

Research Article

A Novel Multiscale Deep Health Indicator with Bidirectional LSTM Network for Bearing Performance Degradation Trend Prognosis

Han Wang ¹, Gang Tang ¹, Youguang Zhou,² and Yujing Huang¹

¹College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, Beijing 100029, China

²Shanghai Aircraft Design and Research Institute, Shanghai 201210, China

Correspondence should be addressed to Gang Tang; tanggang@mail.buct.edu.cn

Received 11 September 2020; Revised 4 October 2020; Accepted 20 October 2020; Published 3 November 2020

Academic Editor: Changqing Shen

Copyright © 2020 Han Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As rolling bearings are the key components in rotating machinery, bearing performance degradation directly affects machine running status. A tendency prognosis for bearing performance degradation is thus required to ensure the stability of operation. This paper proposes a novel strategy for bearing performance degradation trend prognosis, including health indicator construction techniques and a performance degradation trend prediction method. To more accurately represent the degradation trend, the multiscale deep bottleneck health indicator is proposed as a new synthesized health indicator to remove high-frequency detail signals from features, which can reduce possible fluctuations in conventional synthetic health indicators. A suitable method for selecting the statistical characteristics required for fusion is also presented to solve the problem of information redundancy that affects trend representation. In addition, a stacked autoencoder network is used for deep feature extraction of selected statistical features. A bidirectional long short-term memory network prediction model is also proposed for the prediction of degradation trend, which can make full use of historical and future information to improve prediction accuracy. Finally, experiments are carried out to verify the effectiveness of the proposed method.

1. Introduction

Rolling bearings are one of the most critical components in rotating machinery and their operating state will affect the overall performance of mechanical equipment [1–3]. Bearing performance degradation trend prognosis can ensure the stability of mechanical equipment, avoid catastrophic events, and extend the life cycle of the machinery [4–6].

The prediction of bearing performance degradation is based on operating rules derived from the analysis of historical monitoring data [7, 8]. In general, the prediction process can be divided into three stages: (1) data acquisition; (2) health indicator (HI) construction; and (3) performance degradation trend prognosis [4, 9]. Therefore, constructing HI from the monitoring data and using an effective prediction algorithm are both important factors affecting the final prediction result.

Features of the statistical data are most commonly used as HI to determine the true health of the bearing. The selection or construction of a suitable HI is highly important to correctly determine the occurrence and development of bearing degradation. Conventional HI constructions are mainly based on statistical characteristics. Specifically, the traditional method often uses the statistical features extracted from the time domain [10], frequency domain [11], or time-frequency domain [12] of the vibration signal to represent the degradation trend. For example, the root mean square (RMS) value of the time-domain vibration signal is a commonly used HI [7, 13]. However, a simple statistical feature indicator is neither sensitive enough to early bearing defects nor stable enough in later stages. Therefore, a multidimensional HI must be established to fully represent the trend of degradation.

In recent years, synthesized HI has attracted much interest for its ability to more accurately represent the

tendency of bearing degradation and is usually constructed using a data fusion technology. Liao [14] used a genetic algorithm to perform deep feature fusion and extraction on different domain features to construct a new HI. Li et al. [15] employed cointegration theory to fuse the two degradation characteristics of RMS and SampEn and formed a new HI to improve the ability of fault prediction. Wang et al. [16] fused multiple statistical features through support vector data description and constructed an effective fusion HI. However, these articles rarely select the statistical characteristics of fusion, which may lead to redundant information and affect the description of the degradation process. In addition, the synthesized health indicators often contain a lot of detailed fluctuations, which is not conducive to the description of the degradation trend. Therefore, the construction of a non-linear multifeature model that can adapt to bearing trend prediction is urgently required.

After extracting appropriate feature parameters, a suitable prediction model is needed to achieve an accurate prediction of bearing degradation trend. Machine learning models have been widely used in machinery performance assessment and remaining life prediction. For instance, Gebraeel et al. [17] used neural networks for bearing degradation states and life predictions. Tang et al. [8] extracted a variety of characteristic parameters from the bearing vibration data and used the least-squares support-vector machine (LS-SVM) to predict the degradation trend of the bearing. Aye and Heyns [18] proposed an improved Gaussian process regression (GPR) for the prediction of remaining useful life (RUL). However, these traditional machine learning techniques cannot automatically learn multiple representations from the original input data and require hand-coded rules or domain knowledge.

Deep learning has been introduced and widely studied in recent years to solve such issues. Belonging to the field of artificial intelligence, deep learning can automatically and accurately learn deeper features from low-level feature composition [19–21]. Malhi et al. [22] used a continuous wavelet transform to preprocess the bearing vibration signals and then employed recurrent neural networks (RNNs) to further predict the trend of degradation. Deutsch and He [23] proposed a deep belief network-feedforward neural network (DBN-FNN) algorithm for RUL prediction, while Zhu et al. [24] extracted time-frequency features by wavelet transform and then used convolutional neural networks (CNNs) to estimate RUL. These deep learning models can achieve certain effects for the fault diagnosis and prediction of mechanical equipment but do not fully consider the correlation between monitoring data or the correlation between monitoring data and time information. This characteristic means that such models can easily experience overfitting problems, which limits their application in mechanical equipment fault diagnosis and prognosis.

Long short-term memory network (LSTM) is a type of deep learning model that is suitable for processing and predicting important events with relatively long intervals and delays in time series. A memory controller is included in the LSTM structure to determine when to forget, remember, and output data. For instance, Zhang et al. [25] proposed a

method for evaluating bearing performance degradation using LSTM recursive network. Wang et al. [26] solved the gradient explosion problem by using LSTM and realized the time series prediction of rolling bearing vibration signals. Hinch and Tkouat [27] used LSTM to capture the degradation process and predict the time value to estimate the remaining useful life. However, LSTM only utilizes historical information. To improve the prediction ability of the model in this work, a bidirectional LSTM network (BiLSTM) is employed to facilitate the use of both historical information and future information in the prediction of degradation trends.

In this paper, a novel multiscale deep bottleneck health indicator (MDBHI) with BiLSTM for performance degradation trend prognosis of rolling bearing is proposed. A more reasonable synthesized HI is developed to represent and predict the bearing degradation trend. The major contributions of this work are summarized as follows:

- (i) A method for multiscale decomposition of the selected statistical features is proposed. Multiscale decomposition is performed on the selected statistical features to be fused and high-frequency noise components are removed to solve the problem of large feature fluctuations in bearing performance prediction.
- (ii) A multiscale deep health indicator (MDBHI) is proposed. A stacked autoencoder (SAE) network is used for deeper feature extraction of multiple statistical features after multiscale decomposition in order to obtain an HI that better reflects the tendency of bearing degradation.
- (iii) A prediction model of bearing degradation trend based on BiLSTM is established which can make full use of both historical and future information to obtain more accurate forecasting trends.

The remainder of the paper is organized as follows: in Section 2, the BiLSTM network, SAE, and the principle of multiscale decomposition are briefly introduced. In Section 3, an introduction of MDBHI extraction and the framework of the proposed method is provided. Experiments are presented to verify the effectiveness of the proposed method in Section 4, and, finally, conclusions are drawn in Section 5.

2. Related Basic Theories

2.1. Overview of Bidirectional LSTM Networks. The concept of bidirectional circulatory neural network was first put forward by Mike Schuster [28]. In the network, a backward hidden layer is added to the traditional RNN network structure to enable training of input information within a specific time range in the past and future. The BiLSTM network is an improved network structure based on bidirectional RNN. Hidden layers in RNN are transformed into memory modules in the LSTM network and such structures can make full use of temporal sequence information for training. A typical BiLSTM structure is shown in Figure 1.

The LSTM memory module for each time step is marked by green or orange circles in the figure. Compared with the

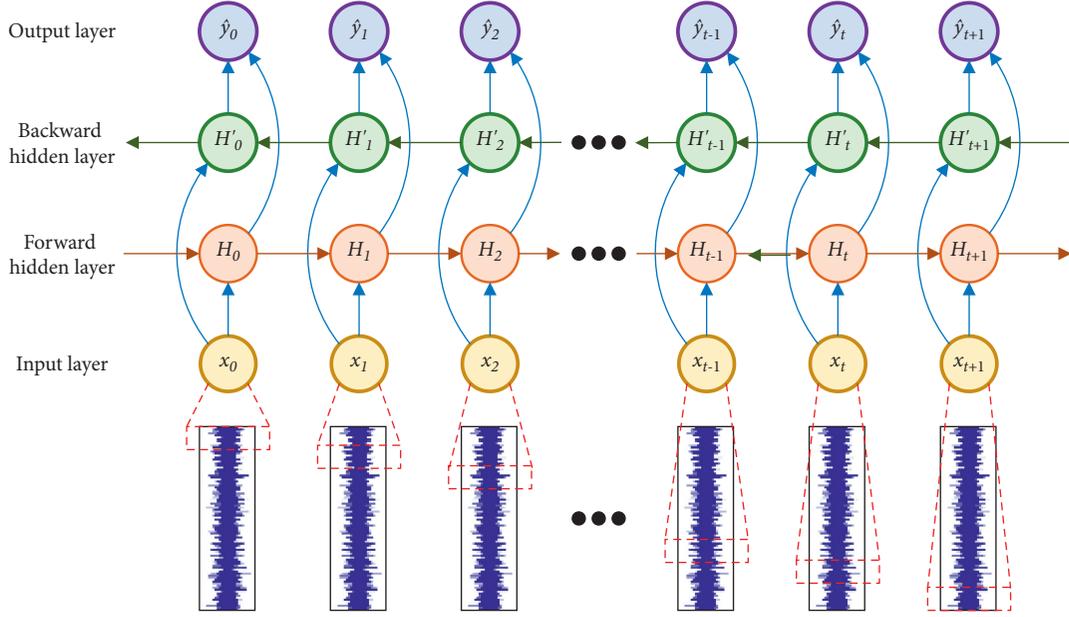


FIGURE 1: Basic structure of a BiLSTM.

conventional LSTM network, BiLSTM contains hidden layers that pass the input information from back to front in each hidden layer node until finally the information is gathered in the output layer. There is no connection between the two hidden layers, so it can be considered as two LSTM networks stacked to work together on one output.

