

Research Article

Fault Diagnosis of Induction Motors Using Recurrence Quantification Analysis and LSTM with Weighted BN

Dengyu Xiao , Yixiang Huang , Chengjin Qin , Haotian Shi, and Yanming Li

State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai, China

Correspondence should be addressed to Dengyu Xiao; xiaodengyu@sjtu.edu.cn

Received 17 October 2018; Revised 15 December 2018; Accepted 3 January 2019; Published 27 January 2019

Academic Editor: José J. Rangel-Magdaleno

Copyright © 2019 Dengyu Xiao 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.

Motor fault diagnosis has gained much attention from academic research and industry to guarantee motor reliability. Generally, there exist two major approaches in the feature engineering for motor fault diagnosis: (1) traditional feature learning, which heavily depends on manual feature extraction, is often unable to discover the important underlying representations of faulty motors; (2) state-of-the-art deep learning techniques, which have somewhat improved diagnostic performance, while the intrinsic characteristics of black box and the lack of domain expertise have limited the further improvement. To cover those shortcomings, in this paper, two manual feature learning approaches are embedded into a deep learning algorithm, and thus, a novel fault diagnosis framework is proposed for three-phase induction motors with a hybrid feature learning method, which combines empirical statistical parameters, recurrence quantification analysis (RQA) and long short-term memory (LSTM) neural network. In addition, weighted batch normalization (BN), a modification of BN, is designed to evaluate the contributions of the three feature learning approaches. The proposed method was experimentally demonstrated by carrying out the tests of 8 induction motors with 8 different faulty types. Results show that compared with other popular intelligent diagnosis methods, the proposed method achieves the highest diagnostic accuracy in both the original dataset and the noised dataset. It also verifies that RQA can play a bigger role in real-world applications for its excellent performance in dealing with the noised signals.

1. Introduction

An induction motor is one of the most critical components in industrial processes due to its high reliability, low cost, and robust performance. It has been widely used as the key machine dynamical equipment to generate electromagnetic torque. However, the mechanical degradation with natural aging process, coupled with the fact that motors are often exposed to multifarious harsh environments, makes motors vulnerable to various sorts of faults [1]. Any unexpected motor failure may lead to the unexpected downtime and repair expense and even cause human casualties. Therefore, accurate, effective, and reliable motor fault diagnosis approaches have received considerable attention from academic research to guarantee the enhancement of security and avoid the downtime losses [2]. Various sensing techniques for signals of vibration, noise, voltage, current, flux, etc. [3–5] have been used in the motor fault diagnosis during the past ten decades. Among those techniques, vibration-

based sensing is most frequently used due to its easier acquisition of more detailed information [3].

Recently, owing to the significant development of the computing ability [6], massive efforts for motor fault diagnosis have been devoted to the data-driven approaches. For instance, Diazet provided an experimental comparative evaluation of various classifiers including k-NN, Bagging, AdaBoost, and SVM for motor fault detection [7]. Zhang presented a motor fault identification method through sparse representation and achieved good robustness to noise [8]. Pan improved the reliability of motors with feature extraction of entropy and SVM classifier [9]. Currently, several deep learning algorithms, which are thought as effective tools for the high-level feature learning based on the nonlinear transformations through multiple layers, are widely used to discover the underlying feature representations [10]. Several deep learning techniques such as sparse autoencoder (SAE) [11], deep stacking network (DSN) [12], and convolutional neural network (CNN) [13] are used for

identifying motor faults. Wang obtained the time-frequency scale-down maps using short-time Fourier transform (STFT) and classified the faulty motors using CNN [13]. Sun proposed a fast diagnosis method through convolutional discriminative learning of a BP network [14]. LSTM, as a significant branch of RNN, capable of addressing data sequences of varying length, encoding temporal information, and capturing long-term dependencies [6], has been successfully used in several machinery fault diagnosis tasks [15–18]. Recurrence quantification analysis (RQA) is a useful tool to deeply investigate the mechanical vibration dynamic properties, owing to its good capability in dealing with nonstationary data sequences, especially the data sequences involving continuous fluctuations and presenting nonlinear characteristics [19]. RQA has proven its effectiveness in bearing fault detections [20, 21].

