IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 24, 2019, accepted November 13, 2019, date of publication November 19, 2019,
date of current version December 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2954091

Rolling Bearing Initial Fault Detection Using Long Short-Term Memory Recurrent Network

HUAITAO SHI[1], LEI GUO[1], SHUAI TAN[2], AND XIAOTIAN BAI[1]

[1]School of Mechanical Engineering, Shenyang Jianzhu University, Shenyang 110168, China
[2]Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China

Corresponding author: Shuai Tan (tanshuai@ecust.edu.cn)

This work was supported in part by the National Key Research and Development Plan under Gant 2017YFC0703903, in part by the National Natural Science Foundation of China under Gant 51705341 and Gant 51905357, and in part by the Natural Science Foundation of Liaoning under Gant 2019-ZD-0654.

ABSTRACT The complete failure of the rolling bearing is a deterioration process from the initial minor fault to the serious fault, it is meaningless for guiding maintenance when the serious fault is alarmed. This work presents a novel initial fault diagnosis framework based on sliding window stacked denoising auto-encoder (SDAE) and long short-term memory (LSTM) model. In this approach, multiple vibration value of the rolling bearings are entered into SDAE by sliding window processing. Then, multiple vibration value of the rolling bearings of the next period is predicted from the signal reconstructed by the trained SDAE in the previous period using LSTM. For the given input data, the reconstruction errors between the next period data and the output data generated by trained LSTM are used to detect initial anomalous conditions. The proposed method not only utilizes the ability of SDAE to learn the inherent distribution of data, but also ensures that LSTM can extract timing relationships between data cycles, and the model is built using only normal data. The initial fault detection as a key difficulty in the operating condition monitoring and performance degradation assessment of the rolling bearing is effectively solved. Experimental and classic rotating machinery datasets have been employed to testify the effectiveness of the proposed method and its preponderance over some state-of-the-art methods. The experiment results indicate that the proposed method can effectively detect the initial anomalies of the rolling bearing and accurately describe the deterioration trend with strong robustness, and have high significance for maintenance guiding.

INDEX TERMS Rolling bearing, fault diagnosis, long-short-term memory.

I. INTRODUCTION

Rolling bearings are one of the key components of rotating machinery. According to [1], more than 40% of motor failures are related to bearing faults. The degradation trend of most rolling bearings usually follows the ''U-shaped curve''. As shown in Figure 1, larger state parameters mean performance degradation and it consists of four stages: (I) Run-in phase, (II) Normal operation phase, (III) Early degradation phase, (IV) Severe failure phase. From the early degradation stage to the severe failure stage, the operating parameters of the machine change significantly in a short period of time. If potential faults are not detected in the early stages of degradation, progressive faults will cause a rapid deterioration of the equipment, it will lead to irreparable and serious consequences.

Hence, accurate initial fault detection [2], [3] and prediction [4], [5] are rewarding to decrease the maintenance cost and reduce costly downtime. There are two main challenges in the early stages of fault initiation and evolution: (1) The degree of initial failure is small, and the fault characteristics are not obvious. (2) The fault signal is hidden in strong background noise, and the fault feature is disturbed by noise and uncorrelated signals, which is difficult to extract.

Many effective methods have been proposed and applied for rolling bearing intelligent diagnosis [6]–[11], these strategies could be roughly classified into two categories [12]: model-driven and data-driven. The model-driven approach

The associate editor coordinating the review of this manuscript and approving it for publication was Xin Luo.

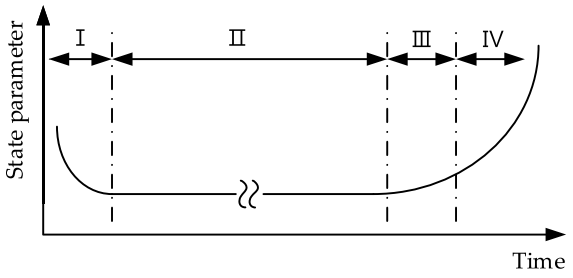


FIGURE 1. Deterioration trend of rotating machinery.

isolates early faults by establishing the system degradation model or analyzing the fault frequencies by kinetic analysis. Diagnostic performance depends on the accuracy of the model. Due to the complexity, nonlinearity and strong coupling of the equipment, it takes a lot of effort to fully understand the mechanical structure of the dynamic system and establish an accurate mathematical model. Data-driven methods indicate early faults by performing various analysis techniques on the collected data, which are more suitable to monitor the modern large-scale industrial process.

In traditional methods, vibration analysis [13]–[15] is the most used one and numerous indicators derived from vibration signals are implemented as criteria in traditional fault diagnosis. Approaches based on frequency domain signals are proposed in [16]. The accuracy of the frequency domain method will decrease when the signal is a non-stationary signal. Thus some signal processing approaches in time-frequency-domain are adopted (e.g. Wigner distribution [17], and wavelet packets [18], Empirical mode decomposition [19]). For early fault detection, The signal acquisition process is accompanied by the signal weakening, superposition and other phenomena. And the early fault signal is weak, the background noise is strong. So the signal processing based methods are limited in detecting early faults.

Some of statistical method based have been utilized to isolate the early faults in recent years, mainly including principal component analysis (PCA), autoregressive model, support vector machine (SVM) and its extensive forms: one-class SVM, support vector data description (SVDD).But these methods rely on well-selected features to classify faults. In recent years, deep learning has become one of the most popular machine learning approaches. In the field of fault diagnosis, Tamilselvan and Wang [20] conduct dimension reduction on signals using compressed sensing and utilize features extracted by deep belief network (DBN) to do fault classification. An end-to-end CNN approach is proposed by Chen *et al.* [21] for rolling bearing fault diagnosis. These methods focus on fault classification and rely on complete fault samples, not for early fault detection.

At present, there are few current algorithms make good use of the long-term dependencies hidden in time-series data. This long-term dependence refers to the temporal or spatial correlation in sequential observations. Normal system behavior strongly depends on the previous context and might evolve over time, in particular, an anomalous pattern is often defined as a series of behaviors that are normal individually but abnormal only collectively. Vibration signal is a time-series data arranged according to the time axis. Focusing on the full historical trajectories of sequential data is necessary to improve early fault detection. LSTM has been demonstrated its effectiveness in a wide range of machine learning problems that involve sequential data.

In summary, for the initial anomaly detection of rolling bearings, there are several problems to be solved.

(1) There is a large amount of raw normal data that is not well utilized.

(2) Many methods do not consider the timing dynamics of retained data when using the SDAE method.

(3) The timing dynamics of the data are not considered.

The essence of fault detection is to distinguish fault data from normal data, but usually, the collected data is non-stationary, non-linear, or damaged. To separate the fault data from the normal data, there are two ways to enhance the characteristics of the fault signal or to learn the robust characteristics of the normal signal. The signal characteristics for the initial anomaly are weak, the noise is strong, and it may be damaged.