In BiLSTM training, forward propagation is basically consistent with conventional LSTM. Taking the output as an example, the basic principle is as follows:

$$\begin{aligned} \vec{h}_t &= f_2(x_t \vec{U} + \vec{h}_{t-1} \vec{W}), \\ \overleftarrow{h}_t &= f_2(x_t \overleftarrow{U} + \overleftarrow{h}_{t+1} \overleftarrow{W}), \\ y_t &= f_3(\vec{V} \vec{h}_t + \overleftarrow{V} \overleftarrow{h}_t). \end{aligned} \quad (1)$$

where \vec{h}_t , \vec{U} , \vec{W} , \vec{V} , and f_1 , respectively, represent the state of the forward hidden layer, the weight of the input, the weight of the hidden layer at time $t-1$, the weight of the output, and the activation function; in the same way, \overleftarrow{h}_t , \overleftarrow{U} , \overleftarrow{W} , \overleftarrow{V} , and f_2 , respectively, represent the state of the backward hidden layer, the weight of the input, the weight of the hidden layer at time $t+1$, the weight of the output, and the activation function; and y_t and f_3 , respectively, represent the output value and the activation function of the output layer.

Backpropagation through time (BPTT) algorithm is still used to update weight parameters in BiLSTM backpropagation. By calculating the derivative of the loss function to the weight of each layer, using the chain rule and gradient descent, the network parameters are constantly updated to carry out network converge. The only difference is that the backward propagation updates the forward hidden layer from $t=T$ to $t=1$ and the backward hidden layer from $t=1$ to $t=T$. By adding a time delay, BiLSTM

ensures that the current output depends on future information in addition to the historical information, which can improve the prediction ability of the model.

2.2. Stacked Autoencoder Networks. Autoencoder network (AE) is a type of unsupervised neural network which employs an encoding and decoding process. The first part is the encoding process, which transforms the input into other dimensions. The second part is the decoding process, which restores the encoded information to input. For AE networks, the final output of the network is usually not included, but encoded low-dimensional features are extracted to achieve the effect of feature reduction.

The forward information dissemination process of the AE network can be expressed as follows: input signal $X = \{x_1, x_2, \dots, x_n\}$ is mapped to a new dimensional space by an encoder and a hidden layer is obtained. The hidden layer is then transformed into an output $\hat{X} = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n\}$ using a decoder, and the process can represent

$$\begin{aligned} H_t &= f_1(X), \\ \hat{X} &= g_1(H_t), \end{aligned} \quad (2)$$

where f_1 is the encoder and g_1 represents the decoder. Encoders and decoders typically perform nonlinear mapping of their inputs. The basic principles are as follows:

$$\begin{aligned} H_1(x) &= s_{f_1}(W_1 X + b_1), \\ \hat{X}(x) &= s_{g_1}(W'_1 H_1 + b'_1), \end{aligned} \quad (3)$$

where s_{f_1} and s_{g_1} represent the nonlinear activation functions of the hidden layer and the output layer, respectively; W_1 and W'_1 represent the weight parameters of the input and hidden layer; and b_1 and b'_1 represent the bias parameters of the hidden layer and the output layer, respectively.

The backward information propagation process of the AE network can be expressed as follows: the error between the input value and the model output value can be back-propagated through the preset loss function, so that the

network weight parameter is continuously updated until the network model loss function reaches the expected threshold. The specific mathematical process is to find the network parameters that minimize the loss function:

$$\begin{aligned}
 L(\theta^*, \theta'^*) &= \arg \min_{\theta, \hat{\theta}} \frac{1}{n} \sum_{i=1}^n L(x_i, \hat{x}_i) \\
 &= \arg \min_{\theta, \hat{\theta}} \frac{1}{n} \sum_{i=1}^n L(X, s_{gi}(W'_i H_i + b'_i)) \\
 &= \arg \min_{\theta, \hat{\theta}} \frac{1}{n} \sum_{i=1}^n L(X, s_{gi}(W'_i s_{fi}(W_i X + b_i) + b'_i)),
 \end{aligned} \tag{4}$$

where $\theta = \{W_1, b_1\}$, $\hat{\theta} = \{W'_1, b'_1\}$, and θ^* and θ'^* indicate the updated weight parameters. The schematic diagram of the three-layer AE network is shown in Figure 2.

On the basis of AE, SAE takes the output of the hidden layer of the AE as the input of the next AE to continue on dimensional reduction learning of features. The SAE is able to learn a variety of expressions of raw data layer by layer, and each layer is based on the expression of the bottom layer, so it is more suitable for tasks such as complex classification. The specific process is shown in Figure 3.

2.3. The Multiscale Decomposition. As a representative multiscale decomposition method, wavelet transform can obtain the time-frequency components of nonstationary signals. The basic idea is to approximate a target signal by using a cluster of wavelet function systems, which are obtained by different scaling and shifting transformations of a basic wavelet function.

For the continuous wavelet transform, the scale factor \mathbf{a} , translation factor \mathbf{b} , and time \mathbf{t} are continuous. It is inconvenient to process the parameters by computer, which will also increase the redundancy of wavelet coefficients. Therefore, it is necessary to discretize the continuous wavelet. The discretization of the scale factor and the translation factor is denoted as follows:

$$\begin{aligned}
 a &= a_0^j, \\
 b &= kb_0 a_0^j,
 \end{aligned} \tag{5}$$

where a_0^j and b_0 are discrete initial values of scale factors a and b , and $j = 0, \pm 1, \pm 2, \dots$, $k = 0, \pm 1, \pm 2, \dots$. The discrete wavelet function $\psi_{j,k}(k)$ and discrete wavelet transform DWT are defined as follows:

$$\begin{aligned}
 \psi_{j,k}(t) &= a_0^{-(j/2)} \psi(a_0^{-j} t - b_0 k), \\
 \text{DWT}_f(j, k) &= \int_{-\infty}^{+\infty} a_0^{-(j/2)} f(t) \bar{\psi}(a_0^{-j} t - b_0 k) dt.
 \end{aligned} \tag{6}$$

Multiscale wavelet analysis is essentially a process of continuous approximation. By approximating the function

$f(t)$ with different scales for function $f(t)$ in any one-dimensional space, the approximation function of the function $f(t)$ will be obtained in the process of approximation. The function is an imitation of the function $f(t)$. As the approximation process continues to increase, this function gradually becomes more precise.

Multiscale decomposition can obtain some low-frequency approximation signals and high-frequency detail signals by approximating nonstationary signals at different time resolutions. The low-frequency approximation signal characterizes the approximate contour and trend of the original signal, while the high-frequency detail signal contains the detailed information in the signal, which can be expressed as

$$\begin{aligned}
 f(x) &= A_1^{d_j} f = A_2^{d_{j+1}} f + D_2^{d_{j+1}}, \\
 f &= A_3^{d_{j+1}} f + D_3^{d_{j+2}} f + D_3^{d_{j+1}}, \\
 f &= \dots,
 \end{aligned} \tag{7}$$

where A is the low-frequency approximation signal and D is the high-frequency detail signal.

3. The Proposed Bearing Trend Prediction Method

3.1. Extracting Deep Bottleneck Health Indicator. The feature extraction of vibration signals is a necessary premise for the degradation trend prognosis of bearing performance. Extracting rich feature samples by monitoring vibration signals can improve the accuracy of trend prediction. Most commonly used statistical feature extraction methods work to explore feature parameters in the time domain, frequency domain, or time-frequency domain, respectively. However, this kind of simple statistical feature parameter cannot fully reflect the running status of bearings.

To solve this problem, it is necessary to select various sensitive characteristic parameters to measure the health of the bearing from various aspects. However, if the features to be fused are not properly selected, information redundancy may diminish the extraction effect. Therefore, depth features must be extracted according to the various sensitive feature

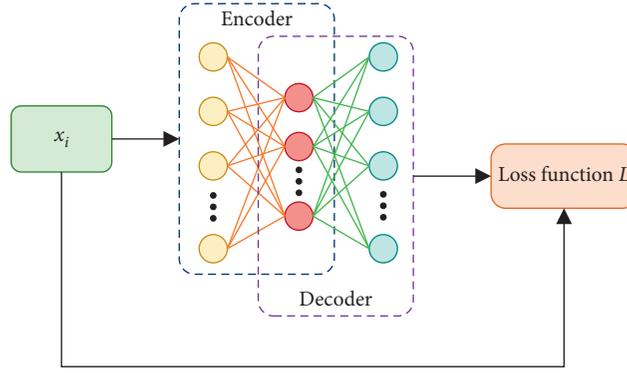


FIGURE 2: Schematic diagram of the AE principle.

