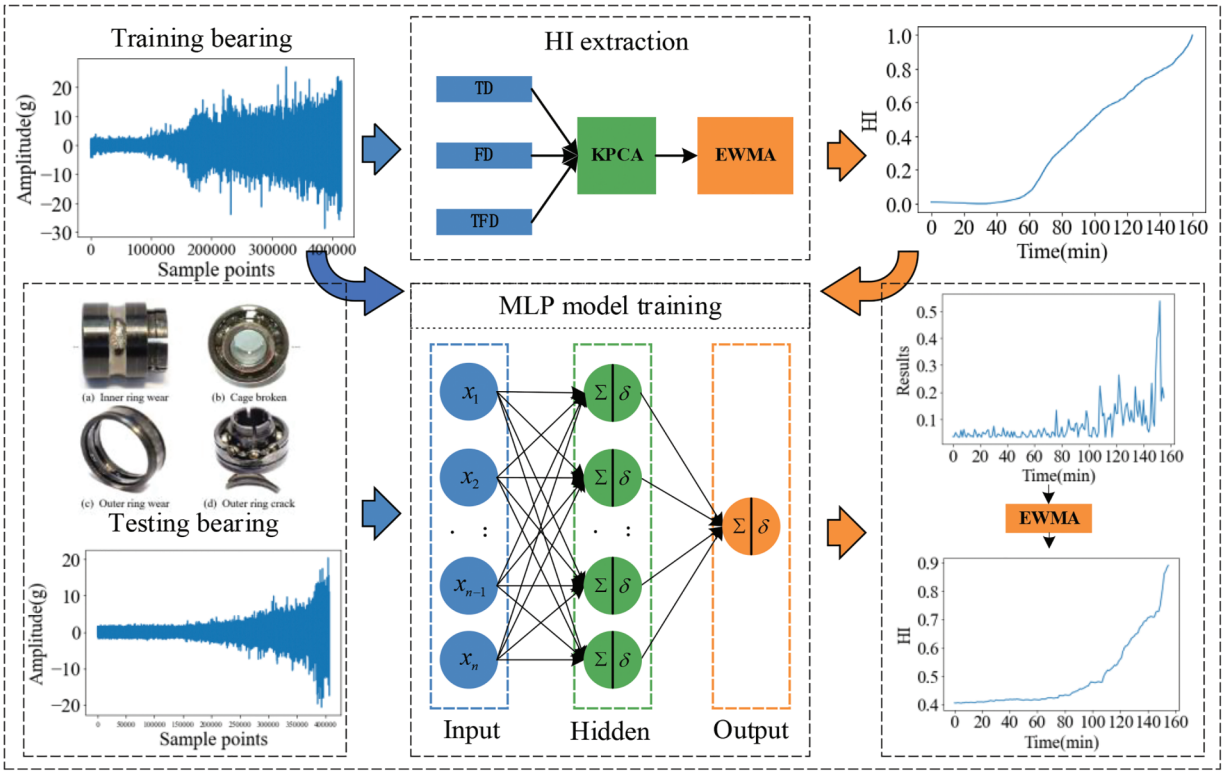

Figure 4: Framework of the MLP-based HI estimation method

3.3 Life Prediction

In this section, the BiLSTM model is employed to predict the future HI of rolling bearings, and the model parameters are optimized through PSO.

The BiLSTM model consists of H hidden layers and K units in each hidden layer and includes training and testing steps. In the training steps, the history HI values of the bearing are utilized to construct the training dataset. The construction of the training dataset adopts sliding time window technology. The sliding time window technology process is shown in Fig. 5, which uses a fixed length time window to sample HI values continuously. Then, the training dataset can be formed as $\{X_t, y_t\}_{t=1}^{T}$, where $X_t = [h_t, h_{t+1}, \cdots, h_{t+w}]$ is the i-th training sample vector, where h_t represents the HI value of the training rolling bearing at time t, and w is the length of the time window. $y_t = h_{t+w+1}$ is the corresponding label. Train BiLSTM by minimizing the mean square error (MSE) function, which can be represented as:

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2 \tag{7}$$

where y_t and $\hat{y}_t$ are actual labels and predicted labels. T represents the total number of training samples. In order to quickly converge the training process, the BiLSTM model is equipped with an Adam optimizer.