To address the above issues, this work presents a novel initial fault diagnosis framework based on Sliding Window Stacked Denoising Auto-Encoder (SDAE) and Long Short-Term Memory (LSTM) model, we called SDLSTM. In this approach, To ensure that the SDAE learns more robust features from the normal state and does not lose the temporal dynamic characteristics of the data, multiple vibration value of the rolling bearings are entered into SDAE by sliding window processing (SWDAE). Then, multiple vibration value of the rolling bearings of the next period is predicted from the signal reconstructed by trained SWDAE in the previous period using LSTM. For the given input data, the reconstruction errors between the next period data and the output data generated by trained LSTM are used to detect initial anomalous conditions. We use the SDLTSM algorithm to learn the robustness and temporal correlation of normal data and learn the characteristics of normal data more completely.

Compared with other existing methods, the proposed method not only considers the nonlinear relationship between data, but also the temporal dynamics of the data is used to construct the prediction model to predict the trend of normal data. What is important is that the model is built using only normal data, and the initial failure of the rolling bearing can be detected by the error between the predicted value and the true value. Deep learning theory is used and does not rely on handcrafted features. and the model is built using only normal data. The initial fault detection as a key difficulty in the operating condition monitoring and performance degradation assessment of the rolling bearing is effectively solved.

The rest of the paper is organized as follows. Section II makes a brief review of basic theory, Section III illustrates proposed methods for rolling bearing initial fault diagnosis, Section IV is used for experiments, and Section V makes the conclusion.

TABLE 1. Notations.

Symbols	Definition
x	Sequence data with T steps or T samples
$\tilde{x}$	Corrupted sequence data
h	Hidden variable of SDAE
r	Output of SDAE
y	Output of LSTM
W, b	Weight matrix and bias for network connections
m	Continuous time steps in data process
ω	The size of the window in the sliding window

II. BASIC THEORY

In this section, notations, which are used in this paper frequently, are summarized in Table 1. We brief structures of denoising auto-encoder (DAE) and standard LSTM.

A. DENOISING AUTO-ENCODER

The basic auto-encoder (AE) is an unsupervised neural network, which contains an input layer, a hidden layer, and an output layer. The final goal of AE is that the output is basically consistent with the input, through the encoder and decoder. DAE is an improved method of AE, which is trained to reconstruct a data sample from its corrupted version. The x is a set of sequence data with T steps or T samples and $x = (x_1, x_2, \cdots, x_t, \cdots, x_T)$, $x_t \in R^{d_x}$, $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_t, \cdots, \tilde{x}_T)$ is corrupted version. Two common choices for data corruption are additive Gaussian (GS) noise and Zero-Masking (ZM) noise where a fraction of input values are randomly forced to 0. The f_e is encoding function that extracts features from the input data, and the formula for converting from the input layer to the hidden layer is

$$h = f_e(a_h) = \sigma(W_{xh}\tilde{x} + b_{xh}) \tag{1}$$

The σ is an activation function of sigmoid, The W_{xh} is the weight matrix from the input layer to the hidden layer, The b_{xh} is the coefficient of offset.

After getting the hidden layer, the output layer is calculated by the decoding function f_d. The formula for converting from the hidden layer to the output layer is

$$r = f_d(h) = \sigma(W_{hr}h + b_{hr}) \tag{2}$$

The σ is an activation function of sigmoid, The W_{hr} is the weight matrix from the input layer to the hidden layer, The b_{hr} is the coefficient of offset.

The goal of AE is to minimize the reconstruction error, therefore, the loss function for AE is written as

$$L_{AE} = \frac{1}{2T}\|r - x\|_F^2 \tag{3}$$

where $\|\|_F$ denotes *Frobenius* norm.

The unprocessed vibration signal contains a lot of noise, and the signals characteristics of the initial fault are weak, so most of the noise in the signal is removed by DAE processing, and the hidden layer extracts robust features from the signal.

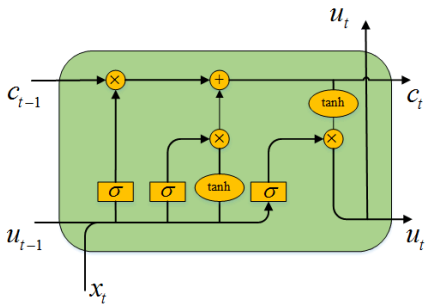


FIGURE 2. LSTM structure.

B. LONG-SHORT-TERM MEMORY

Recurrent neural networks (RNNs) are a family of neural networks for processing sequential data. The Long Short-Term Memory (LSTM) networks are a gated RNNs that effectively solves the problem of long-term dependence. As shown in Figure 2, the first sigmoid function determines what information is forgotten, called forget gate. What information is to be updated is determined by the second sigmoid function and the tanh function, called the input gate. The third sigmoid function determines which information is output, called the output gate. The updating equations are given as follows:

$$f_t = \sigma\left(W_f[u_{t-1}, x_t] + b_f\right) \tag{4}$$
$$i_t = \sigma(W_i[u_{t-1}, x_t] + b_i) \tag{5}$$
$$c_t = f_t \times c_{t-1} + i_t \times \tanh(W_c[u_{t-1}, x_t] + b_c) \tag{6}$$
$$o_t = \sigma(W_o[u_{t-1}, x_t] + b_o) \tag{7}$$
$$u_t = o_t \times \tanh(c_t) \tag{8}$$

where f_t, i_t, o_t are forget gate, input gate, output gate, respectively. LSTM can choose to forget redundant information, memorize useful information, and output important information through the gating unit. More details about LSTM can be found in the literature [20], [21].

III. PROPOSED METHOD

As mentioned before, the signals of the initial fault are weak and nonlinearly correlated and temporally dependent. In order to increase robustness against noise and local variations in original data, DAE is used by many methods. But each data sample at each time instance is regarded as independent. Therefore, the important timing correlation characteristics inherent in time series data are ignored. This will lead to reduced reliability of the fault detection model and unsatisfactory results. To address this problem, a fault detection method based on sliding window DAE and LSTM (SDLSTM) is proposed by this work. As shown in Figure 3, this method includes two stages: off-line construction and online monitoring.