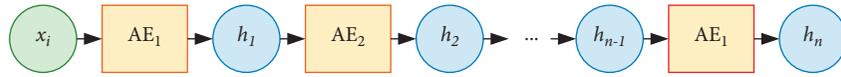


FIGURE 3: The basic process of SAE.

parameters selected. This process ensures that the features are representative during bearing degradation.

Based on the monotonous feature evaluation indicator, this paper utilizes the correlation between feature parameters to further select sensitive feature parameters. This correlation indicator based on feature parameters is referred to as a consistent feature evaluation indicator in literature [29]. The consistency evaluation indicator Con is defined as follows:

$$\text{Con}(F_1, F_2) = \frac{\sum_{i=1}^k (f_{1,k} - \bar{F}_1)(f_{2,k} - \bar{F}_2)}{\sqrt{\sum_{i=1}^k (f_{1,k} - \bar{F}_1)^2} \sqrt{\sum_{i=1}^k (f_{2,k} - \bar{F}_2)^2}}, \quad (8)$$

where F_1 and F_2 are two characteristic parameters, $f_{1,k}$ and $f_{2,k}$ represent the values of F_1 and F_2 at time t_k , respectively, and \bar{F}_1 and \bar{F}_2 represent the mean values of F_1 and F_2 , respectively.

Six characteristic parameters can be obtained through monotonic evaluation indicators: absolute mean value (AMV), variance, RMS, peak value, energy spectrum (ES), and wavelet decomposition of second layer energy (WDSLE). The consistency evaluation indicator between each characteristic parameter is then calculated and several characteristic parameters with high consistency of degradation laws are selected.

The steps for extracting deep bottleneck health indicators are as follows: first, vibration signals measured by sensors are collected; different signal characteristics are then classified and processed; and, finally, deep learning algorithms are used for deep feature extraction of various features. Traditional multifeature extraction algorithms such as principal component analysis (PCA) can reduce the feature dimension and accelerate operation. However, the transmission path of a bearing vibration signal is very complicated, and the signal obtained by sensors has strong nonlinear characteristics. Therefore, although PCA achieves dimensionality reduction by studying a linear transformation of the data, it cannot learn deeper nonlinear features. The SAE can obtain nonlinear

characteristics of a bearing vibration signal by setting a nonlinear activation function. Therefore, in this paper, deep features are obtained from multiple features through SAE; then, the output of the hidden layer provides the deep feature parameters required. The extraction process of a deep bottleneck health indicator (DBHI) is shown in Figure 4.

3.2. Multiscale Decomposition Parameter Selection. The DBHI obtained using the process above often contains many fluctuating details. This has an impact on bearing performance predictions, especially in the later stages of bearing degradation. Therefore, the extraction of multifeature sensitive parameters is attempted in this work using a multiscale decomposition. Through multiscale wavelet decomposition of the multifeature sensitive set, the low-frequency approximate component of the feature is extracted and high-frequency detailed components with large fluctuations are eliminated. Multiscale decomposition can be used to optimize the sensitive feature sample set, which is more conducive to trend prediction in the later stage of bearing degradation.

The multiscale decomposition method contains two parameters. The first one is the wavelet function and the second is the decomposition scale, also called the decomposition layer number. The difference between these two parameters will affect the final extracted trend term, so it is necessary to select an appropriate wavelet function and decomposition scale.

3.2.1. Selection of Optimal Decomposition Function. Wavelet transform is proposed to improve Fourier transform in this work. Compared with the trigonometric functions used in the Fourier transform, the wavelet function has a characteristic of diversity. Moreover, different wavelet functions will produce different effects on the analysis of the same signal, so it is necessary to select an appropriate wavelet function to analyze the corresponding signal.

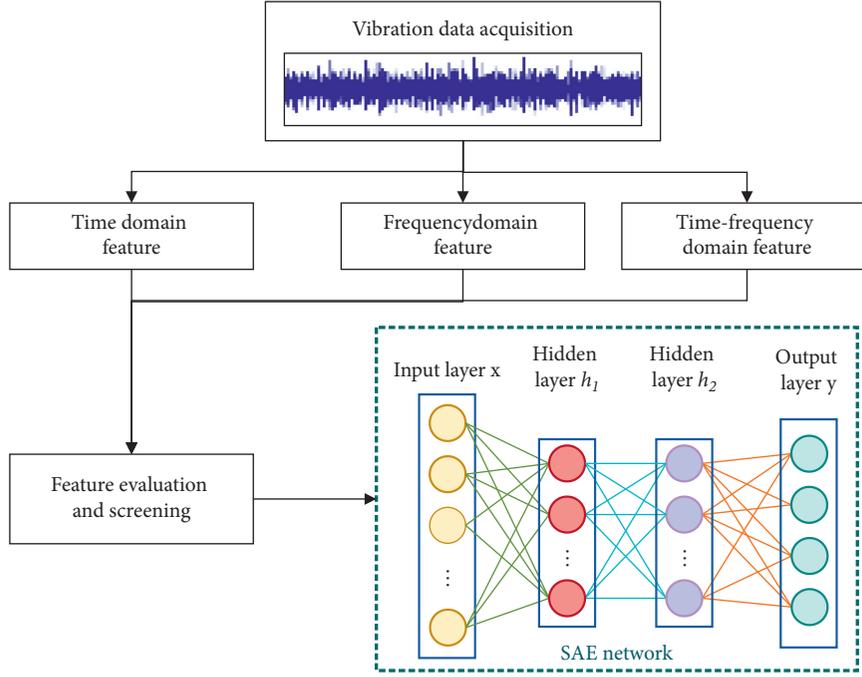


FIGURE 4: Extraction process of DBHI based on SAE.

Generally, there are five commonly used wavelet functions: Haar wavelet, Coiflet (coifN) wavelet, Dmeyer (dmey) wavelet, Daubechies (dbN) wavelet, and Symlets (symN) wavelet. The selection criteria for the generally accepted optimal wavelet function can be roughly divided into orthogonality, tight support, regularity, symmetry, and vanishing moment. However, the five common wavelet functions mentioned above contain individual advantages and disadvantages and should be selected according to the actual needs. In this section, multiscale wavelet decomposition is used to optimize DBHI of bearing degradation trend and the trend term (low-frequency approximation component) of sensitive feature parameters is extracted. Therefore, the selected wavelet basis function should be able to separate the low-frequency approximation component from the high-frequency detail components in the sensitive feature parameters. For this purpose, a trending optimal wavelet function selection criterion is defined [9], and the principle is as follows:

$$Tre = \frac{k(\sum_{k=1}^k f_k t_k) - (\sum_{k=1}^k f_k)(\sum_{k=1}^k t_k)}{\sqrt{k \sum_{k=1}^k f_k^2 - (\sum_{k=1}^k f_k)^2} \sqrt{k \sum_{k=1}^k t_k^2 - (\sum_{k=1}^k t_k)^2}}, \quad (9)$$

where t_k represents t_k moment and f_k is the value of the characteristic parameter at time t_k . The closer the trending criterion Tre is to 1, the better the trend of the characteristic parameters is.

For the selection of an optimal wavelet function in MDBHI extraction proposed in this paper, the five most common wavelet functions described above are selected as alternative wavelet functions. Considering that some wavelet functions consist of a series of functions, such as dbN and coifN, the vanishing moment N should be determined. The

larger the vanishing moment, the higher the smoothness of the wavelet, and the better the wavelet decomposition band division effect. However, an excessive vanishing moment will cause the tight support in the time domain to weaken, increase the calculation requirements, and provide poor real-time performance. Two-layer wavelet decomposition is then performed on the multifeature sensitive parameters selected in the upper part and the trend Tre value extracted by each wavelet function is calculated. Finally, the optimal wavelet function can be selected.

3.2.2. Selection of the Optimal Number of Decomposition Layers. Multiscale wavelet decomposition is a constant approximation of the target signal in which some high-frequency components will be discarded. As the number of decomposition layers increases, the high-frequency details of the signal will decline progressively, and the low-frequency trend information will become increasingly detailed. This is conducive to the prediction of bearing degradation trend, but too many decomposition layers will lead to a deformation of the trend term obtained by the final decomposition. Therefore, it is important to select an appropriate number of wavelet decomposition layers.

A fusion indicator is proposed in this paper for the selection criteria of wavelet decomposition layers. Through a weighted summation of consistency indicator and trend indicator, a fusion indicator Fus is obtained, which is defined as follows:

$$Fus = w_1 Con + w_2 Tre, \quad (10)$$

where w_1 and w_2 , respectively, represent the weight of the consistency indicator Con and the trend indicator Tre .

Similar to the above wavelet function selection, the trend indicator represents the trend of sensitive feature parameters after multiscale decomposition. In theory, the higher the number of decomposition layers, the smoother the trend curve. However, too many decomposition layers will cause the feature to be deformed, thus increasing the work of consistency indicator to assist in the adjustment. Consistency indicator calculates the correlation between sensitive feature parameters and original feature parameters after multiscale decomposition, and, in theory, the more the decomposition layers, the smaller the consistency indicator. By combining these two indicators, the optimal wavelet decomposition layer can be appropriately selected.