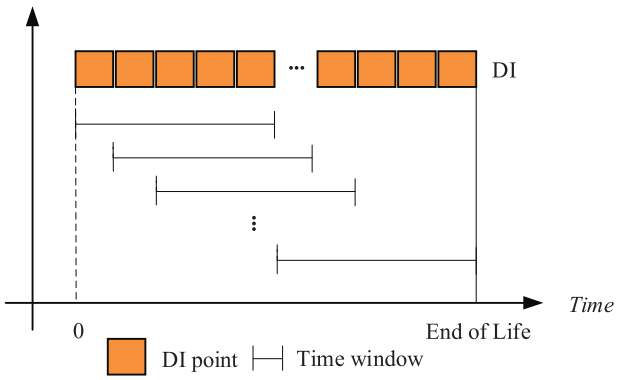

Figure 5: Processing method for sliding time windows illustrated

4 Experimental Verification

4.1 Data Set

The XJTU-SY rolling bearing dataset is a set of vibration state signals for the entire life cycle of rolling bearings [19]. The sampling interval is set to 1 minute, with each sampling lasting 1.28 s, and each sample contains 32,768 data points. The horizontal vibration signals are employed in this study as they contain more useful information. The dataset includes 15 sets of bearing data, which were obtained under three different operating conditions. Table 2 shows the total number of samples and failure locations for all bearings. When training MLP, the following four bearings are selected as the training set: bearings 1_2, 1_4, 3_2, and 3_3. When training BiLSTM, the data of the health stage and the initial degradation stage of the remaining bearings are used as the training set, and the data of the degradation stage is utilized as the test set.

4.2 HI Design

According to the expression in Table 1, 36 statistical features are extracted from the time domain, frequency domain, and time-frequency domain to constitute the original feature set. Then, the KPCA is applied to combine these features, and the first principal component after fusion is used as a HI to

describe the degradation process of bearing. It is worth noting that the Gaussian kernel function is employed as the kernel function, the kernel radius is chosen as 0.01, and the number of principal components to retain is set to 3. Then, EWMA was employed to further modify HI, with a smoothing parameter selection of 1/50. Fig. 6 shows the raw vibration data and health indicators of bearings 1_2 and 3_2 used to construct the HI. HI is relatively stable in the initial phase, starts to decline in the intermediate stage, and rises rapidly towards the end of the bearing's service life. The results show that the HI can effectively capture the degradation trend, and the actual degradation process of the rolling bearing is described. The HI of bearing 3_2 fluctuates repeatedly in the rapid rise stage. Still, the deterioration process of bearing is irreversible and should not have a downward trend, which reflects the limitations of artificial feature extraction.

Table 2: Working environment and fault situation of bearings

Operating condition	Speed/(r/min)	Radial force/(kN)	Bearing dataset	Fault element
1	2100	12	Bearing1_1	Outer race
			Bearing1_2	Outer race
			Bearing1_3	Outer race
			Bearing1_4	Cage
			Bearing1_5	Inner race and outer race
2	2250	11	Bearing2_1	Inner race
			Bearing2_2	Outer race
			Bearing2_3	Cage
			Bearing2_4	Outer race
			Bearing2_5	Outer race
3	2400	10	Bearing3_1	Outer race
			Bearing3_2	Inner race, ball, cage and outer race
			Bearing3_3	Inner race
			Bearing3_4	Inner race
			Bearing3_5	Outer race

4.3 MLP Model

An MLP model was constructed for the HI calculation of online rolling bearings. The structure of the MLP model greatly affects the performance of the network. The 2560 central data points from the original vibration signal of the bearing are employed as input to the MLP model. The HI calculated by KPCA-EWMA is selected as the training label of the MLP model. The ReLU activation function is used between fully connected layers. During training, the MSE function is used as the loss function of MLP. The optimal model parameters are obtained after 50 epochs using the Adagrad optimizer.

To effectively evaluate HI and screen suitable network hyperparameters, an effective evaluation method is needed. However, most existing methods directly assess HI itself. Since the performance degradation of rolling bearings is a stochastic process, HI can be divided into trend indicators and residual parts to better measure the degradation of rolling bearings. Therefore, a rolling bearing HI evaluation criterion that integrates trend and robustness is proposed, which can be expressed as:

$$\begin{cases} Y(t_k) = Y_T(t_k) + X_R(t_k) \\ V_{\text{corr}}(Y(t_k), T(t_k)) = \dfrac{\left|\sum_{k=1}^{K}\left(Y(t_k)-\tilde{Y}\right)\left(T(t_k)-\tilde{T}\right)\right|}{\sqrt{\sum_{k=1}^{K}\left(Y(t_k)-\tilde{Y}\right)^2\sum_{k=1}^{K}\left(T(t_k)-\tilde{T}\right)^2}} \\ V_{rob}(Y(t_k)) = \exp\left(\dfrac{-\text{std}(X_R(t_k))}{\text{mean}\left|Y(t_1)-Y(t_K)\right|}\right) \\ V = \omega_1 V_{\text{corr}}(Y(t_k), T(t_k)) + \omega_2 V_{rob}(Y(t_k)) \end{cases} \tag{8}$$