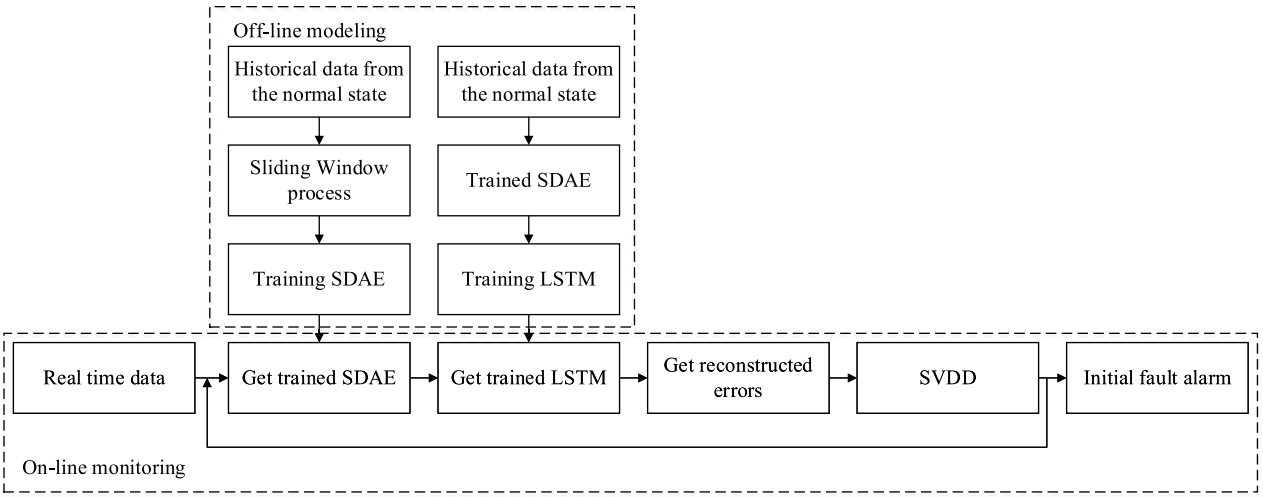


FIGURE 3. Flowchart of initial fault detection consisting of offline training phase and online monitoring phase.

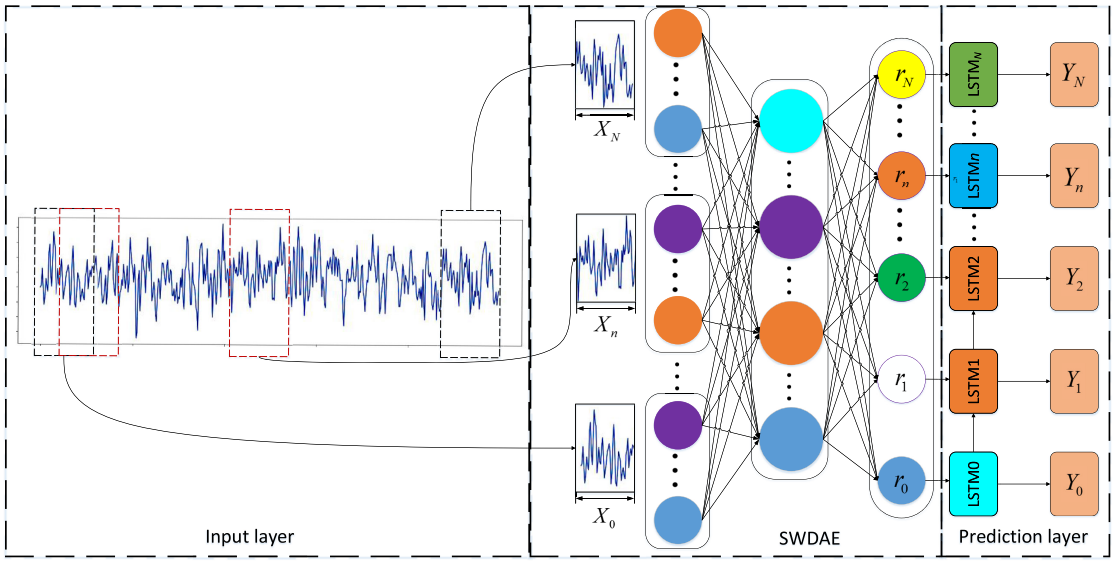


FIGURE 4. Framework of SDLSTM.

In the off-line modeling stage, the historical data collected from the normal state is first processed by the sliding window, and then the SWDAE is trained. The trained SWDAE can not only learn the robust features of the normal data but also have no loss of the timing dynamics of the data, and then fix the parameters of the SWDAE. The data reconstructed by SWDAE is used as the input of LSTM to learn the temporal correlation characteristics.

In the on-line monitoring stage, real-time data is used as input to the trained SDLSTM, trained LSTM will predict the value of the next time based on the reconstructed value by SWDAE at the current time. Then, the original signal acquired at the next time is reconstructed by SWDAE, The deviation value between the predicted value and the reconstructed value is used to measure the state of the system.

A. MODEL-STRUCTURE

SDLSTM can be understood as the use of deviations between predicted and actual values to measure whether the system deviates from normal operation. Figure. 4 shows the graphical structure of SDLSTM, which consists of three main modules, input layer, SWDAE, and prediction layer, respectively.

Let $X = \{x_i\}_{i=1}^{n} \in \mathrm{R}^{m \times n}$ be the collected normal sensor data from rolling bearing, where m is the number of sensor variables and n denotes the number of data samples. Since the motion of the rolling bearing is a periodic rotation, these variables have timing correlation characteristics. And the data collected by the sensor is mixed with noise. On the one hand, the noise will cause the characteristics of the data to be inconspicuous. And on the other hand, some data will be missing to some extent.

B. SLIDING-WINDOW

When training the DAE, it is very important to select the input size of the data. After the signal is reconstructed by the DAE, the characteristics of the signal are already obvious. For DAE without cyclic structure, the nonlinear characteristics of the data should be learned and the temporal correlation of LSTM learning should not be lost. So the form of the data input to the DAE is very important. However, small lengths of input data may result in loss of information, and too long input data may cause a large increase in model complexity and training time. Even if the input data has a suitable length, if the input data are independent of each other. The timing correlation of the data will be lost.

Thus, the sliding window is used to process the data and get the appropriate input data for the DAE. We first run an overlapping sliding window with length ω on the original multivariate time-series data. As shown in Figure. 4, we can obtain sets of overlapping windowed subsequences $s_j = \left(x_j, x_{j+1}, \ldots, x_{j+\omega 1}\right) \in \mathrm{R}^{m\times\omega}$, s_j can be reviewed as a small subset of accounting for temporally correlated information in several successive observed values for each variable, the represents the sequence of the windows. All subsequences constitute a new augmented time-series data $Y \in \mathrm{R}^{M\times N}$, $M = m\omega$, $N = n - \omega + 1$. It is obvious that contains the current and previous information in a small time window. Typically, if $\omega = 1$, is simply the original time-series data.

C. SDAE MODEL TRAINING

Afterward, we perform (1)–(3) on the data Y to learn optimal parameters W_{xh}, b_{xh}, W_{hr}, b_{hr}, and obtain the hidden representation h and the reconstructed output r. Since the new augmented data Y incorporates temporal information for each variable, the trained DAE model not only captures the nonlinear correlation between variables and removes noise, but also preserves the time dependence of the variables. Then, the learned relations are embedded in learned weight metrics and bias vectors.

The trained SWDAE already has the ability to denoise, and the reconstructed data does not lose its own timing correlation, then fix its parameters, and then train the LSTM.