3.3. Summary of the Proposed Method. A trend prediction method based on MDBIH and BiLSTM is proposed in this paper to solve the problem of value fluctuations in bearing degradation feature parameters. First, the statistical features from the time domain, frequency domain, and time domain are, respectively, extracted. The appropriate statistical characteristics are selected based on the consistency evaluation indicator, and a multifeature sensitive parameter set is obtained. To extract the low-frequency approximate component of the feature, the high-frequency detailed components with large feature value fluctuations are removed. Multiscale decomposition is then carried out for the multifeature sensitive parameter set and the decomposed multifeature set is input into the SAE to extract MDBIH. Finally, the BiLSTM model is used to predict the trend of bearing degradation. By adding time delay information, the current output is dependent on not only the historical information but also the future information, so as to improve the prediction ability of the model. The whole framework of the proposed method is shown in Figure 5.

4. Experiment Verification

Rolling bearing degradation trend prediction requires a full-life test data of the bearing. Usually, the actual bearing normal degradation process takes too much time; thus, a bearing acceleration degradation test is often used. The experimental data in this section uses data from the degraded database of four roller bearings provided by the University of Cincinnati and shared on the website of the Prognostic Data Repository of NASA [30]. The test rig for bearing acceleration degradation is shown in Figure 6.

The test rig was comprised of an alternating current dynamo and a shaft with four bearings mounted on the shaft. The dynamo transmitted power through the belt to make the shaft rotate at a speed of 2000 r/min. The bearing on the test rig was subjected to a constant load of 6000 lbs along the shaft diameter direction.

The experiment employed a total of three sets of data. Each set contained complete degradation process data from the start of bearing testing to failure, and the second data set was used. The sampling frequency was set to 20 kHz and the sampling time was 1 s. A total of 20,480 acceleration data points were collected during each sampling time and the

sampling interval duration was 10 min. The bearing accelerated degradation experiment was run for eight days and a total of 984 sets of data were collected until a serious outer ring failure occurred in bearing 1. The full-life cycle vibration time domain of a rolling bearing is shown in Figure 7.

In general, the life cycle of bearing degradation can be divided into five stages, which are running-in stage, steady state stage, defect initiation, crack initiation to crack opening state, and defect propagation [31]. This is because the bearing wear evolution process will produce some surface topological changes which have a significant influence on bearing state monitoring. The test data used in this work incorporates this type of bearing degradation growth process; thus, the objective is to obtain a more accurate prognosis of the degradation trend, especially near the changing points of different stages.

4.1. Establishment of Feature Fusion Health Indicators. The six statistical characteristics mentioned in the third section are AMV, variance, RMS, peak, ES, and WDSLE, and their characteristic parameter diagrams are shown in Figure 8. As shown in the figure, the degradation trend represented by different features, the starting point of the fault, and the degree of degradation are different. In other words, a simple feature cannot fully reflect the bearing running status. Intuitively, the extraction of multiple characteristic parameters is an alternative choice to solve this problem. In this work, the consistency of different characteristics is calculated according to the consistency evaluation indicator. The purpose is to determine the most suitable features for the establishment of subsequent health indicators. The results are provided in Table 1 and Figure 9.

It can be clearly observed from Table 1 that, among the six characteristic parameters selected by monotonicity evaluation indicators, the two-pair consistency evaluation indicators of the five characteristic parameters, that is, absolute mean value, variance, effective value, peak value, and energy spectrum, are very large, indicating that the degradation rules contained in the five characteristic parameters are basically consistent. However, the consistency evaluation indicator between the second layer energy of wavelet decomposition and the other five characteristic parameters is relatively small, indicating that the second layer energy of wavelet decomposition cannot accurately represent the bearing degradation information. Therefore, in this paper, the five characteristic parameters of absolute mean value, variance, effective value, peak value, and energy spectrum are selected as the original characteristic sample set of DBHI.

4.2. Extracting Deep Bottleneck Health Indicator. In order to select the best wavelet function required in multiscale decomposition, the five common wavelet functions mentioned in the third section are selected as candidate wavelet functions and the vanishing moment M is set as 4 to make the contrast more complete. Two-layer wavelet decomposition is then performed on the multifeature sensitive parameters selected above. Finally, the trending Tre corresponding to the trend term extracted by each wavelet

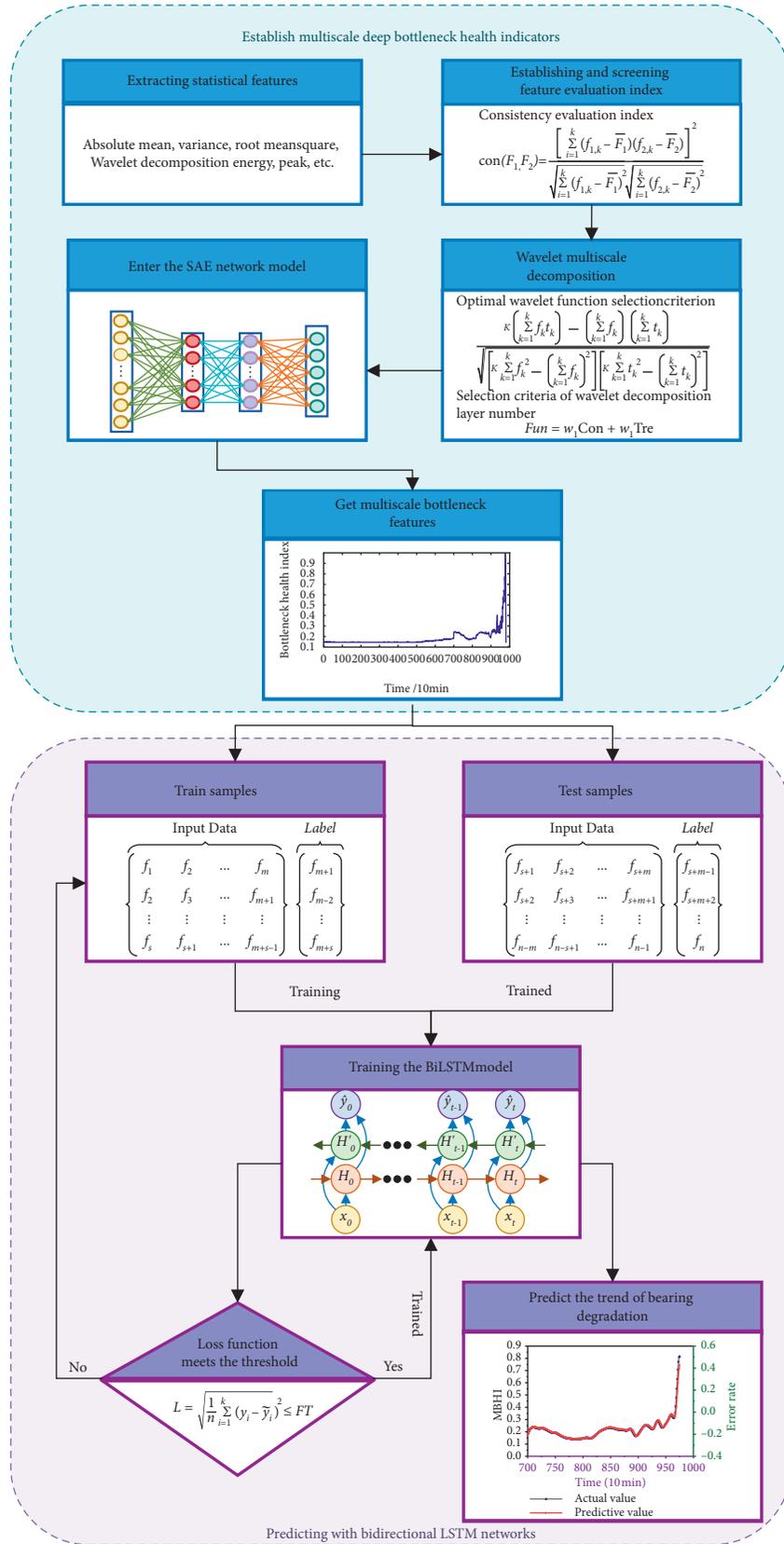


FIGURE 5: The framework of the proposed method.

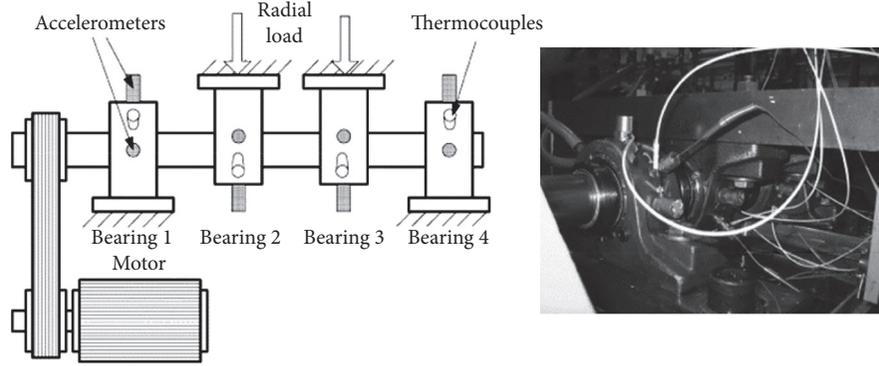


FIGURE 6: Test rig for accelerated degradation of rolling bearing.