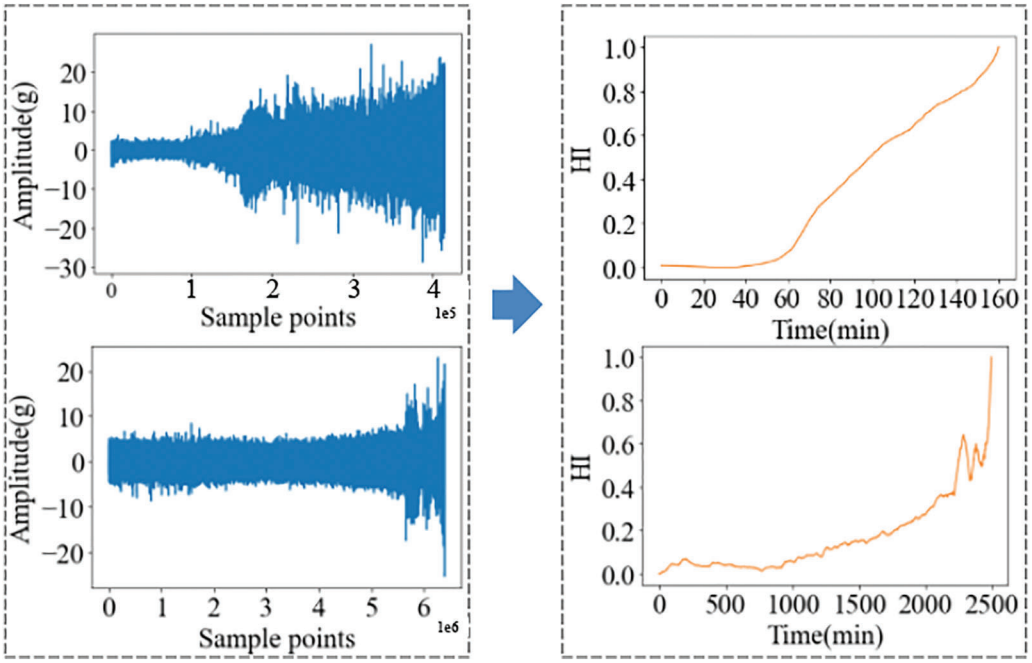

Figure 6: Raw vibration data and HI of bearings 1_2 and 3_2

The HI curve is decomposed into the trend part $Y_T(t_k)$ and residual part $X_R(t_k)$, k is the length of the time series, $Y(t_k)$ is the HI. $V_{\text{corr}}(Y(t_k),\ T(t_k))$ is the trend value, $\tilde{Y}$ and $\tilde{T}$ are the mean values of health index $Y(t_k)$ and time vector $T(t_k)$, $V_{rob}(Y(t_k))$ is the robustness value, $Y(t_1)$ and $Y(t_K)$ are the starting and ending values of the HI $Y(t_k)$. ω_1 and ω_2 are two weight coefficients greater than zero and with a sum of 1. According to the effect of HI on the prediction of remaining life, the trend and robustness of HI were weighted by 0.5 and 0.5.

Taking bearing 1 as an example, experimental verification of HI and MLP model construction was carried out. The comparison of HI constructed by KPCA-EWMA and HI constructed by MLP models with different parameters are shown in Table 3 and Fig. 7. The results indicate that the HI constructed by the MLP model has signifcant advantages in trend and robustness, and the HI score of the MLP model is the highest when using the 2560-160-16-1 parameter. Therefore, the model trained with this parameter is employed to carry out subsequent calculations.

Using the remaining 11 bearing vibration data as the online data, each sampled vibration data is sequentially input into the trained MLP model and then smoothed by EWMA to obtain the following results. As shown in Table 4, all bearings have high HI scores, which fully reflects the effectiveness of

the model. Fig. 8 shows the HI results of the three bearings under condition 1. f It can be seen that HI is relatively stable in the healthy state of the bearings and gradually increases after degradation begins, showing a monotonic upward trend overall, effectively capturing its degradation trend.

Table 3: Evaluation indicators for different methods

Method		Trendliness	Robustness	Indicators
KPCA-EWMA		0.375	0.967	0.671
MLP model	2560-1600-16-1	0.266	0.975	0.620
	2560-1280-80-1	0.677	0.933	0.805
	2560-160-16-1	0.693	0.935	0.814