D. LSTM MODEL TRAINING

The data reconstructed by well-trained SWDAE is r, so the input form of LSTM is shown in Figure 4.

$$Y = f\left(r;\theta\right), \theta = \left\{W_f, W_i, W_c, W_o, b_f, b_i, b_c, b_o\right\} \quad (9)$$

where r represents the denoised data reconstructed by SWDAE, θ represents the parameters of the model and Y is output. Generally, LSTM is constructed by minimizing the loss function as follows:

$$J\left(\theta\right) = \frac{1}{\omega}\sum_{i}\frac{1}{2l}\left\|y_i - r_i\right\|^2 + \frac{\lambda}{2}\sum\left(w_{ij}\right)^2 \quad (10)$$

where r_i^t and y_i^t represent the denoised reconstructed value of the original data and its predicted value for i-th dimension at time t, respectively. To avoid the overfitting problem, an additional regularization term is usually added to the cost function. The goal is to obtain minimum of cost function where the parameters are the optimal values for the model. Many optimization algorithms such as stochastic gradient descent (SGD) and RMSProp can serve for getting the minimum of the cost function. Considering about the non-stationarity of vibration signals, RMSProp approach is applied in this paper.

E. DIAGNOSIS STRATEGY

For the purpose of initial anomaly detection, SDLSTM is trained using normal data to have the ability to predict normal behavior of the system along the time axis. Therefore, when online data is input of the well-trained SDLSTM, it can reconstruct each input data and predict the value of the next moment. The different SDLSTM trained by datasets of various vibration patterns contain different characteristics. For instance, when the system is normal, the reconstruction error is very small. When the system is abnormal, the reconstruction error will be significantly bigger. More importantly, the reconstruction errors of different types of initial failures are also different. So, there is reason to believe that the operating state of the system can be judged by the reconstruction error. Various types of initial abnormal vibration signals and normal state vibration signals are input to the SDLSTM and the reconstruction error is obtained, and an SVDD is trained using the abnormal reconstruction error and the normal reconstruction error to indicate the operating state of the system.

The initial fault signal is weak and the signal may be damaged. And there are fewer fault samples in the actual process, so it is not suitable to use a supervised model for monitoring. So we consider estimating the health pattern by learning historical normal data and then measuring the damage based on the distance indicator

SDAE can learn more powerful data distribution characteristics by entering corrupted data and expecting to output data without damage. The gated design of the LSTM gives it a powerful advantage in analyzing the overall logic between sequence data. So SDAE can be used to learn the distribution characteristics of possible damaged signals, and LSTM is used to estimate the health pattern. The SADE cannot learn the temporal correlation from the data, which will cause the LSTM to estimate the health mode to be inaccurate. Therefore, it is necessary to perform sliding window processing on the data when training SDAE.

IV. EXPERIMENT RESULTS AND ANALYSIS

In order to verify the effectiveness of our proposed method in early fault detection, two classic data sets are used in this paper, i.e. a rolling bearing fault data set and a run-to-failure rolling bearing data set.

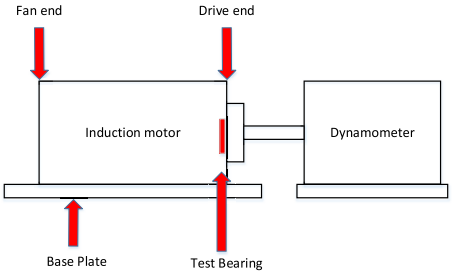


FIGURE 5. Testbed for CWRU rolling bearing data set.

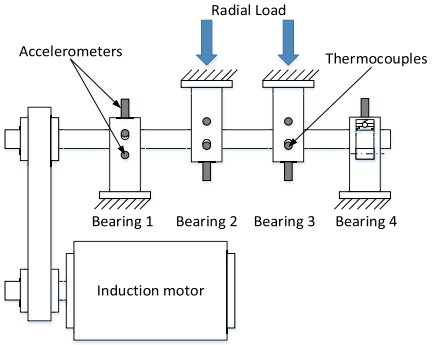


FIGURE 6. Testbed for IMS rolling bearing data set.

A. DATA DESCRIPTION

1) ROLLING BEARING FAULT DATA SET

The data set was acquired from the rolling bearing data center of the Case Western Reserve University (CWRU). This test stand consists of a motor, a torque transducer/encoder, a dynamometer, and control electronics (shown in Figure. 5). The type of rolling bearing used is SKF, single-point faults were introduced to the test rolling bearings using electro-discharge machining. The fault data set consists of four categories: normal state (N), and fault in the inner race (FI), the outer race (FO), and the ball (FB), respectively. For the same fault, the degree is 0.007, 0.014, 0.028, and the motor speeds of 1797 RPM. The test rolling bearings were reinstalled into the drive end. Accelerometers were placed at the 12 o'clock position at both the drive end and fan end of the motor housing, and the motor supporting base plate. Digital data was collected at 12,000 samples per second.

2) RUN-TO-FAILURE ROLLING BEARING DATA

The experimental data set was acquired from intelligent maintenance system (IMS) of University of Cincinnati [22]. The rolling bearing test rig hosts four rolling bearings on one shaft, which is driven by an ac motor (shown in Figure 6). The rotation speed kept 2000 r/min. And a radial load of 6000 lbs is added to the shaft and rolling bearing by a spring mechanism. An oil circulation system measures the flow and the temperature of the lubricant. Besides, a magnetic plug installed in the oil feedback pipe collects debris from the oil as evidence of rolling bearing system degradation. The system will stop while the accumulated debris adhered to the magnetic plug exceeds a certain level. A high sensitivity accelerometer was installed on each rolling bearing housing. Vibration data were recorded at 10-min intervals. The data sampling rate is 20 kHz and each vibration signal snapshot length contains 20480 points. At the end of the test-to-failure experiment, outer race failure occurred in rolling bearing 1 and 984 snapshots were collected. And first 500 snapshots represent system normal state and its detailed information is listed in Table 2.

TABLE 2. The running condition and fault type of run-to-failure rolling bearings.

Bearing	Set No.	Failure type	Rotation speed (rpm)	Load (lbs)	Failure time (h)
TB 1	2	Outer race defect	2000	6000	88
TB 3	1	Inner race defect	2000	6000	302
TB 4	1	Roller element defect	2000	6000	262

B. MODEL ESTABLISHMENT

1) INPUT PREPARATION

The vibration signal is different from the field in Natural Language Processing. The length of the vibration signal collected per second usually exceeds one thousand points, usually depending on the sampling frequency of the sensor. As the input sequence length for SDLSTM will directly affect the complexity and performance of the whole model, an effective input preparation strategy becomes quite important. In order to reduce the complexity of the model, and better use the useful information hidden in the data. So input data needs to be fused appropriately via the same input preparation process. This data-fusion strategy makes our method be extended to multi-sensor data. For a time-series signal segment from n sensors, a preparation function f_p is defined to form sampling points in m continuous time steps into one vector as the input at time step:

$$x_t = f_p\left(s_{mt:mt+m}\right) \tag{11}$$

In this work, we simply take the identity function as f_p and $x_t \in R^{m\times n}$ will be reshaped into R^{mn} as an input vector.