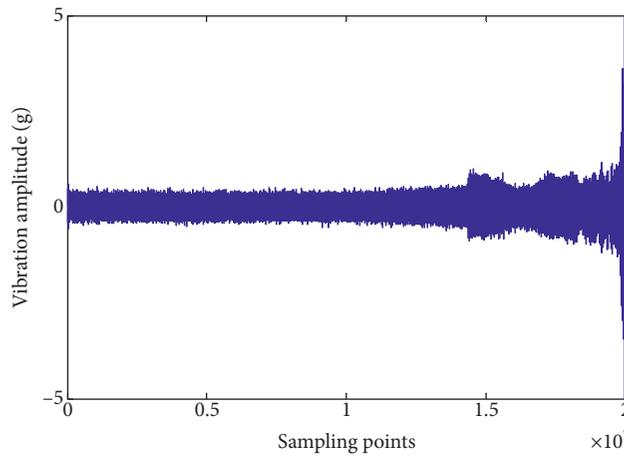


FIGURE 7: Time-domain diagram of rolling bearing vibration during its whole life cycle.

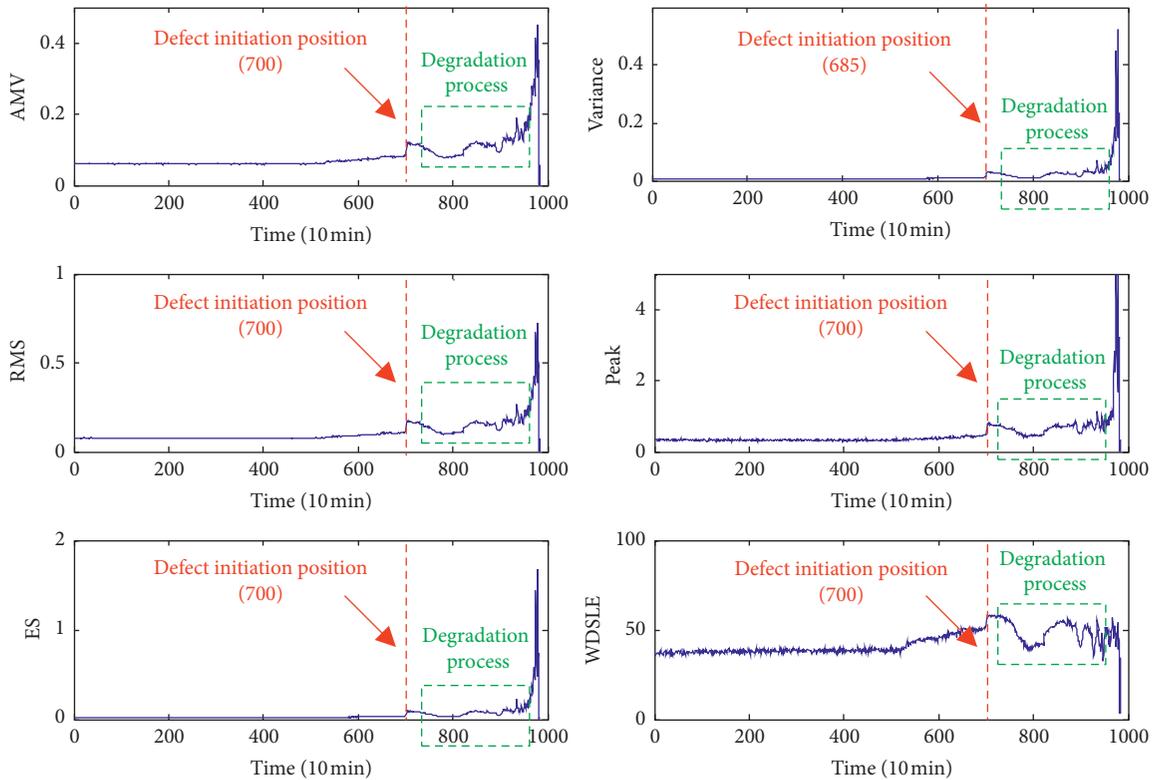


FIGURE 8: Six characteristic parameters selected by monotonic evaluation indicators.

TABLE 1: Table of consistency indicator value for six characteristic parameters.

	AMV	Variance	RMS	Peak value	ES	WDSLE
AMV	1.0000	0.8915	0.9945	0.9360	0.8915	0.5210
Variance	0.8915	1.0000	0.9158	0.9645	1.0000	0.2591
RMS	0.9945	0.9158	1.0000	0.9599	0.9158	0.5292
Peak value	0.9360	0.9645	0.9599	1.0000	0.9645	0.4056
ES	0.8915	1.0000	0.9158	0.9645	1.0000	0.2591
WDSLE	0.5210	0.2591	0.5292	0.4056	0.2591	1.0000
Average	0.8724	0.8385	0.8859	0.8717	0.8385	0.4956

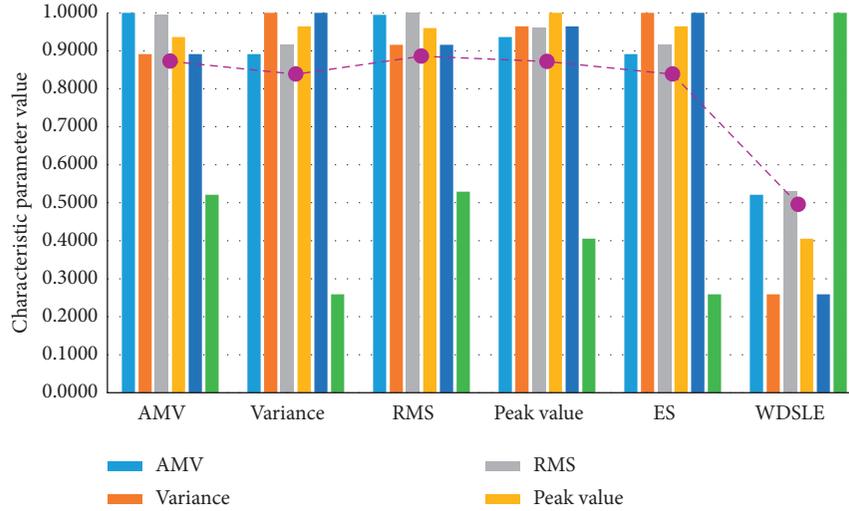


FIGURE 9: Diagram of consistency indicator value for six characteristic parameters.

TABLE 2: Table of trend values for different characteristic parameters decomposed by different wavelet functions.

	Haar	Coif4	Dmey	dbN	symN
AMV	0.6486	0.6916	0.6438	0.6538	0.6400
Variance	0.4425	0.4781	0.4301	0.4276	0.4287
RMS	0.6464	0.6903	0.6412	0.6504	0.6366
Peak value	0.5451	0.5844	0.5405	0.5434	0.5380
ES	0.4425	0.4781	0.4301	0.4276	0.4287
Average	0.5450	0.5845	0.5371	0.5406	0.5344

function after processing each feature parameter is calculated, and then the optimal wavelet function is selected. The results are shown in Table 2 and Figure 10.

According to Table 2, for these five wavelet functions, the *coif4* wavelet function has the largest mean value among the samples of sensitive characteristic parameters after two-layer wavelet decomposition. Moreover, the independent trend of each sensitive feature obtained by *coif4* wavelet decomposition is higher than that of other wavelet decomposition. Therefore, *coif4* wavelet is selected as the optimal wavelet function to extract MDBHI features of bearings.

Next, the number of wavelet decomposition layers must be selected according to the fusion indicator mentioned above. In this paper, the two weight values in the fusion indicator are set to be 0.5.

The next step is to determine an appropriate number of decomposition levels. The fusion indicators proposed in this article can help in this selection. Table 3 and Figure 11 show the calculation results of the fusion indicator of the one to six

layers of the wavelet decomposition. It can be clearly seen from the histogram that when the number of decomposition layers $N=3$, the fusion indicator value of each sensitive feature is larger than that of the other decomposition layer numbers. The average value of the fusion indicator is also the largest when $N=3$, so the optimal number of decomposition layers $N=3$ is selected. The optimal wavelet function is determined as *coif4* wavelet based on trend evaluation indicators, the optimal wavelet decomposition layer is selected as three layers based on fusion indicators, and then the sensitive feature samples are extracted, as shown in Figure 12. Comparing Figures 12 and 8, it can be observed that the trend of feature parameters after multiscale decomposition is more clear. This is helpful for the extraction of bottleneck health indicators in the next step.