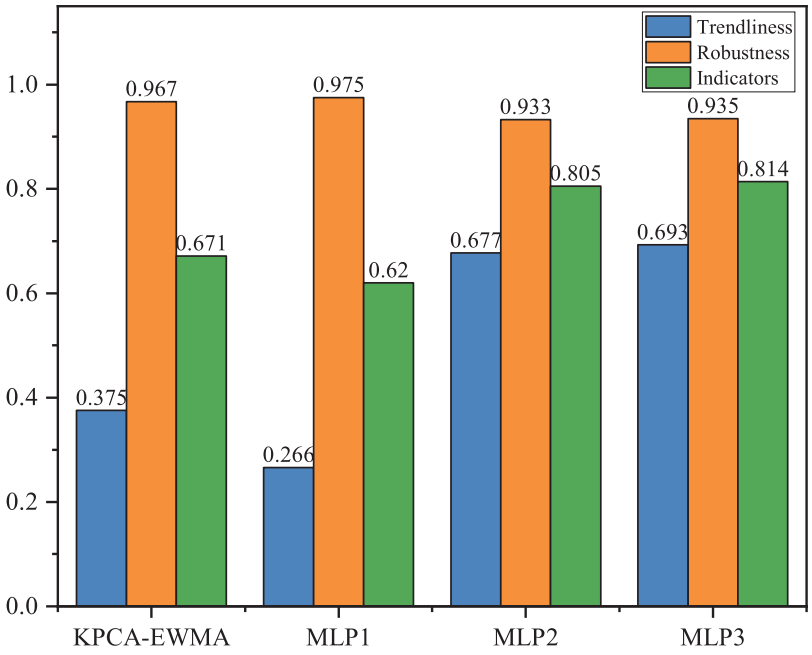

Figure 7: Evaluation indicators for different methods

Table 4: Evaluation indicators for test set bearings

Bearing	Trendliness	Robustness	Indicators
1_1	0.995	0.966	0.980
1_3	0.992	0.955	0.974
1_5	0.693	0.935	0.814
2_1	0.592	0.888	0.740
2_2	0.878	0.972	0.925
2_3	0.997	0.963	0.980

(Continued)

Table 4 (continued)

Bearing	Trendliness	Robustness	Indicators
2_4	0.996	0.965	0.981
2_5	0.844	0.968	0.906
3_1	0.313	0.910	0.612
3_4	−0.002	0.933	0.465
3_5	0.999	0.985	0.992

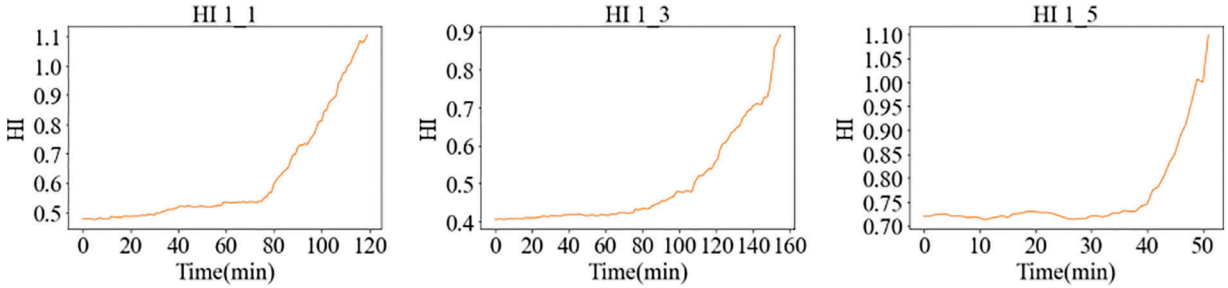

Figure 8: HI results of the remaining three bearings under condition 1

4.4 BiLSTM Model

A BILSTM model for predicting future HI has been established. In order to comprehensively evaluate the performance of the proposed method, three metrics are adopted, including normalized mean square error (NMSE), root mean square error (RMSE), and maximum absolute error (MaxAE), defined as follows:

$$NMSE = \frac{\sum_{t=1}^{T} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{T} \hat{y}_t^2} \tag{9}$$

$$RMSE = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2} \tag{10}$$

$$MaxAE = \max(|y_t - \hat{y}_t|) \tag{11}$$

where $\hat{y}_t$ and y_t represent actual and predicted values, respectively.

4.4.1 PSO Optimization Hyperparameter

Taking bearing 1_1 as an example, 80% of its HI is used as training data, and the last 20% is used as testing data. The results processed by sliding time window technology are used as training and testing sets. During training, the MSE function is used as the loss function of BiLSTM. The optimal model parameters are obtained after 50 epochs using the Adam optimizer.

The number of hidden layers H, and the unit K in each hidden layer have a significant impact on the model performance. In this paper, these two essential hyperparameters are optimized by the bird swarm algorithm. The bird swarm particles can be set to 10, iterations to 50, hidden layers to 3, and the number

of hidden layers to [1, 100] to predict the bearing directly. Table 5 compares BiLSTM and PSO-optimized models with different H and K values. Table 5 shows that when the H and K values are different, the performance of BiLSTM will vary. Obviously, the PSO optimized BiLSTM model has the best performance.

Table 5: Performance of the proposed method with different architectures

Model structure	MaxAE	RMSE	NMSE
5-20-40-10-1	0.0707	0.0081	0.0019
5-10-20-10-1	0.0453	0.0044	0.0005
5-20-10-5-1	0.1137	0.0129	0.0049
5-50-10-26-1 (PSO)	0.0438	0.0013	0.0005

The size of the sliding window is another important factor that affects the performance of the model. In this article, the window size is set to 5. The comparison between the actual HI of bearing 1_1 and the one-step expected HI is shown in Fig. 9. The projected HI is very similar to the actual HI. The experimental details and fitting errors for all tested bearings are shown in Table 6. The largest error is the MaxAE of bearing 3_4, which is only 0.3473. The results indicate that the model has significant advantage in capturing long-term correlations hidden in time series signals and mining bearing degradation trends, demonstrating the effectiveness and reliability of the proposed health prediction method.