C. SDLSTM BASED INITIAL FAULT DIAGNOSIS

1) RESULTS ON ROLLING BEARING FAULT DATA SET

The CWRU data set is used to evaluate the fault diagnosis performance of the algorithm. As described in Section 4.1, the CWRU data contains four operating states common to the rolling bearing (N, FI FO, FB). This section uses a part of the normal data training model, and then data of four operating states are entered into the trained model and judges the state of the system by the reconstructed error. After the data preprocessing step, the data from the two sensors (DE, FE) is fused, in order to better evaluate the performance of the algorithm, the sample with the least fault diameter is

TABLE 3. Detailed descriptions of datasets, Num. denotes the number of samples for each type of.

Fault Location	Fault diameter (inches)	Num. (DE/FE)
Normal	0.000	120000
Inner Raceway	0.007	120000
Outer Raceway	0.007	120000
Ball	0.007	120000

TABLE 4. Structural hyperparameters of the proposed method for CRWU testing.

Hyper.	Value	Hyper.	Value
m	300	LSTM input size	800
ω	300	LSTM layers	1
DAE input size	800	LSTM hidden layer size	200
DAE layers	3	LSTM output size	800
DAE hidden layer size	400/200/100	Dense layers	1
DAE output size	800	Dense output size	800

selected, and its detailed information is listed in Table 3. For the proposed method, all structural hyperparameters are shown in Table 4.

In the model, we use the SDAE algorithm. Section 2.1 describes the SDAE algorithm in detail. In the face of two data collection difficulties that often exist in the actual process, namely noise influence and data loss, the SDAE algorithm can achieve more robust feature learning. Figure 7 shows the reconstruction results of the SDAE. In order to simulate the data loss problem, we randomly select the sample points of the two sensors in one cycle and set them to 0 with a probability of 0.2, and then reconstruct them by SDAE. Figure 7 shows, from top to bottom, raw normal data, damage data, and reconstructed data. It can be seen that SDAE can still reconstruct a distribution similar to the original data when the data is damaged.

In this experiment, the sampling frequency of the sensor is 12 kHz, and the motor speed is 1797 RPM, so 30 cycles of data can be acquired per second, and there are 400 sampling points per cycle. In order to learn the timing correlation between data period and period, the input and output size of the model is 400. Because the data collected by two sensors are used, the input and output size is 800. Each sensor selected a total of 120,000 sample points for 10 seconds, and 80,000 sample points were used to train the model, 20,000 for verification, and 20,000 for testing. After training the model with normal data, we test the model using different types of fault samples.

In the model, we use the SDAE algorithm. Section 2.1 describes the SDAE algorithm in detail. In the face of two data collection difficulties that often exist in the actual process, namely noise influence and data loss, the SDAE algorithm can achieve more robust feature learning. Figure 7 shows

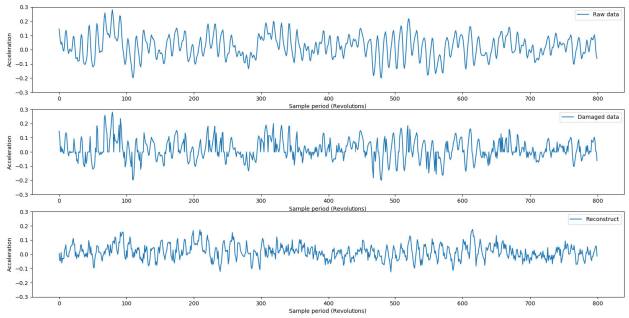


FIGURE 7. Reconstructed results after SDAE processing.

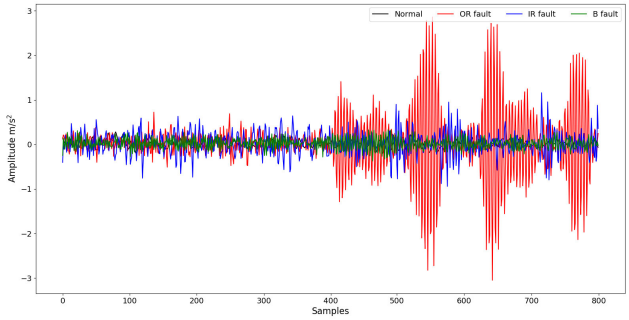


FIGURE 8. CWRU sample.

the reconstruction results of the SDAE. In order to simulate the data loss problem, we randomly select the sample points of the two sensors in one cycle and set them to 0 with a probability of 0.2, and then reconstruct them by SDAE. Figure 7 shows, from top to bottom, raw normal data, damage data, and reconstructed data. It can be seen that SDAE can still reconstruct a distribution similar to the original data when the data is damaged.

In Figure. 8, data for four cases of one cycle collected by two sensors. Black lines represent normal conditions, green lines represent rolling element faults, blue lines represent inner ring faults, red lines represent outer ring faults. Obviously, the rolling element failure is the easiest to distinguish, but the other two types of failure are difficult to distinguish from the normal situation. Figure 9 shows the test results of the model.

In Figure. 9, black lines represent normal conditions, green lines represent rolling element faults, blue lines represent inner ring faults, and red lines represent outer ring faults. As can be seen from Figure 9. (1) The reconstruction error of the normal (black line) conditions fluctuates around 2.5, the ranges of the reconstruction error of the abnormal state (IR, OR, B) are from 3.7 to 20. There is a clear distinction between the normal state and the abnormal conditions. Representing the ability of the model to detect anomalous conditions. (2) The ranges of the reconstruction error of rolling element faults (green line) are from 3.7 to 4.5, the reconstruction error of the inner ring faults (blue line) fluctuates around 7.5, the ranges of the reconstruction error of outer ring faults (red line) are from 3.7 to 4.5. Three different fault conditions

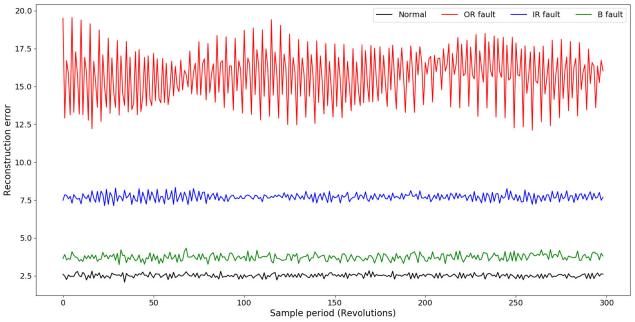


FIGURE 9. Results of CWRU sample diagnosis.

(IR, OR, B) can also be well distinguished, representing the ability of the model to diagnose faults.

And the advanced feature of this method is: (1) The construction and training of the model only use normal situation data. The distribution and characteristics of normal data are well learned by the model, and is sensitive to the abnormal situation of the change, indicating that the model can cope with the lack of fault data in the actual process. (2) After collecting enough types of fault data, we can label the data and add it to the model database to update the model and improve the model's fault diagnosis capabilities.