Finally, through the screening of trend indicator and fusion indicator, the optimal wavelet function is determined as *coif4*, the optimal wavelet decomposition layer number is three, and the sensitive feature samples after multiscale

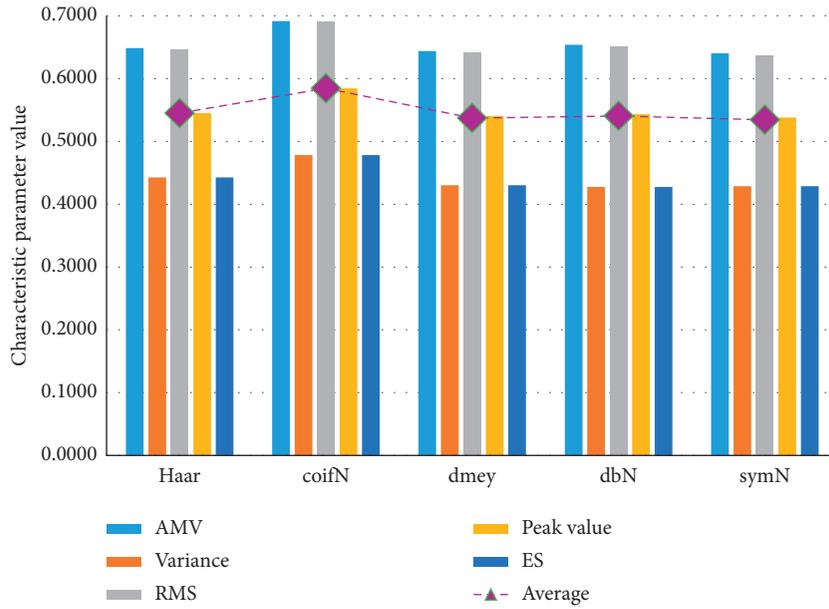


FIGURE 10: Diagram of trend values for different characteristic parameters decomposed by different wavelet functions.

TABLE 3: Table of fusion indicator values for decomposing characteristic parameters of different decomposition layers.

	$N=1$	$N=2$	$N=3$	$N=4$	$N=5$	$N=6$
AMV	0.7643	0.7607	0.7951	0.7752	0.7208	0.6141
Variance	0.5500	0.5399	0.6747	0.6401	0.5657	0.4242
RMS	0.7476	0.7418	0.7839	0.7472	0.6913	0.5842
Peak value	0.6474	0.6028	0.7055	0.6776	0.6312	0.5087
ES	0.5500	0.5399	0.6747	0.6401	0.5657	0.4242
Average	0.6519	0.6370	0.7268	0.6960	0.6350	0.5111

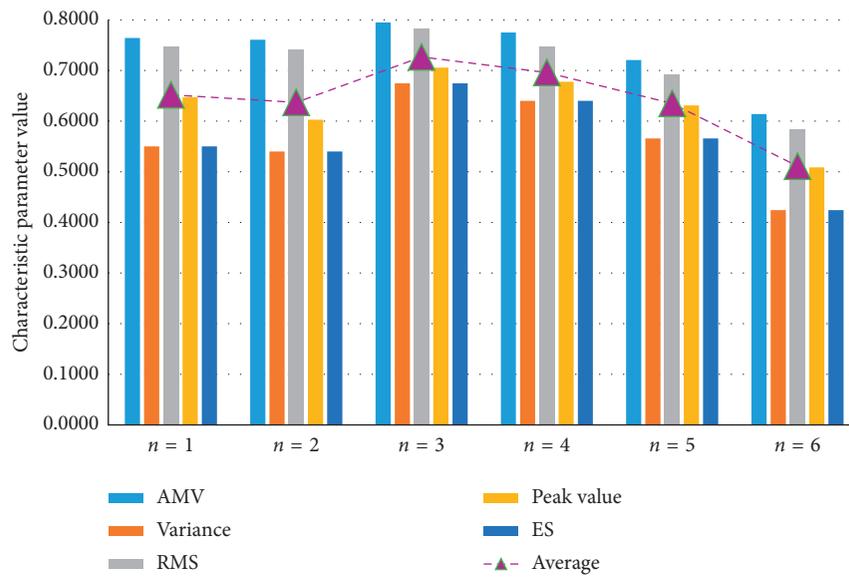


FIGURE 11: Diagram of fusion indicator values for decomposing characteristic parameters of different decomposition layers.

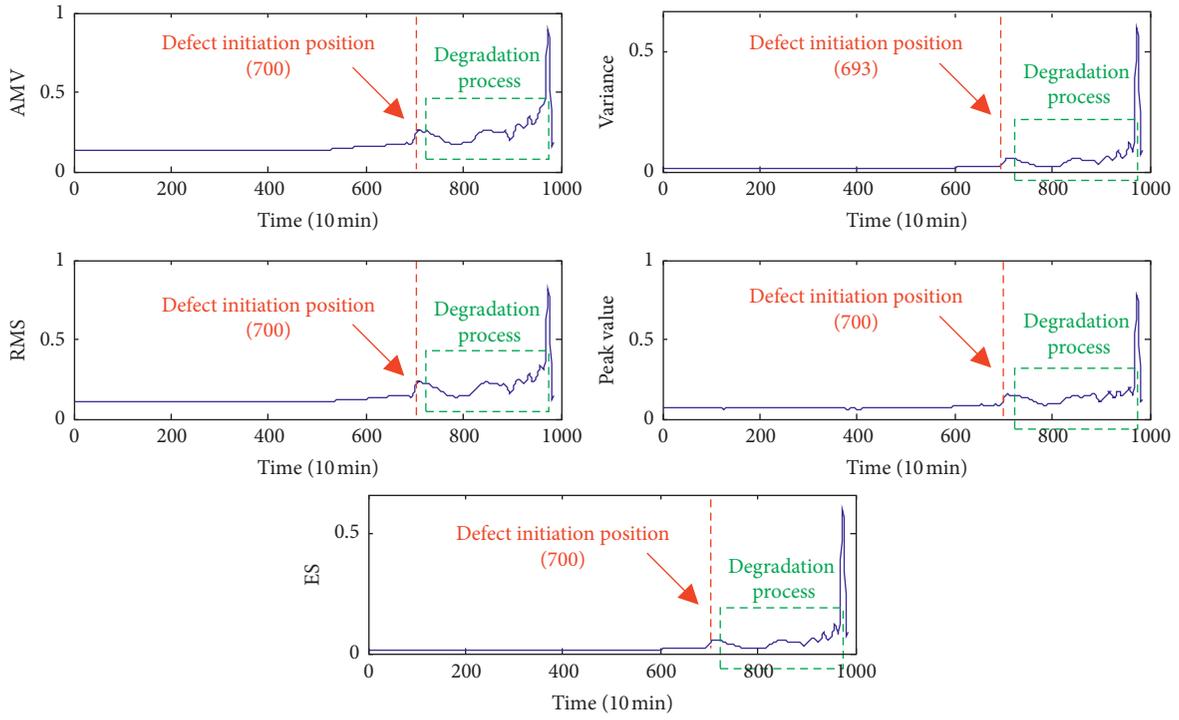
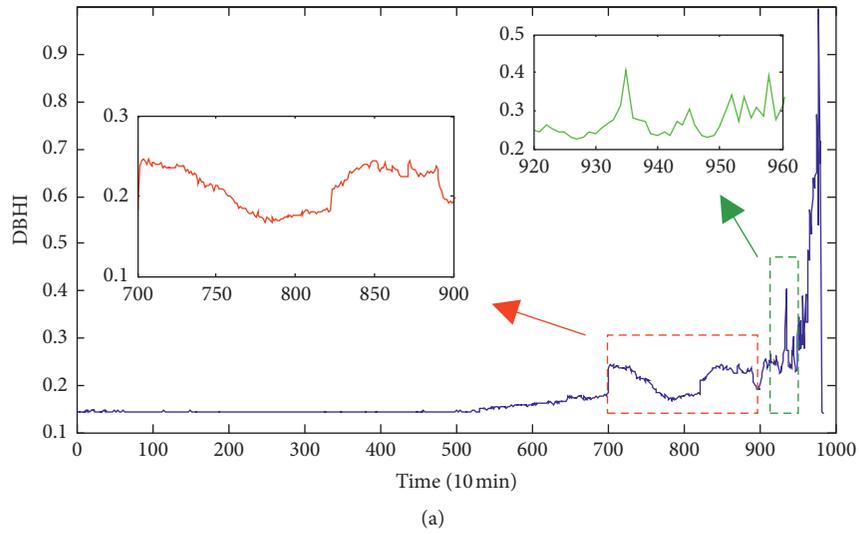


FIGURE 12: Characteristic parameters after multiscale decomposition.



(a)
FIGURE 13: Continued.

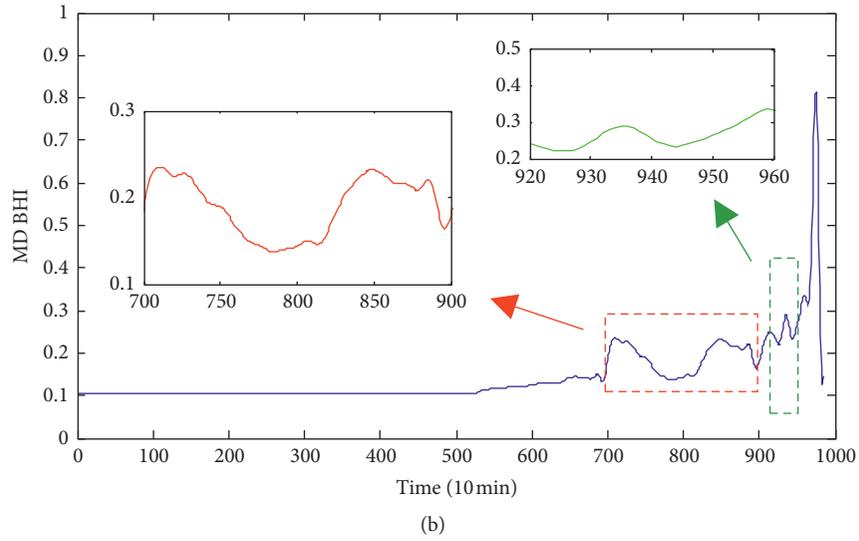


FIGURE 13: Description of bearing degradation trends: (a) DBHI and (b) MDBHI.