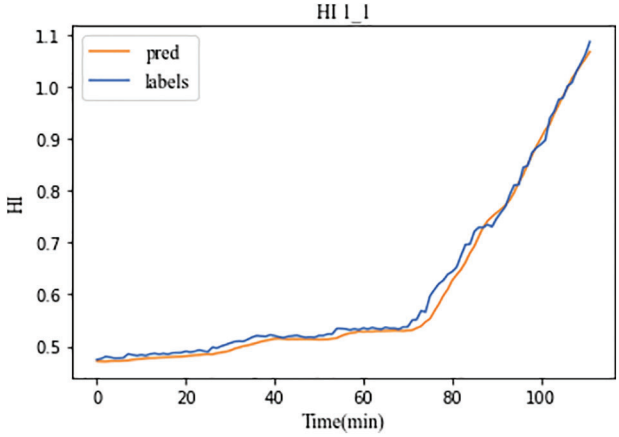

Figure 9: The comparison between single step predicted HI and actual HI of bearings 1_1

Table 6: The MaxAE, RMSE, and NMSE values for all predicted HI

Bearing	Training data	Testing data	Model structure (PSO)	MaxAE	RMSE	NMSE
1_1	95	25	50-10-26	0.0438	0.0013	0.0005
1_3	120	36	24-41-5	0.0860	0.0009	0.0005
1_5	42	8	1-100-97	0.0604	0.0164	0.0008
2_1	468	20	50-14-45	0.0340	0.0003	0.0003

(Continued)

Table 6 (continued)

Bearing	Training data	Testing data	Model structure (PSO)	MaxAE	RMSE	NMSE
2_2	120	40	17-50-50	0.0826	0.0019	0.0007
2_3	482	50	70-16-100	0.1109	0.0083	0.0008
2_4	32	8	21-59-54	0.0208	0.0118	0.0004
2_5	286	50	62-12-65	0.3226	0.0252	0.0035
3_1	2486	50	66-30-88	0.3196	0.0175	0.0127
3_4	1462	50	30-100-90	0.3473	0.0200	0.0092
3_5	62	50	75-55-89	0.2670	0.0175	0.0027

4.4.2 Life Prediction

The bearing test data is input into the PSO optimization model for iterative prediction until the predictes HI reached the fault threshold. The number of steps in iteration is the remaining life, and the predicted result is shown in Fig. 10. The results indicate that the predicted RUL almost coincides with the actual RUL, and and over time, the predicted ULL becomes closer to the actual RUL.

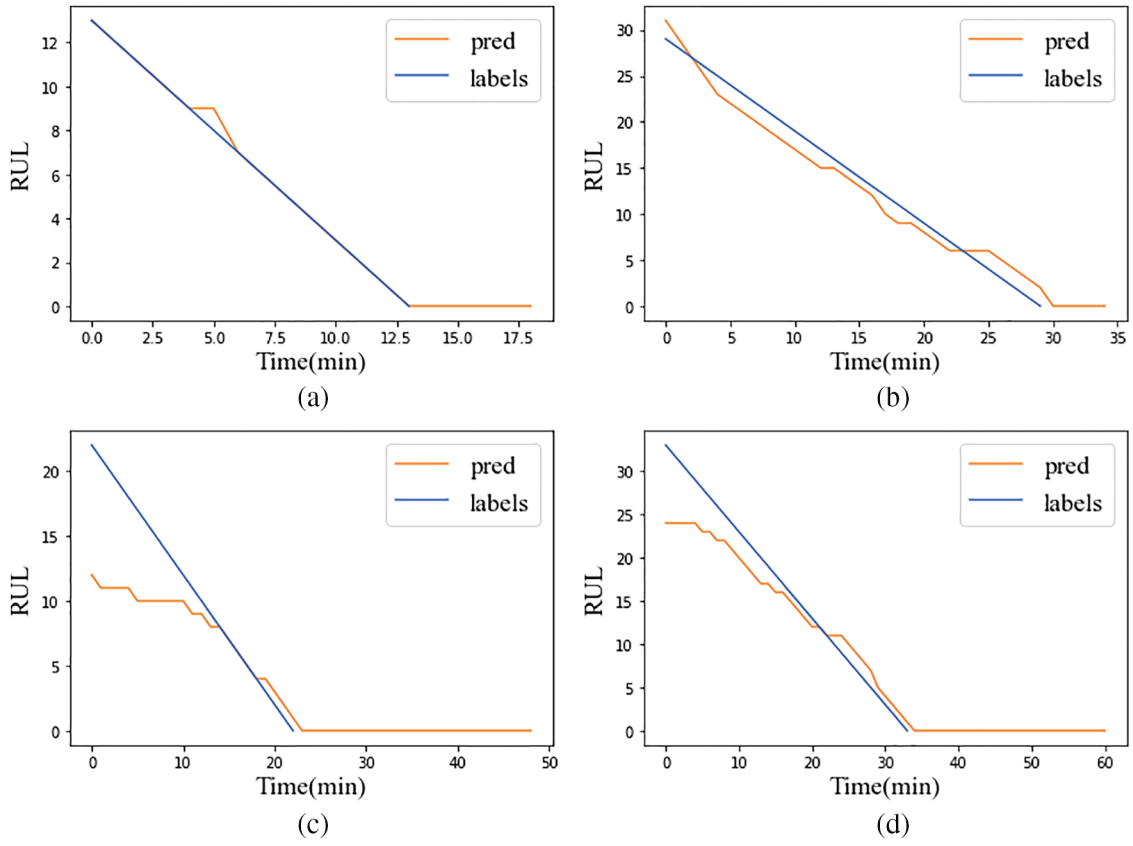

Figure 10: Partial life prediction results: (a) Bearing1_1 (b) Bearing 1_3 (c) Bearing 2_3 (d) Bearing 3_5