2) RESULTS ON RUN-TO-FAILURE ROLLING BEARING DATA SET

Unlike the CWRU data, the IMS data obtained through the accelerated aging test of the rolling bearing contains all the processes from the healthy running state to the failure of the rolling bearing. Figure 8 shows the acquisition data of rolling bearing 1 in the second group of experiments in IMS. After the end of the experiment, the outer ring fault occurred in the rolling bearing. It can be seen that as the sample point is increased, that is, the rolling bearing running time increases. The rolling bearing undergoes a process of gradual degradation from the initial state of health to the onset of an initial anomaly, followed by a failure, and finally a serious failure.

Figure 10 shows the original time domain vibration signal of the entire life cycle of TB1. It can be seen from the figure 10 that the time-domain vibration signal collected before about 740 files (about 100 hours) is relatively stable, and there is a very obvious impact phenomenon after about 750 files, the bearing has obvious damage, and short-lived troughs appear in several times of shock signals. The appearance of such troughs is a typical ''cure'' in the fault of the outer ring of a rolling bearing. The position where the damage occurs is gradually smoothed to smooth under the impact force. The degradation trend of the bearing is roughly seen from the original signal, but the initial stage of degradation of the rolling bearing and the stage of degraded performance cannot be directly obtained from the original time-domain signal.

We select the first 400 files to get normal data and only use that data to train the model. The remaining 584 files

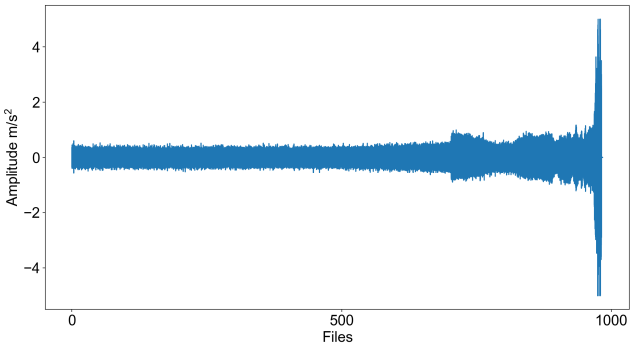


FIGURE 10. IMS rolling bearing degradation signal data.

TABLE 5. Structural hyperparameters of the proposed method for IMS testing.

Hyper.	Value	Hyper.	Value
	600	LSTM input size	600
	500	LSTM layers	1
DAE input size	600	LSTM hidden layer size	200
DAE layers	3	LSTM output size	600
DAE hidden layer size	300/200/100	Dense layers	1
DAE output size	600	Dense output size	600

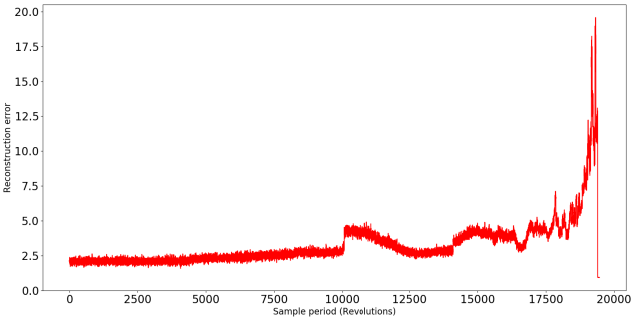
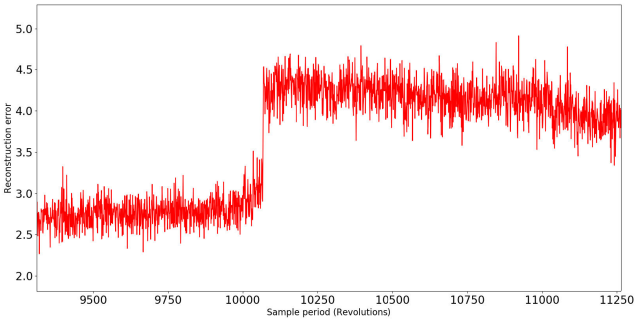


FIGURE 11. TB1 of IMS rolling bearing life cycle curve.

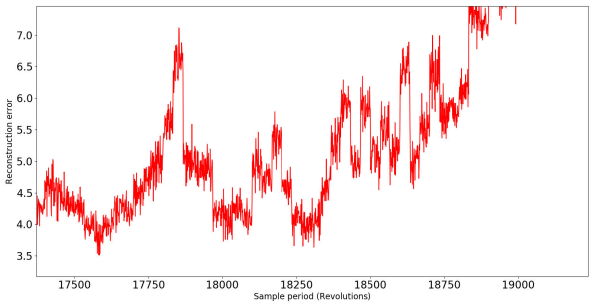
contain normal and fault conditions for evaluating the performance of the model. all structural hyperparameters are shown in Table 5.

The essence of fault detection is to distinguish fault data from normal data, but usually, the collected data is non-stationary, non-linear, or damaged. To separate the fault data from the normal data, there are two ways to enhance the characteristics of the fault signal or to learn the robust characteristics of the normal signal. The signal characteristics for the initial anomaly are weak, and the noise in the signal is strong, and the signal may be damaged during transmission. We use the SDLTSM algorithm to learn the robustness and temporal correlation of normal data and learn the characteristics of normal data more completely.

Figure 11 shows the performance of the SDLSTM algorithm in the IMS dataset. By using the normal data training model, the variation of the data under normal operating



(a) Local enlargement between sample period 10000 and 10500



(b) Local enlargement between sample period 18250 and 18750

FIGURE 12. Partial magnification of the IMS TB1 rolling bearing life curve.

conditions is well learned by the model, and the reconstruction errors of the predicted and true values are used to measure the status of the system. Since the IMS data set was obtained through the accelerated aging test of the rolling bearing, the rolling bearing finally suffered an outer ring failure, and it can be seen that the model first appeared abnormally in the 10070th cycles (As shown in Figure 12(a)) of the test data (About 90 hours from the beginning). And then the data tends to normal, which also meets the process of the rolling bearing from health to complete failure. When the rolling bearing has an initial abnormality in the outer ring, as the rolling bearing continues to operate, the small fault of the outer ring will be polished by the moving rolling element to smooth, and the abnormal vibration of the rolling bearing will gradually weaken, so there will be a short-term normal-like data. The rolling bearing with the initial fault continues to run, and this abnormal and similar normal data will be repeated (As shown in Figure 12(a)), but the time interval will gradually become shorter, and the degree of each abnormal signal will gradually increase until the devastating complete fault occurs.

Table 6 shows the test results of the SDLSTM algorithm proposed in this paper on the entire IMS data set. The TB1 rolling bearing has an abnormality around 88h, and finally develops an outer ring fault. The model detected an abnormality in about 90 hours, which was 2 hours later than the actual situation. The TB3 rolling bearing showed an abnormality around 302h, and finally developed into an inner ring fault. The model detected anomalies around 305 hours, 3 hours later than the real one. The TB4 rolling bearing showed an abnormality at about 262h, and finally developed into a rolling element failure. The model detected an abnormality of about 263 hours, which was one hour later than the real one.