However, the abovementioned methods still have some shortcomings. The performances of the traditional methods [7–9] generally depend on the manual feature extraction techniques, which may be incapable of fully exploiting the raw data and may omit some important information. Besides, they have limited adaptability when facing the new diagnosis issues or application objects. For the deep learning methods [11–14], the black-box characteristics make the intermediate representations hard to explain, and the lack of domain expertise weakens the interpretability. In some papers like [13, 18], the deep learning techniques only act as classifiers, which are overqualified and constrain the deep learning's capability to discover the underlying features in an end-to-end fault diagnosis. LSTM and RQA have proven effective in machinery fault diagnosis, mainly aiming at bearings [15–21]. So far, neither RQA nor LSTM has been applied on motors. Furthermore, the deeper research about RQA's adaptability and robustness to different levels of noise has never been considered.

To tackle those issues, a deep learning framework is integrated with manual feature learning techniques to preserve the advantages of both sides. RQA is seamlessly embedded into the stacked LSTM architecture. RQA's antinoise capability is verified by a weighted BN layer. The major contributions of this paper are summarized as follows:

- (i) Propose a new hybrid feature learning approach that combines statistical parameters, RQA and LSTM neural network, for motor fault detection. Neither RQA nor LSTM has been applied on motors before.
- (ii) A modification of BN named weighted BN is designed to assign dynamic weights to the hybrid feature set and then perform batch normalization. It shares BN's advantages and possesses the capability to evaluate the contributions of the three various feature learning approaches [22].
- (iii) Two datasets, respectively, the original dataset and the noised dataset, are acquired from tests to verify the proposal's adaptability and robustness to different levels of noise.

The remainder of this paper is organized as follows. In Section 2, tests of 8 three-phase induction motors with

different fault types on the drivetrain diagnostics simulator platform are introduced. Then, the related theories of RQA and LSTM are presented, and the specific procedures of the proposed method are described in detail in Section 3. Then, in Section 4, the performance of the proposed method is illustrated and discussed. Several other methods are tested and contrasted. Finally, Section 5 provides concluding remarks.

2. Description of Motor Tests and Acquired Data

2.1. Tests Description. In order to experimentally verify the performance of this proposed approach, tests of 8 three-phase induction motors with different fault types under a uniform operation condition were carried out in a drivetrain diagnostics simulator platform. As shown in Figure 1(a), this system is composed of a three-phase induction motor, a magnetic brake for loading, a two-stage fixed-axis gearbox, and a two-stage planetary gearbox. This test rig meets the all configuration's requirement of vibration analysts and provides a practical and reliable test environment for motor diagnosis. A data acquisition card NI-9234, shown in Figure 1(b), is selected as the dynamic signal acquisition (DSA) module for making high-accuracy measurements from sensors. The ADC resolution is 24 bits, the dynamic range is 102 dB, and the sampling rate per channel is 51.2 kS/s. The selected acceleration sensor BW-BJ14530, shown in Figure 1(c), belongs to the voltage-output type. The sensibility ($\pm 5\%$) is 100 mV/g, and the measurement range is 1–5 kHz.

Specifically, the tests were under the operating condition of 33.90 N m (25 lbf. ft.) load. The motor frequency was set to 15 Hz, and thus, the rotation rate was 900 r/min. Under this uniform operating condition, 8 motors with different fault types (1 healthy and 7 faulty) were used to generate the required dataset. Data from the healthy motor are used as a benchmark for comparison with the experimental data from other faulty motors. The faulty types are as follows: (1) built-in broken rotor bars (BB), (2) built-in bowed rotor (BR), (3) rotor misalignment (RM), (4) stator winding faults (SW), (5) voltage unbalance and single phasing (VP), (6) built-in rotor unbalance (RU), and (7) faulted bearings (FB). The faulty types and the corresponding causes are listed in Table 1.

The acceleration sensor was used in the experiment and installed at the shell of motors to measure the vibration signals in radial direction. The acceleration signals were collected by the data acquisition card, and the sampling frequency was set to 10.24 kHz. Each test lasted approximately 120 s, and thus, the number of raw data points acquired from one motor is approximately 1228800. Figure 2 displays a set of time waveforms of the measured acceleration signals.