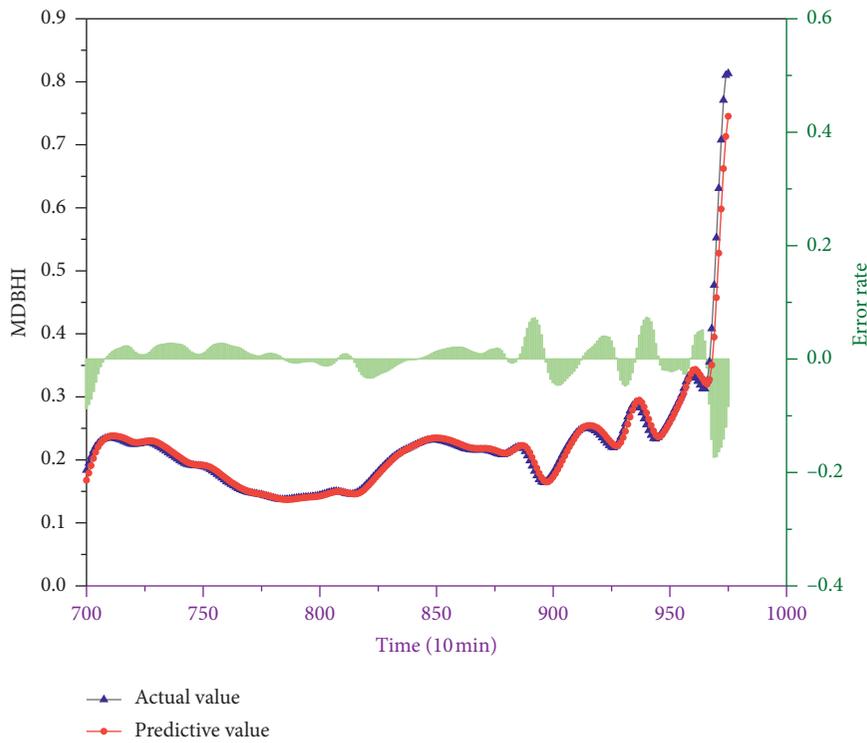


FIGURE 14: The prediction results and prediction error rate of the proposed method.

decomposition are calculated. Then, the five selected features after multiscale decomposition are input to SAE for deep feature extraction. Finally, MDBHI proposed in this paper can be obtained, and the results are shown in Figure 13. A comparison between MDBHI and DBHI without multiscale decomposition is also provided in the same figure.

Compared with the DBHI parameters without multiscale decomposition, MDBHI obviously removes some detail signals with large fluctuations. Removing these details can improve the accuracy of the prediction of bearing degradation trends. In the late stage of degradation, this corresponds more accurately with the degradation state of the

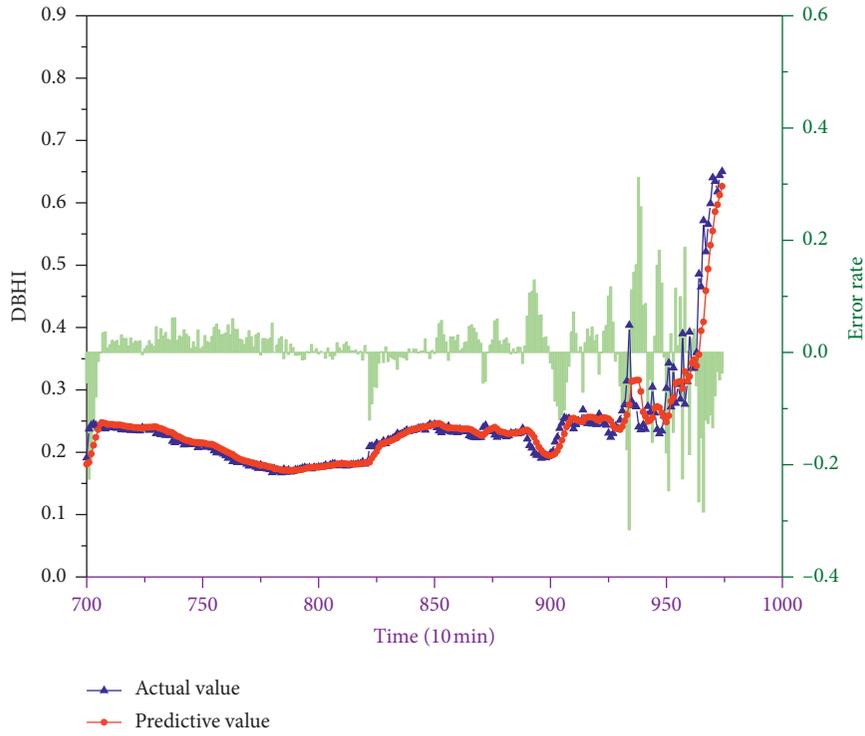


FIGURE 15: Prediction results and prediction error rates based on DBHI and LSTM.

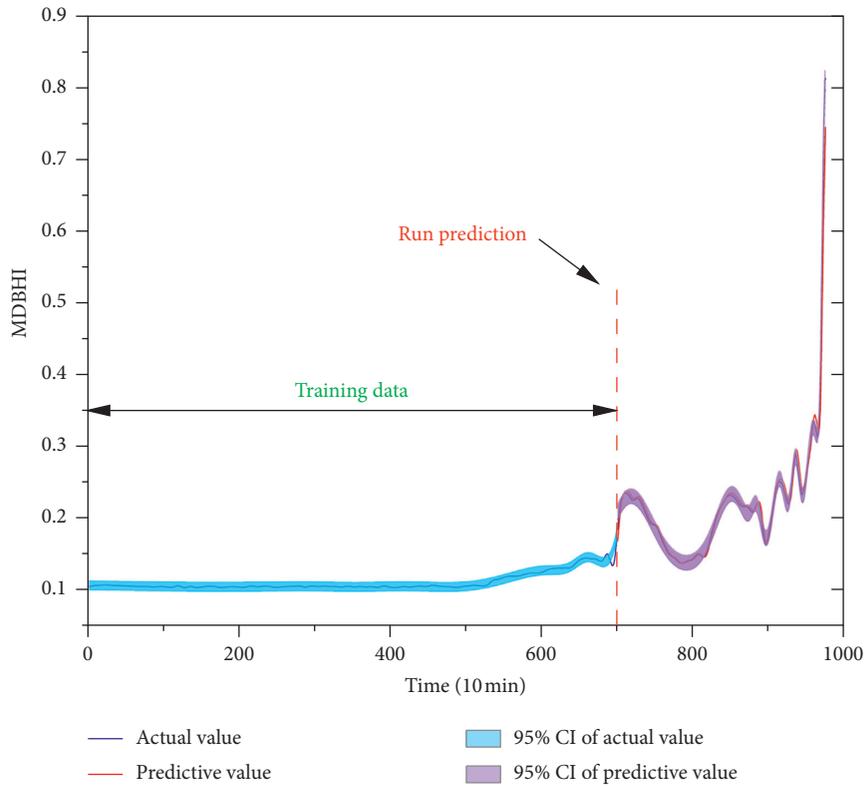


FIGURE 16: Confidence intervals for the prediction results of the proposed method.

TABLE 4: Comparison of five prediction performance measures.

Method of prediction	RMSE	MAE	R^2
MDBHI-BiLSTM	0.0168	0.0070	0.9655
SAE-LSTM	0.0247	0.0127	0.9065
PCA-LSTM	0.0283	0.0149	0.8903
RMS-RNN	0.0433	0.0213	0.8169
RMS-BP	0.0689	0.0453	0.5360

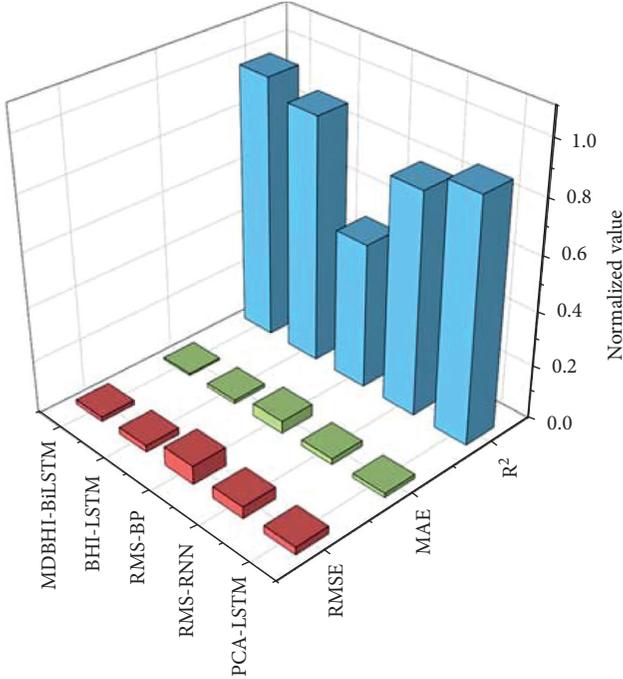


FIGURE 17: Comparison of five prediction performance measures.

bearing. Therefore, MDBHI is more conducive to the prediction of the degradation trend of the bearing in the overall trend.