TABLE 6. Fault detection results.

Bearing	Set No.	Failure type	Failure time (h)	Results (h)
TB 1	2	Outer race defect	88	90
TB 3	1	Inner race defect	302	305
TB 4	1	Roller element defect	262	263

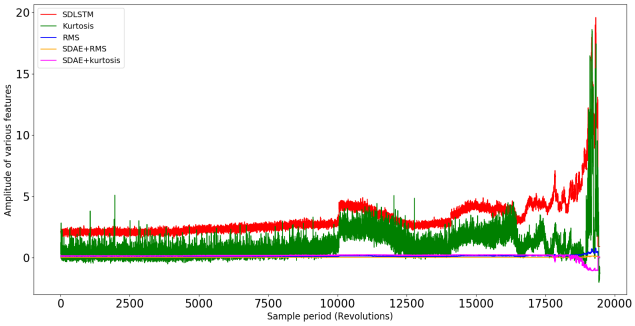


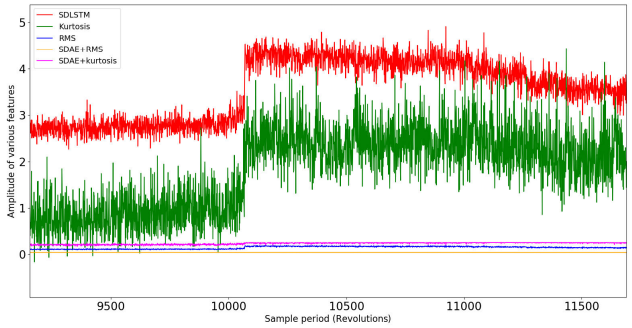
FIGURE 13. Comparison of various features.

D. COMPARISON WITH OTHER METHODS

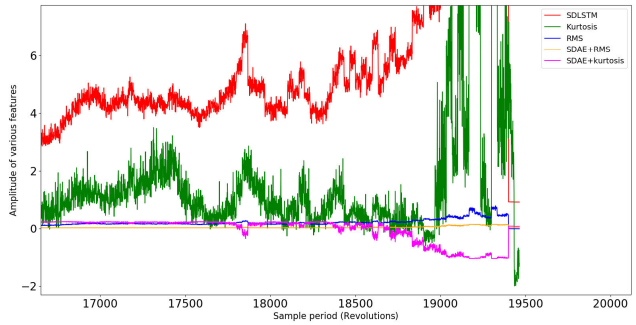
In order to verify the effectiveness of the SDLSTM model, the proposed method is compared with several fault detection and anomaly detection methods, namely, Kurtosis factor, and SADE features.

Figure 13 shows the trend graphs of the features extracted by several fault detection algorithms on the IMS dataset. Since we use the original time-domain data for fault detection, we compare the SDLSTM algorithm proposed in this paper with the commonly used method of extracting time-domain features. In the figure 13, the red line represents the SDLSTM algorithm, the green line represents the Kurtosis statistical indicator, the blue line represents the RMS statistical indicator, the orange line represents the characteristics of SDAE and RMS extraction, and the purple line represents the features extracted by SDAE and Kurtosis.

As shown in Figure 13, It is clear that SDLSTM can describe the development of damage, and very sensitive to initial anomalies through Figure 14 (a) and Figure 14 (b).For the Kurtosis, it is sensitive to abnormal changes in the signal, because Kurtosis characterizes the number of peaks of the probability density distribution curve at the mean value, reflecting the sharpness of the peak. When the running rolling bearing is abnormal, the vibration signal will have a transient spike, so Kurtosis will have a more obvious change. It can be seen from Figure 14 (a) that Kurtosis and the method proposed in this paper have similar effects on the diagnosis of initial abnormalities, but it can be seen from Figure 14 (b) that Kurtosis does not reflect the deterioration trend of the rolling



(a) Local enlargement between sample period 9500 and 11500



(b) Local enlargement between sample period 17000 and 19500

FIGURE 14. Partial magnification of the IMS TB1 rolling bearing life curve.

TABLE 7. Fault detection results of other methods in TB1.

Method	Result	Method	Result
SDLSTM	10070	RMS	17170
SWDAE+RMS	10752	Kurtosis	17103
SWDAE+Kurtosis	10152	SVM	10809

bearing very well. It is not possible to provide a very valuable diagnostic feature for the evaluation of the condition of the rolling bearing. The other two methods, through SDAE, have learned stable features, though, but are not sensitive to state changes.

SDLSTM method proposed in this work uses SVDD to indicate the signal position at the time of the fault. Therefore, we use SVDD in combination with several other methods in Table 7 to evaluate the effectiveness of different methods by comparing the operating time when a fault alarm occurs. According to the results, SDLSTM could declare the fault alarm at 10070 cycles, this result is better than other methods. In addition, the combination of SWDAE-processed data and time-domain statistical indicators has a better diagnostic effect than direct use of statistical indicators, which also indicates that SWDAE extracted data features are more stable.

V. CONCLUSION

This paper designed an end-to-end diagnostic strategy that uses only raw time-domain signals, independent of expert knowledge and signal processing knowledge. We called the SDLSTM algorithm, first, the sliding window algorithm is used in the SDAE algorithm, then, learning the trend of normal data over time by the LSTM algorithm. The reconstruction error of the predicted value and the true value is used to measure the system status. The algorithm was verified by experimental data, SDAE algorithm not only learns the distribution of data but also preserves the temporal correlation characteristics between the data, prediction of normal data behavior through LSTM. Compared with the SVM method, the initial anomaly was detected about 7 hours earlier by our proposed method, and the degradation process of rolling bearings is well described by the proposed model, which is important for the degradation monitoring of rolling bearings. By comparing the diagnostic results of SWDAE and traditional methods, it is verified that the SDAE algorithm processed by the sliding window can more effectively describe the data distribution with time-correlation characteristics and significantly improve the accuracy of diagnosis.

REFERENCES

[1] J. Zarei and J. Poshtan, "Bearing fault detection using wavelet packet transform of induction motor stator current," *Tribol. Int.*, vol. 40, no. 5, pp. 763–769, May 2007, doi: 10.1016/j.triboint.2006.07.002.

[2] L. Cui, J. Huang, F. Zhang, and F. Chu, "HVSRMS localization formula and localization law: Localization diagnosis of a ball bearing outer ring fault," *Mech. Syst. Signal Process.*, vol. 120, pp. 608–629, Apr. 2019, doi: 10.1016/j.ymssp.2018.09.043.

[3] E. El-Thalji and E. Jantunen, "A summary of fault modelling and predictive health monitoring of rolling element bearings," *Mech. Syst. Signal Process.*, vols. 60–61, pp. 252–272, Aug. 2015, doi: 10.1016/j.ymssp.2015.02.008.