2.2. Datasets Generation. From the 120 s acceleration signals, the middle 100 s stable signals are selected as the training and test data. A sliding window is used to obtain the samples of the same length. Suppose the length of window as l , the step size s , and the number of data points N . One sample is generated for every step. Thus, the sample number n is

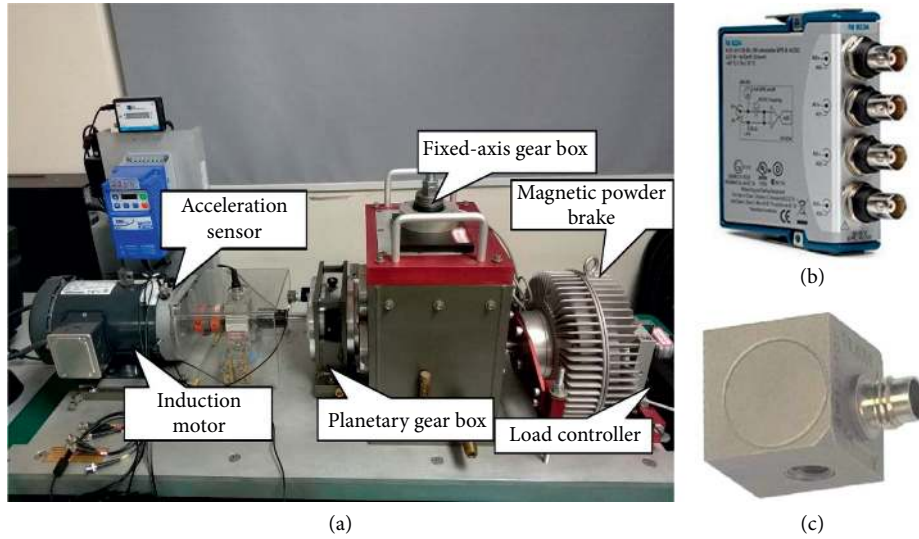


FIGURE 1: The test platform and data acquisition devices. (a) Test rig: drivetrain diagnostic simulator (DDS). (b) Data acquisition card NI-9234. (c) Acceleration sensor BW-BJ14530.

TABLE 1: Motors with 8 fault types in tests.

Fault condition	Abbreviations	Class	Description
Normal	NO	1	Healthy state
Broken rotor bars	BB	2	Fitted with an already broken rotor bar
Bowed rotor	BR	3	Consists of a centrally intentionally bent rotor
Rotor misalignment	RM	4	Caused by custom-machined end bells with asymmetric structure
Stator winding faults	SW	5	Copper wires around the stator with shorted stator winding turns
Voltage unbalance and single phasing	VP	6	Controlled by the control console to disrupt voltage balance and to disconnect one phase
Rotor unbalance	RU	7	Intentionally removing one of the balanced rotors from induction motor and destroying the inner balance
Faulted bearings	FB	8	Composed of one inner race faulted bearing and one outer race faulted bearing

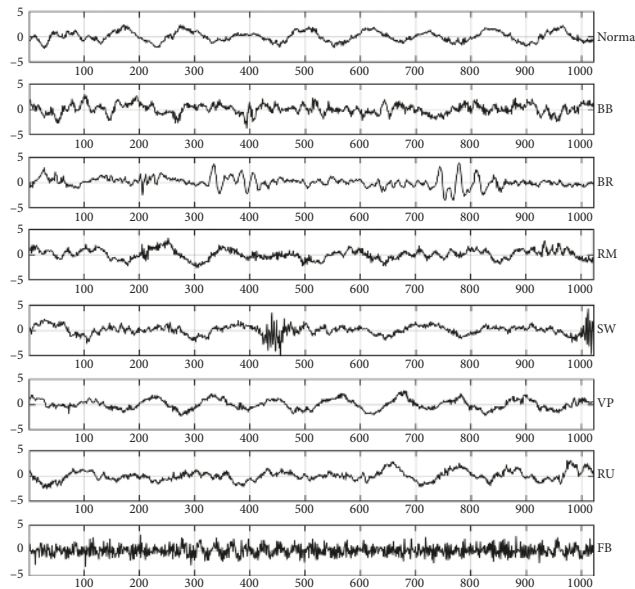


FIGURE 2: The waveforms of the measured acceleration signals of 8 motors.

$$n = \frac{N-l}{s} + 1. \quad (1)$$

In this paper, the step size and window length are both set to 1024. It means that every 0.1-second signal which contains 1024 data points (10.24 kHz) represents a sample. Thus, there are 1000 samples of each fault type, and those samples form the original dataset named dataset 1.