4.3. Bidirectional LSTM Networks Trend Prediction Results. Finally, BiLSTM is used to predict the bearing degradation trend. The specific implementation steps are as follows.

Step 1 The construction of the predictive model is first carried out. In order to verify the validity of the model, MDBHI parameters must be divided into two parts; then the appropriate sample pairs are built based on the embedding theory.

Step 2 The BiLSTM model is trained using training samples to achieve convergence. The predicted samples are then input into the trained BiLSTM model to obtain predicted values.

According to the method proposed in this paper, 100 epoch models were trained to achieve convergence. The model prediction results and prediction error rates in the test samples are shown in Figure 14. According to the results, it can be seen that the prediction method proposed in this

paper can accurately predict the development of bearing degradation.

The prediction results and prediction error rates of the LSTM based on DBHI are shown in Figure 15. Compared with the two methods, the prediction trend using DBHI and LSTM is less accurate. And the prediction error is larger than the proposed method.

To further demonstrate the effectiveness of the presented method, confidence intervals (CI) for the prediction results obtained using the method are presented in Figure 16. It can be seen from the figure that the prediction component is almost entirely within the 95% CI, demonstrating that the prediction of bearing degradation tendency with the proposed method is effective.

Root mean square error (RMSE), mean absolute error (MAE), and the decision coefficient (R^2) are used to evaluate the accuracy of the prediction and compare it with traditional bearing trend prediction methods such as LSTM network based on SAE (SAE-LSTM), LSTM network based on PCA (PCA-LSTM), circulatory neural network based on statistical characteristic root mean square (RMS-RNN), and BP neural network based on statistical characteristic root mean square (RMS-BP). The results are shown in Table 4 and Figure 17:

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

$$e_{\text{MAE}} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}.$$

From the perspective of prediction errors, compared with traditional methods, the prediction effect based on MDBHI and BiLSTM is better than others. This result indicates that the health indicator (BHI) after removing detailed fluctuation components can successfully reflect the tendency of bearing degradation and will improve the final prediction effect. In addition, the introduction of BiLSTM can effectively utilize the characteristic information of bearing degradation trends, thereby minimizing prediction errors and improving the prediction effect.

5. Conclusions

A method of bearing degradation trend prediction based on MDBHI and BiLSTM was proposed in this paper. The method can solve the problem that a simple feature cannot accurately describe the bearing degradation process. In terms of HI extraction, the MDBHI extracted from the selected multiple features by SAE has a superior ability to represent the bearing degradation process. And after multiscale decomposition, MDBHI solves the problem of high-frequency components in DBHI that affect the trend representation. In terms of prediction methods, based on the BiLSTM network, the past and future information is fully

utilized to improve prediction accuracy. According to the experimental results, the proposed method based on MDBHI and BiLSTM obtains higher prediction accuracy. Therefore, this method has broad application prospects in bearing condition assessment and trend prediction.

As future works, after HI are extracted, some appropriate methods should be used to judge the starting point of the fault. This will help determine a more accurate time to start prediction. In addition, end-to-end forecasting methods should also be explored.

Data Availability

The manuscript uses public data from the degradation database of four roller bearings provided by the University of Cincinnati.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was partially supported by the National Natural Science Foundation of China (Grant no. 52075031) and the Central Funds Guiding the Local Science and Technology Development (Grant no. 206Z5001G). This support is gratefully acknowledged.

References

- [1] A. Ghods and H.-H. Lee, "Probabilistic frequency-domain discrete wavelet transform for better detection of bearing faults in induction motors," *Neurocomputing*, vol. 188, pp. 206–216, 2016.
- [2] G. Li, G. Tang, G. Luo, and H. Wang, "Underdetermined blind separation of bearing faults in hyperplane space with variational mode decomposition," *Mechanical Systems and Signal Processing*, vol. 120, pp. 83–97, 2019.
- [3] X. Jiang, C. Shen, J. Shi, and Z. Zhu, "Initial center frequency-guided VMD for fault diagnosis of rotating machines," *Journal of Sound and Vibration*, vol. 435, pp. 36–55, 2018.
- [4] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, pp. 98–109, 2017.
- [5] J. Z. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," *Mechanical Systems and Signal Processing*, vol. 25, no. 5, pp. 1803–1836, 2011.
- [6] G. Tang, Y. Wang, Y. Huang, N. Liu, and J. He, "Compound bearing fault detection under varying speed conditions with virtual multichannel signals in angle domain," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 8, pp. 5535–5545, 2020.
- [7] B. Yang, R. Liu, and E. Zio, "Remaining useful life prediction based on a double-convolutional neural network architecture," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9521–9530, 2019.
- [8] G. Tang, Y. Zhang, and H. Wang, "Multivariable LS-SVM with moving window over time slices for the prediction of bearing performance degradation," *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 6, pp. 3747–3757, 2018.
- [9] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mechanical Systems and Signal Processing*, vol. 104, pp. 799–834, 2018.
- [10] Y. Wang, Y. Peng, Y. Zi, X. Jin, and K.-L. Tsui, "A two-stage data-driven-based prognostic approach for bearing degradation problem," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 924–932, 2016.
- [11] J. Yu, "A hybrid feature selection scheme and self-organizing map model for machine health assessment," *Applied Soft Computing*, vol. 11, no. 5, pp. 4041–4054, 2011.
- [12] H. Ocak, K. A. Loparo, and F. M. Discenzo, "Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: A method for bearing prognostics," *Journal of Sound and Vibration*, vol. 302, no. 4-5, pp. 951–961, 2007.
- [13] W. Ahmad, S. A. Khan, and J.-M. Kim, "A hybrid prognostics technique for rolling element bearings using adaptive predictive models," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1577–1584, 2018.
- [14] L. Liao, "Discovering prognostic features using genetic programming in remaining useful life prediction," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 5, pp. 2464–2472, 2014.
- [15] H. Li, Y. Li, and H. Yu, "A novel health indicator based on cointegration for rolling bearings' run-to-failure process," *Sensors*, vol. 19, no. 9, p. 2151, 2019.
- [16] D. Wang, P. W. Tes, W. Guo, and Q. Miao, "Support vector data description for fusion of multiple health indicators for enhancing gearbox fault diagnosis and prognosis," *Measurement Science and Technology*, vol. 22, no. 2, Article ID 025102, 2011.
- [17] N. Gebraeel, M. Lawley, R. Liu, and V. Parmeshwaran, "Residual life predictions from vibration-based degradation signals: A neural network approach," *IEEE Transactions on Industrial Electronics*, vol. 51, no. 3, pp. 694–700, 2004.
- [18] S. A. Aye and P. S. Heyns, "An integrated Gaussian process regression for prediction of remaining useful life of slow speed bearings based on acoustic emission," *Mechanical Systems and Signal Processing*, vol. 84, pp. 485–498, 2017.
- [19] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, 2018.
- [20] C. Shen, Y. Qi, J. Wang, G. Cai, and Z. Zhu, "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 170–184, 2018.
- [21] H. Wang, S. Li, L. Song, L. Cui, and P. Wang, "An enhanced intelligent diagnosis method based on multi-sensor image fusion via improved deep learning network," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 2648–2657, 2020.
- [22] A. Malhi, R. Yan, and R. X. Gao, "Prognosis of defect propagation based on recurrent neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 3, pp. 703–711, 2011.
- [23] J. Deutsch and D. He, "Using deep learning-based approach to predict remaining useful life of rotating components," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 1, pp. 11–20, 2018.
- [24] J. Zhu, N. Chen, and W. Peng, "Estimation of bearing remaining useful life based on multiscale convolutional neural network," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 4, pp. 3208–3216, 2018.

- [25] B. Zhang, S. Zhang, and W. Li, "Bearing performance degradation assessment using long short-term memory recurrent network," *Computers in Industry*, vol. 106, pp. 14–29, 2019.
- [26] F. Wang, X. Liu, G. Deng, X. Yu, H. Li, and Q. Han, "Remaining life prediction method for rolling bearing based on the long short-term memory network," *Neural Processing Letters*, vol. 50, no. 3, pp. 2437–2454, 2019.
- [27] A. Z. Hinchí and M. Tkiouat, "Rolling element bearing remaining useful life estimation based on a convolutional long-short-term memory network," *Procedia Computer Science*, vol. 127, pp. 123–132, 2018.
- [28] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.
- [29] Z. Liu, M. J. Zuo, and Y. Qin, "Remaining useful life prediction of rolling element bearings based on health state assessment," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 230, no. 2, pp. 314–330, 2016.
- [30] J. Lee, H. Qiu, G. Yu et al., *Rexnord Technical Services. "Bearing Data Set", NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA, USA, 2007.*
- [31] I. El-Thalji and E. Jantunen, "A descriptive model of wear evolution in rolling bearings," *Engineering Failure Analysis*, vol. 45, pp. 204–224, 2014.