[4] G. Prakash, S. Narasimhan, and M. D. Pandey, "A probabilistic approach to remaining useful life prediction of rolling element bearings," *Struct. Health Monit.*, vol. 18, no. 2, pp. 466–485, Mar. 2019, doi: 10.1177/1475921718758517.

[5] W. Guangbin, D. Moujun, H. Liangpei, and L. Long, "Prediction of bearing damage in wind turbines based on the quadratic root mean square of sub-band manifold," *Proc. Inst. Mech. Eng., C, J. Mech. Eng. Sci.*, vol. 232, no. 18, pp. 3213–3223, Sep. 2018, doi: 10.1177/0954406217735553.

[6] K. Zheng, J. Luo, Y. Zhang, T. Li, J. Wen, and H. Xiao, "Incipient fault detection of rolling bearing using maximum autocorrelation impulse harmonic to noise deconvolution and parameter optimized fast EEMD," *ISA Trans.*, vol. 89, pp. 256–271, Jun. 2019, doi: 10.1016/j.isatra.2018.12.020.

[7] M. Elforjani and S. Shanbr, "Prognosis of bearing acoustic emission signals using supervised machine learning," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5864–5871, Jul. 2018, doi: 10.1109/Tie.2017.2767551.

[8] L. Song, P. Chen, and H. Wang, "Vibration-based intelligent fault diagnosis for roller bearings in low-speed rotating machinery," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 8, pp. 1887–1899, Aug. 2018, doi: 10.1109/Tim.2018.2806984.

[9] Y. Hao, L. Song, L. Cui, and H. Wang, "A three-dimensional geometric features-based SCA algorithm for compound faults diagnosis," *Measurement*, vol. 134, pp. 480–491, Feb. 2019, doi: 10.1016/j.measurement.2018.10.098.

[10] H. T. Shi, X. T. Bai, K. Zhang, Z. N. Wang, and Z. M. Liu, "Spalling localization on the outer ring of hybrid ceramic ball bearings based on the sound signals," *IEEE Access*, vol. 7, pp. 134621–134634, 2019, doi: 10.1109/access.2019.2941982.

[11] H. T. Shi, X. T. Bai, K. Zhang, Y. H. Wu, and G. D. Yue, "Influence of uneven loading condition on the sound radiation of starved lubricated full ceramic ball bearings," *J. Sound Vib.*, vol. 461, Nov. 2019, Art. no. 114910, doi: 10.1016/j.jsv.2019.114910.

[12] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6418–6428, Nov. 2014, doi: 10.1109/Tie.2014.2301773.

[13] S. N. Chegini, A. Bagheri, and F. Najafi, "Application of a new EWT-based denoising technique in bearing fault diagnosis," *Measurement*, vol. 144, pp. 275–297, Oct. 2019.
[14] X. Yan, Z. Sun, J. Zhao, Z. Shi, and C.-A. Zhang, "Fault diagnosis of rotating machinery equipped with multiple sensors using space-time fragments," *J. Sound Vib.*, vol. 456, pp. 49–64, Sep. 2019, doi: 10.1016/j.jsv.2019.05.036.
[15] L. Cui, B. Li, J. Ma, and Z. Jin, "Quantitative trend fault diagnosis of a rolling bearing based on sparsogram and Lempel-Ziv," *Measurement*, vol. 128, pp. 410–418, Nov. 2018, doi: 10.1016/j.measurement.2018.06.051.
[16] Y. Xin and S. Li, "Novel data-driven short-frequency mutual information entropy threshold filtering and its application to bearing fault diagnosis," *Meas. Sci. Technol.*, vol. 30, no. 11, Nov. 2019, Art. no. 115006, doi: 10.1088/1361-6501/ab2ff3.
[17] A. B. Zoubi, S. Kim, D. O. Adams, and V. J. Mathews, "Lamb wave mode decomposition based on cross-Wigner-Ville distribution and its application to anomaly imaging for structural health monitoring," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 66, no. 5, pp. 984–997, May 2019, doi: 10.1109/tuffc.2019.2903006.
[18] W. Hu, H. Chang, and X. Gu, "A novel fault diagnosis technique for wind turbine gearbox," *Appl. Soft Comput.*, vol. 82, Sep. 2019, Art. no. 105556, doi: 10.1016/j.asoc.2019.105556.
[19] R. U. Maheswari and R. Umamaheswari, "Fault diagnostics of wind turbine drive-train using multivariate signal processing," *Int. J. Acoust. Vib.*, vol. 24, no. 2, pp. 334–342, Jun. 2019.
[20] P. Tamilselvan and P. Wang, "Failure diagnosis using deep belief learning based health state classification," *Rel. Eng. Syst. Saf.*, vol. 115, pp. 124–135, Jul. 2013, doi: 10.1016/j.ress.2013.02.022.
[21] Z. Chen, C. Li, and R.-V. Sanchez, "Gearbox fault identification and classification with convolutional neural networks," *Shock Vib.*, vol. 2015, Apr. 2015, Art. no. 390134, doi: 10.1155/2015/390134.
[22] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *J. Sound Vib.*, vol. 289, nos. 4–5, pp. 1066–1090, 2006.

HUAITAO SHI received the B.S. and Ph.D. degrees in control engineering from Northeastern University, Shenyang, China, in 2005, and 2012, respectively. He has been a Professor with the Faculty of Mechanical Engineering, Shenyang Jianzhu University, Shenyang, where he has also been the Vice Dean, since 2014. He is the author of more than 30 articles, (16 articles were indexed by SCI), and five patents. His current research interest includes hybrid ceramic ball-bearing.

LEI GUO was born in 1994. He received the B.S. degree in mechanical design manufacture and automation major from Jiamusi University, China, in 2017. He is currently pursuing the master's degree in mechanical engineering with the School of Mechanical Engineering, Shenyang Jianzhu University, Shenyang, China. His research interest includes faults detection of mechanical systems.

SHUAI TAN received the B.S. degree in automation and the Ph.D. degree in control theory and control engineering from Northeastern University, China, in 2005 and 2012, respectively. She is currently an Associate Professor with the East China University of Science and Technology, Shanghai, China. Her research interests include operation state evaluation for complex industrial process, fault monitoring and diagnosis, and machine learning of image information.

XIAOTIAN BAI was born in Wanghua, Fushun, Liaoning, China, in 1989. He received the B.S. degree in mechanical engineering from the Dalian University of Technology, Dalian, in 2011, and the M.S. and Ph.D. degrees in mechanical engineering from the Shenyang University of Technology, in 2013 and 2016, respectively. He has been a Lecturer with the faculty of Mechanical Engineering, Shenyang Jianzhu University, Shenyang, China, since 2016, where he carried out his Postdoctoral Research work in the postdoctoral station, from 2016 to 2018. He has also been in Romania for three months as a Visiting Scholar. He is the author of more than 15 articles, and two patents. His current research interest is about the vibration and sound radiation of bearings.