Comparative Experiment 1:

In order to verify the effectiveness of the health indicators constructed by the proposed MLP model for bearing life prediction, the artificial indicators constructed by downscaling and fusing statistical features through KPCA-EWMA are used as the bearing health indicators, and the training set is also built by the sliding time window method to train the BiLSTM network model, and the PSO algorithm is used to optimize the structure of the model, and then the residual life prediction results are compared with the above methods were compared. The RUL prediction results of the two methods are shown in Fig. 11. From the figure, it can be seen that the remaining life results predicted by the MLP method at different time points are more accurate, while the artificial indicator method shows a larger deviation. In addition, by calculating the RMSE of the remaining life prediction results of the two methods, the RMSE of the manual indicator method is 2.1023, which is much higher than the 0.0556 generated by the MLP method, further proving that the health indicators generated by the MLP network have a positive impact on the accuracy of life prediction. The neural network architecture of the MLP model has strong feature learning and nonlinear mapping capabilities, which better capture the dynamic changes during the degradation process of bearing. The generated health indicators can more accurately describe the current health status of the bearing and are more effective.

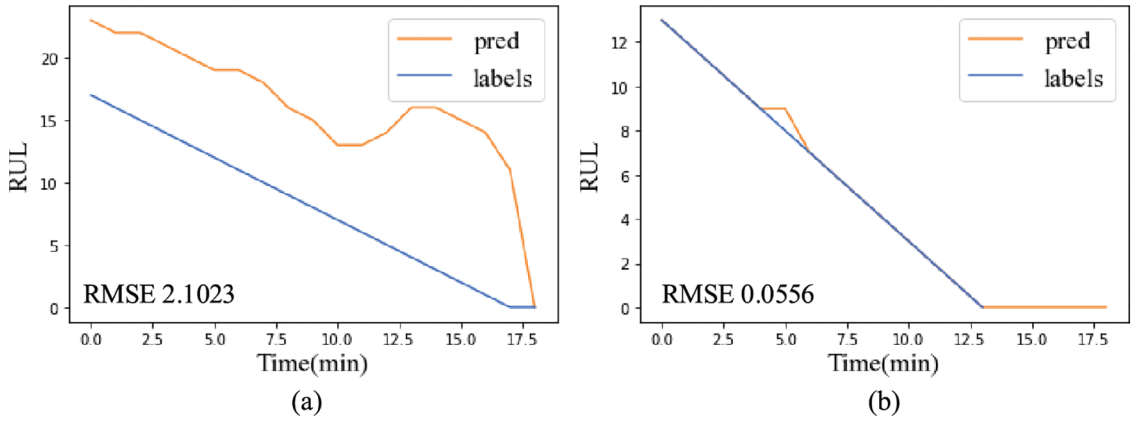

Figure 11: Comparison of bearing 1_1 (a) KPCA-EWMA (b) MLP

Comparative Experiment 2:

In this section, to demonstrate the superiority of the proposed method, the predicted results of the proposed method were compared with the results of five published studies, including the CNN-BiLSTM network of [20], TCN-RSA of [21], RBM-BiLSTM of [22], CNN-SRU of [23] and Bayesian of [24]. Each of these studies selected a portion of the bearings in the dataset as the training set and another portion as the test set, which is slightly different from the proposed method. In this method, the data of each bearing at the beginning of the normal state and degradation stage are used as the training set, and the data that rapidly degradates to the fault stage is used as the testing set. Among them, Cheng et al. build a CNN model to calculate the health indicator value of the bearing, and then use the health indicator to make a BiLSTM model for future health indicators and remaining life prediction. Cao et al. build a temporal convolutional network with a self-attention mechanism to obtain the feature contributions of different moments during the degradation of bearings, and then achieve the remaining life prediction. Hou et al. construct health metrics through an unsupervised learning approach and then obtain the remaining life expectancy through a difference-based target label generation method. Yao et al. combine an improved one-dimensional convolutional neural network and a simple regression unit to

predict the remaining life of bearings. Gao et al. fuse time-domain features to establish health indicators and then build a Bayesian model for the prediction of the remaining life of bearings. The performance comparison between the proposed method and five published studies on bearing life prediction based on the XJTU-SY dataset is detailed in Table 7 and Fig. 12.