As the motor frequency is 15 Hz, one sample consists of 1.5 cycle of motor periodic rotation signals. Therefore, the difference between samples is more distinct, compared with the samples containing integral cycles of signals in previous works [11, 12, 14]. In this way, it can be demonstrated that our proposed framework attains the goal of fault diagnosis by revealing the underlying representative features instead of overfitting and memorization of training samples when the samples are almost the same.

For the sake of investigating the antinoise capability of the proposed algorithm and verifying its effectiveness in real-world fault diagnosis applications, Gaussian noise with a signal-to-noise ratio of 5 dB is artificially embedded into the dataset 1. This noising method is based on the assumption that the acceleration signals of motors in real world contain a higher-level Gaussian noise [23]. The processed dataset is named as dataset 2, which obviously elevates the difficulty level of fault diagnosis.

To obtain a precise diagnostic accuracy, a five-fold cross-validation method is adopted to split the training and test data. The dataset is segmented into 5 subsets, and the holdout method is repeated 5 times. Each time, one of the subsets is used as the test set and the rest as the training set.

3. The Proposed Method

This section describes the specific procedures of the proposed fault diagnosis method, illustrates the basic principles of the RQA, LSTM, and weighted BN, and presents the details of the whole neural network architecture. The flow diagram of the proposal is shown in Figure 3. The main steps are shown as follows:

- (1) Manual feature learning: draw the recurrence plots (RPs) of every sample and extract 10 RQA features from RPs; extract 29 empirical statistical features from time domain and frequency domain.
- (2) Deep learning feature learning: construct three-layer stacked LSTM of 0.25 dropout with hidden layer sizes of 256, 128, and 64.
- (3) Form the hybrid feature set with aforementioned features and put it into the weighted BN block which consists of a weight assignment layer and a BN layer.
- (4) Put the outputs of the previous step into a three-layer fully connected neural network with layer sizes of 64, 32, and 8, and the output is the diagnostic result.

3.1. RQA. RQA is a kind of nonlinear analysis for the dynamical system based on the view of a phase space concept,

aiming at quantifying the recurrence plots (RPs) [19]. Supposing a series $\{t_1, t_2, \dots, t_n\}$ with length n , it can be reconstructed to a new phase space T according to the Takens embedding theorem [24] and time-delay approach. T can be described as a reconstructed matrix:

$$\begin{cases} T(1) = \{t_1, t_{1+\tau}, \dots, t_{1+(m-1)\tau}\}, \\ \dots \\ T(i) = \{t_i, t_{i+\tau}, \dots, t_{i+(m-1)\tau}\}, \\ \dots \\ T(n-(m-1)\tau) = \{t_{n-(m-1)\tau}, t_{n-(m-2)\tau}, \dots, t_n\}, \end{cases} \quad (2)$$

where m is the embedding dimension, τ is the time delay, T is an $m \times m$ matrix, and $T(i)$ presents the i th row of T . Consequently, the recurrence matrix and RPs can be calculated as follows:

$$\begin{aligned} R_{i,j} &= \Theta(\varepsilon - \|T(i) - T(j)\|) \\ &= \begin{cases} 1 : \varepsilon > \|T(i) - T(j)\|, \\ 0 : \varepsilon < \|T(i) - T(j)\|, \end{cases} \quad i, j \in [1, n - (m-1)\tau], \end{aligned} \quad (3)$$

where $\|\cdot\|$ denotes the L2 norm, $\Theta(\cdot)$ represents the Heaviside function, ε is the recurrence threshold parameter, τ is the time delay, and $R_{i,j}$ is a square matrix with $[n - (m-1)\tau]$ length. This formula means that if the L2 distance between $T(i)$ and $T(j)$ is less than ε , then $R_{i,j} = 0$. Otherwise, $R_{i,j} = 1$ and a black dot is located at the (i, j) in a two-dimensional space. After all $T(i)$ and $T(j)$ are processed through this function, the drawn graph is a RP. As a visualization tool, RP can graphically describe the dynamic characteristics in a qualitative way and reveal the latent time-correlated signatures. 8 RPs of 8 samples each with one fault are selected as examples, which are shown in Figure 4.