Table 7: RMSE comparison with different methods

RMSE	Proposed approach	CNN-BiLSTM	TCN-RSA	RBM-BiLSTM	CNN-SRU	Bayesian
Bearing 1_1	0.0556	–	–	2.62	–	4.8843
Bearing 1_3	0.3093	–	–	–	–	–
Bearing 1_5	0	1.27	–	–	–	–
Bearing 2_1	0.7158	–	–	3.56	–	–
Bearing 2_2	0.3859	–	–	–	–	–
Bearing 2_3	0.4855	–	–	–	–	–
Bearing 2_4	1.1304	0.63	–	–	0.1699	–
Bearing 2_5	0.6738	12.97	–	–	–	–
Bearing 3_1	0.7009	–	–	5.27	–	–
Bearing 3_4	0.7753	9.18	–	–	–	–
Bearing 3_5	0.1573	11.24	0.0659	–	–	–

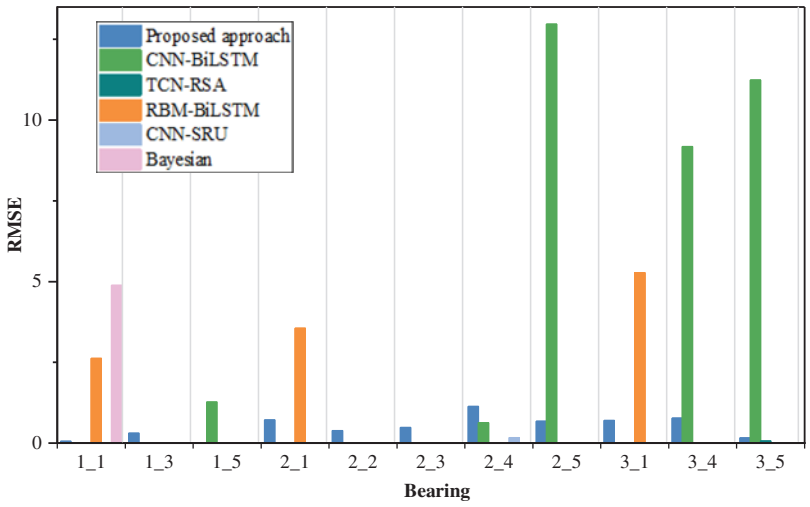

Figure 12: Comparison between the proposed approach and other literature results

From the table, it can be seen that although the prediction errors for bearings 2_4 and 3_5 are slightly higher than those of the CNN-BiLSTM method and the TCN-RSA method, the proposed method obtains smaller prediction errors for all the other bearings, indicating that the proposed method has stable prediction performance for bearings under different operating conditions and fault types. The results of Fig. 12 indicated that compared with other methods, the proposed method can obtain smaller RMSE values and more accurate remaining life prediction results in all bearings, with better accuracy and robustness.

5 Conclusion

In this paper, a bearing life prediction method based on the MLP and BILSTM network model is proposed. The MLP model is trained by raw bearing vibration data and HI constructed by the KPCA-EWMA method. The common hidden features of the original vibration signals of different bearings and the mapping relationship between HI can be effectively explored by this model. Then a rolling bearing HI prediction model based on particle swarm optimization of BiLSTM is established and experimentally validated on the XJTU-SY bearing dataset through experiments and comparisons, the following conclusions have been drawn:

(1) The MLP network is constructed by using raw vibration data and HI to achieve real-time automatic HI calculation. The model has generality and robustness and can be used immediately for similar other bearings.

(2) A bearing life prediction method based on BiLSTM is proposed, which is based on real-time HI of test bearings and can effectively capture the hidden long-term correlation between bearing time series state signals, achieving accurate prediction of HI and remaining life.

(3) During the training process, only a small number of bearing data samples are used to complete the model training, providing an effective technical solution for predicting the remaining life of bearings under conditions such as a small number of data samples, multiple operating conditions, and multiple types of faults. As a further research direction, it is worth studying how to determine the initial degradation point and fault threshold of different types of bearings, as well as how to predict the effective life of bearings that degrade too quickly after the initial degradation point.

Acknowledgement: None.

Funding Statement: This work was supported by the National Key Research and Development Project (Grant Number 2023YFB3709601), the National Natural Science Foundation of China (Grant Numbers 62373215, 62373219, 62073193), the Key Research and Development Plan of Shandong Province (Grant Numbers 2021CXGC010204, 2022CXGC020902), the Fundamental Research Funds of Shandong University (Grant Number 2021JCG008), the Natural Science Foundation of Shandong Province (Grant Number ZR2023MF100).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Yongfeng Tai, Xingyu Yan, Faye Zhang; data collection: Yongfeng Tai, Xiangyi Geng; algorithm design: Xingyu Yan, Lin Mu; analysis and interpretation of results: Yongfeng Tai, Xingyu Yan; draft manuscript preparation: Xingyu Yan, Mingshun Jiang, Faye Zhang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The authors confirm that the data supporting the findings of this study are available within the article.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Vrignat P, Kratz F, Avila M. Sustainable manufacturing, maintenance policies, prognostics and health management: a literature review. Reliability Eng Syst Saf. 2022;218:108140. doi:10.1016/j.ress.2021.108140.
2. Zhao L, Li Q, Feng J, Zheng S. Service life prediction method for wheel-hub-bearing under random multi-axial wheel loading. Eng Fail Anal. 2021;122:105211. doi:10.1016/j.engfailanal.2020.105211.