In RQA, three major parameters, respectively, embedding dimension m , time delay τ , and threshold ε , need to be determined. m is set to 4 based on the false nearest neighbours (FNNs), and τ is set to 5 based on the mutual information [25]. ε is empirically set to 0.8.

It is highly impracticable to directly use RPs to classify faulty types for their low resolution. Thus, RQA appears as a good tool to quantify RPs with recurrence statistics. The core of RQA is to identify and quantify the transient recurrent patterns which characterize the dynamic change behaviors. In this paper, 10 recurrence parameters named $R_1 - R_{10}$ are derived from the distribution of points and vertical lines of the RPs.

The recurrence rate (RR) is the simplest parameter of cross-correlation sum. It is defined as

$$RR = \frac{1}{n^2} \sum_{i,j=1}^n R_{i,j}, \quad (4)$$

where n is the number of RPs points and $R_{i,j}$ is the binary matrix. RR represents the density and trajectory of PRs points and physically indicates the probability of recurring a specific state.

Determinism (DET) is a criterion of the predictability of a system. It is given by

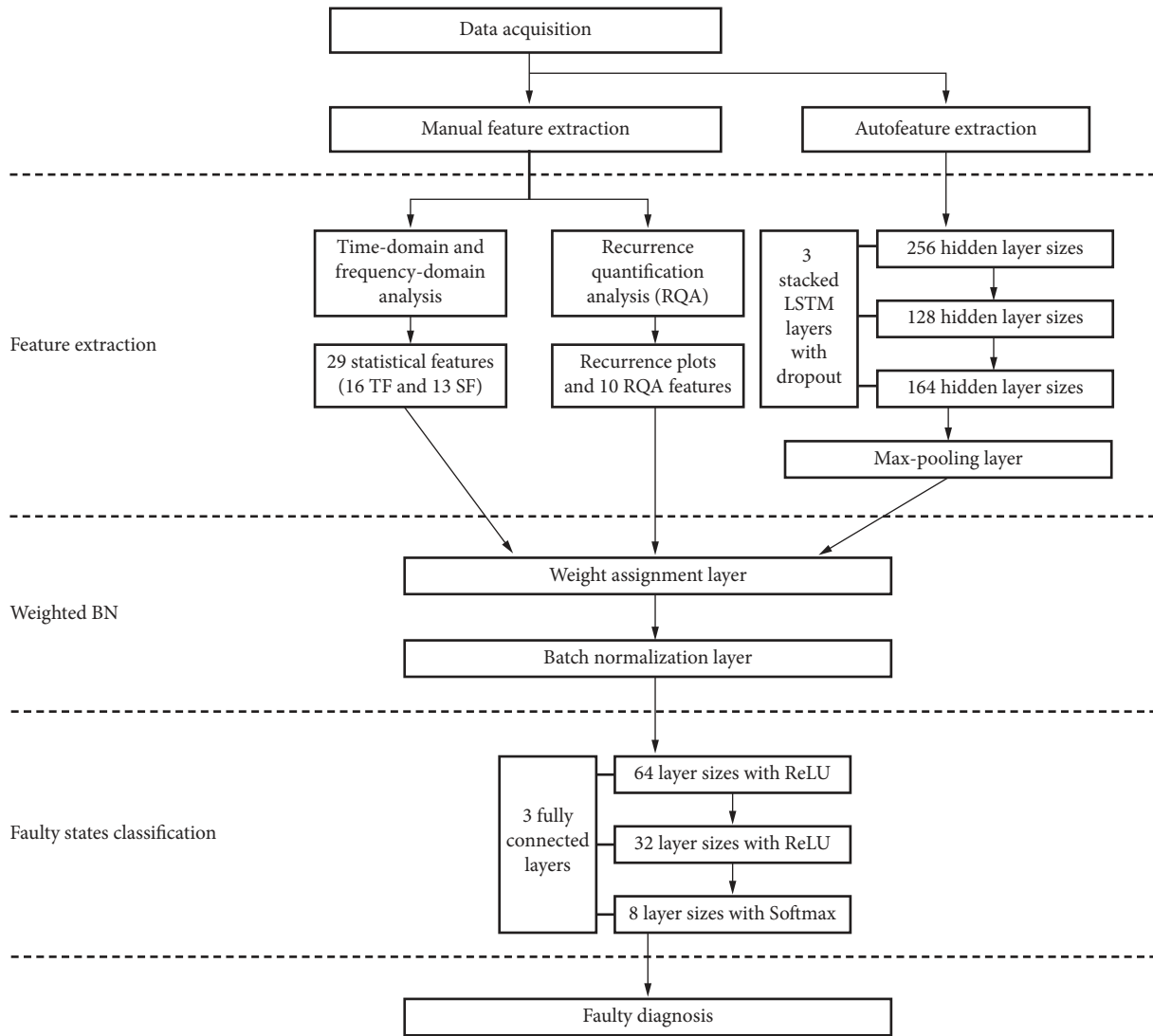


FIGURE 3: The flow diagram of the proposed method.

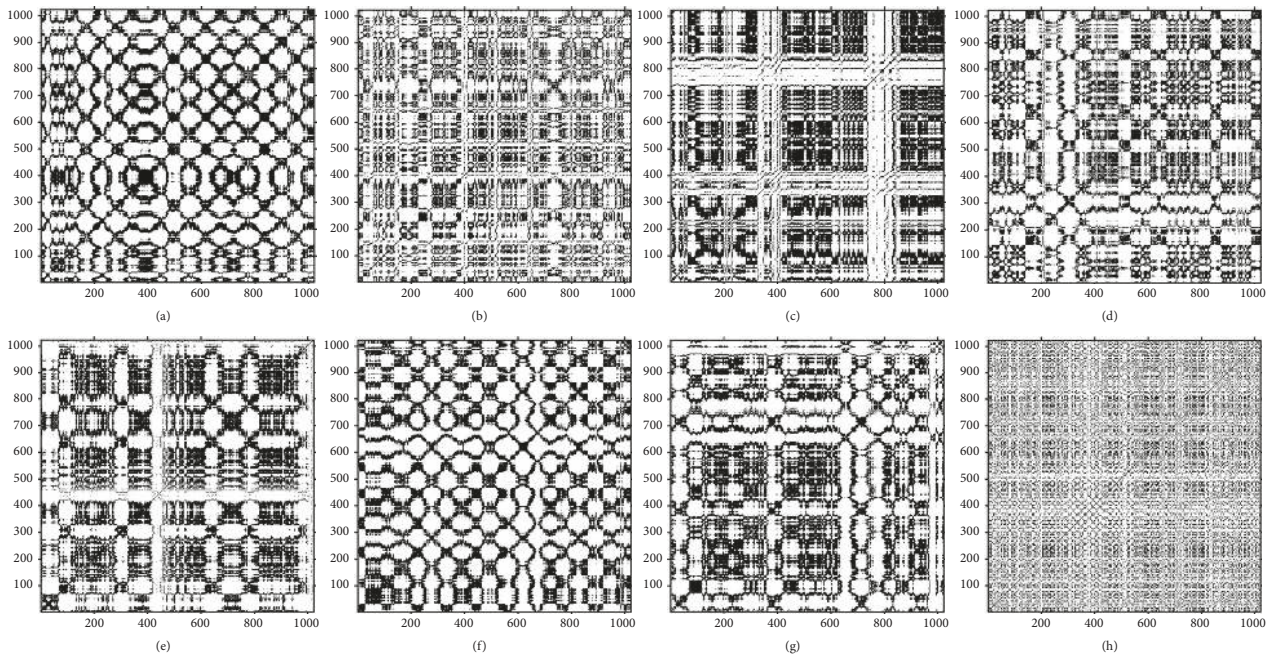


FIGURE 4: 8 RPs of 8 faulty motors. (a) NO. (b) BB. (c) BR. (d) RM. (e) SW. (f) VP. (g) RU. (h) FB.

$$\text{DET} = \frac{\sum_{l=l_{\min}}^n P(l)}{\sum_{i=1}^n \sum_{j=1}^n R_{i,j}}, \quad (5)$$

where l_{\min} is the minimum value of the diagonal line length and $P(l)$ denotes the distribution of the diagonal line length. DET is used to distinguish the organized RPs points from the dispersed ones. A high value of DET reflects a stable system while a low value indicates a stochastic system.

Shannon entropy (LENTR) denotes the complexity of a system. It is calculated as

$$\text{LENTR} = - \sum_{l=l_{\min}}^n P(l) \ln P(l), \quad (6)$$

where l_{\min} is the minimum value of the diagonal line length and $P(l)$ denotes the distribution of diagonal line length. LENTR indicates the complexity of RPs and can be used to estimate how much information is needed to recover the RPs.

Laminarity (LAM) corresponds with the number of laminar phases of a system. It is defined as

$$\text{LAM} = \frac{\sum_{v=v_{\min}}^n vP(v)}{\sum_{v=1}^n P(v)}, \quad (7)$$

where v_{\min} is the minimum value of the vertical line length and $P(v)$ denotes the distribution of the vertical line length. LAM is related to the intermittency of RPs points.

Trapping time (TT) represents the average vertical line length. It is given by

$$\text{TT} = \frac{\sum_{v=v_{\min}}^n vP(v)}{\sum_{v=v_{\min}}^n P(v)}. \quad (8)$$

TT estimates the average time in which a state of RPs is trapped.

Average diagonal length (ADL) is the average diagonal line length. It is the mean length of diagonal parallel lines. Similarity, longest diagonal length (LDL) is the longest diagonal parallel line length, longest vertical length (LVL) is the longest vertical line length, and average vertical length (AVL) is the average vertical line length. Besides these 9 parameters from the distribution of RPs, recurrence time is chosen as the 10th parameter, which evaluates the complexity of RPs though the calculation time.

3.2. LSTM. Recurrent neural network (RNN) is one of the deepest neural networks, which can address the data series with arbitrary length and has been applied successfully in many end-to-end tasks [26]. For the sake of avoiding the gradient vanishing or exploding problem, a modified RNN architecture named long short-term memory (LSTM) which involves a memory cell is created [27]. A concept of forget gates is introduced in LSTM to deal with the long-term dependency problem. These forget gates can screen the useful information of the historical network cells to capture more meaningful and distinct information in data series.

At time step t , the hidden state h_t of the LSTM cell is updated with the following terms: the input data x_t , the input gate i_t , the output gate o_t , the forget gate f_t , the memory cell c_t , and the hidden state h_{t-1} of the previous LSTM cell at time

step $t-1$. The schematic diagram of the architecture of a LSTM cell is shown in Figure 5. Those parameter-updating formulas are expressed as follows:

$$\begin{aligned} i_t &= \sigma(w_i x_t + v_i h_{t-1} + b_i), \\ f_t &= \sigma(w_f x_t + v_f h_{t-1} + b_f), \\ o_t &= \sigma(w_o x_t + v_o h_{t-1} + b_o), \\ c_t &= f_t \odot c_{t-1} + i_t \odot \tanh(w_c x_t + v_c h_{t-1} + b_c), \\ h_t &= o_t \odot \tanh(c_t), \end{aligned} \quad (9)$$

where σ is the activation function of sigmoid and \odot represents the element-wise product. W , V , and b are, respectively, a $d \times k$ matrix, a $d \times d$ matrix, and a d -dimensional vector, where d is the input series length and k is the hidden layer size.

In order to establish a better LSTM architecture with higher diagnostic accuracy, several parameters, such as layer number, time steps, and learning rate, need to be determined. The quantification of these parameters is mainly based on a comparative evaluation of the performances of various optional values in dataset 1. The accuracy and cost time at different layer numbers are shown in Figure 6, with the layer size set to 64. It can be observed that as the layer number increases from 1 to 5, the accuracy curve has a distinct rise at first 3 layers and then tends to be stable, while the cost time always increases as the layers get deeper. By comprehensive consideration of both accuracy and efficiency, the optimal layer number is selected as 3.

During the construction of the stacked LSTM, the backpropagation is used for the update of weights, of which the learning rate is the main parameter. Different learning rates and corresponding results of 10 times of repeat tests are visualized in the boxplot in Figure 7(a). It shows that the test accuracy is relatively low as the learning rate is too small or too big. Thus, 0.001 is selected as the learning rate. Then, the selection of time steps is investigated. Figure 7(b) shows that it attains the highest accuracy as the number of time steps is 4 and the performances vary little when the number of time steps is 2 and 4. It means that a sample with 1024 data points is segmented into 4 parts and converted to a tensor $S \in R^{4 \times 256}$.

The hidden layer size of each stacked layer is empirically, respectively, 256, 128, and 64 in a descending order, and thus, the output size of each layer is 4×256 , 4×128 , and 4×64 .