3. Chen C, Li B, Guo J, Liu Z, Qi B, Hua C. Bearing life prediction method based on the improved FIDES reliability model. Reliability Eng Syst Saf. 2022;227:108746. doi:10.1016/j.ress.2022.108746.
4. Rathore MS, Harsha SP. Rolling bearing prognostic analysis for domain adaptation under different operating conditions. Eng Fail Anal. 2022;139:106414. doi:10.1016/j.engfailanal.2022.106414.
5. Ni Q, Ji J, Feng K. Data-driven prognostic scheme for bearings based on a novel health indicator and gated recurrent unit network. IEEE Trans Ind Inform. 2023;19(2):1301–11. doi:10.1109/TII.2022.3169465.
6. Yang C, Ma J, Wang X, Li X, Li Z, Luo T. A novel based-performance degradation indicator RUL prediction model and its application in rolling bearing. ISA Trans. 2022;121:349–64. doi:10.1016/j.isatra.2021.03.045.
7. Li X, Jiang H, Xiong X, Shao H. Rolling bearing health prognosis using a modified health index based hierarchical gated recurrent unit network. Mech Mach Theory. 2019;133:229–49. doi:10.1016/j.mechmachtheory.2018.11.005.
8. Xu Z, Wang L, Liu Y, Cai K, Chen Z, Chen B. A prediction method for remaining life of rolling bearing using improved regression support vector machine. J Xi'an Jiaotong Univ. 2022;56(3):197–205 (In Chinese).
9. Li L, Xu J, Li J. Estimating remaining useful life of rotating machinery using relevance vector machine and deep learning network. Eng Fail Anal. 2023;146:107125. doi:10.1016/j.engfailanal.2023.107125.
10. Teng W, Zhang X, Liu Y, Andrew K, Ma Z. Prognosis of the remaining useful life of bearings in a wind turbine gearbox. Energies. 2018;10(1):10010032.
11. Yan M, Wang X, Wang B, Chang M, Isyaku M. Bearing remaining useful life prediction using support vector machine and hybrid degradation tracking model. ISA Trans. 2020;98:471–82. doi:10.1016/j.isatra.2019.08.058.
12. Jiang C, Liu X, Liu Y, Xie M, Liang C, Wang Q. A method for predicting the remaining life of rolling bearings based on multi-scale feature extraction and attention mechanism. Electronics. 2022;11:3616. doi:10.3390/electronics11213616.
13. Yang J, Peng Y, Xie J, Wang P. Remaining useful life prediction method for bearings based on LSTM with uncertainty quantification. Sensors. 2022;22:4549. doi:10.3390/s22124549.
14. Zhu H, Cheng J, Zhang C, Wu J, Shao X. Stacked pruning sparse denoising autoencoder basedintelligent fault diagnosis of rolling bearings. Appl Soft Comput. 2020;88:106060. doi:10.1016/j.asoc.2019.106060.
15. Ma M, Mao Z. Deep convolution-based LSTM network for remaining useful life prediction. IEEE Trans Ind Inform. 2021;17(3):1658–67. doi:10.1109/TII.9424.
16. Yang C, Liu J, Zhou K, Jiang X, Ge M, Liu Y. A node-level pathgraph-based bearing remaining useful life prediction method. IEEE Trans Instrum Meas. 2022;71:1–10.
17. Zhang D, Stewart E, Ye J, Entezami M, Roberts C. Roller bearing degradation assessment based on a deep MLP convolution neural network considering outlier regions. IEEE Trans Instrum Meas. 2020;69(6):2996–3004. doi:10.1109/TIM.19.
18. 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 Safe. 2023;230:108914. doi:10.1016/j.ress.2022.108914.
19. Lei Y, Han T, Wang B, Li N, Yao T, Yang J. XJTU-SY rolling element bearing accelerated life test datasets: a tutorial. J Mech Eng. 2019;55(16):1–6. doi:10.3901/JME.2019.16.001.
20. Cheng Y, Hu K, Wu J, Zhu H, Shao X. A convolutional neural network based degradation indicator construction and health prognosis using bidirectional long short-term memory network for rolling bearings. Adv Eng Inform. 2021;48:101247. doi:10.1016/j.aei.2021.101247.
21. Cao Y, Ding Y, Jia M, Tian R. A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings. Reliab Eng Syst Safe. 2021;215:107813. doi:10.1016/j.ress.2021.107813.
22. Hou M, Pi D, Li B. Similarity-based deep learning approach for remaining useful life prediction. Measurement. 2020;159:107788. doi:10.1016/j.measurement.2020.107788.
23. Yao D, Li B, Liu H, Yang J, Jia L. Remaining useful life prediction of roller bearings based on improved 1D-CNN and simple recurrent unit. Measurement. 2021;175:109166. doi:10.1016/j.measurement.2021.109166.
24. Gao T, Li Y, Huang X, Wang C. Data-driven method for predicting remaining useful life of bearing based on Bayesian theory. Sensors. 2021;21(1):182.