The max-pooling function is used to eliminate the dimensionality curse and retain the most useful information of a region by returning the maximum value. In this method, a max-pooling layer is added after the stacked LSTM layers to convert the 4×64 output to a flatten 64-dimensional vector. To prevent model form overfitting, dropout based on early stopping mechanism is adopted in the training process [28]. Probability of dropping out neurons in the convolution layers is set to 0.25.

3.3. Statistical Features. Twenty-nine statistical features, including 16 time-domain features and 13 frequency-domain features, are extracted [29]. It is shown in Table 2, where f_m is the frequency of the m th spectrum component; M is the total

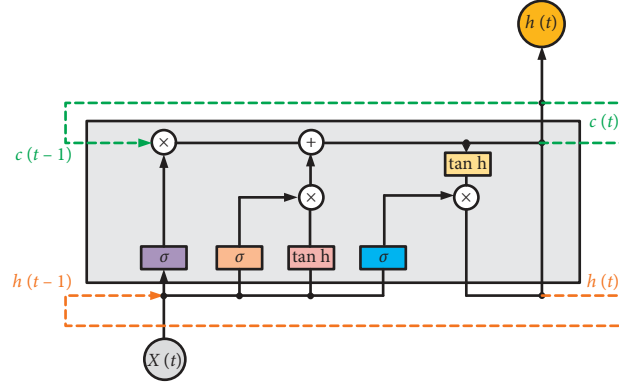


FIGURE 5: The schematic diagram of LSTM architecture.

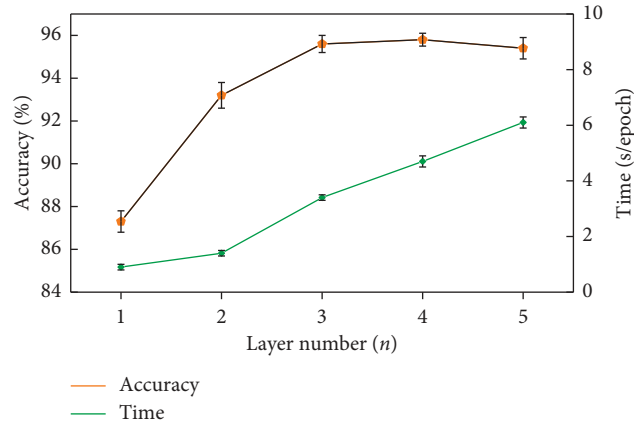


FIGURE 6: Accuracy and cost time at different layer numbers. The accuracy curve has a distinct rise at the first 3 layers and then tends to be stable, while the cost time always increases as the layers get deeper.

number of spectrum components; and $F(m)$ is a spectrum function at $m = 1, 2, \dots, M$. 16 popular empirical statistical time-domain features TF_1 – TF_{16} , including maximum, minimum, variance, skewness, kurtosis, and root mean square, are extracted. Features TF_1 – TF_{10} describe the vibration amplitude and energy distribution. For example, RMS represents the magnitude of varying values, and kurtosis denotes the spikiness degree of time series. 6 dimensionless parameters TF_{11} – TF_{16} reflect the distribution of time sequence. In frequency domain, feature FF_1 represents the average vibration energy; features FF_2 – FF_5 and FF_{10} – FF_{13} indicate the convergence of the spectrum power; and features FF_6 – FF_9 reflect the magnitude of position shift of the dominant frequencies.

3.4. Weighted BN. BN is a popular type of normalization method which can transform the distribution of any neuron's input during a batch of iterations to Gaussian distribution in deep NNs [22]. BN has displayed its impressive impact on reducing overfitting and avoiding gradient dispersion. The weighted BN is a modification of the BN layer, which consists of a weighting layer and a BN layer. It is designed to assign dynamic weights to the hybrid features and then perform batch normalization. It targets not only on reducing overfitting and eliminating the phenomenon of

internal covariate shift but also introducing an assessment mechanism of feature importance and facilitating evaluation of the contributions of the three feature learning approaches.

The weighting layer can be regarded as a customized neural network layer before BN layer and optimized during the whole training process. The role of the weighting layer is to assign dynamic weights to the hybrid features. For the RQA feature set $\{R_1, R_2, \dots, R_{10}\}$ and statistical feature set $\{TF_1, \dots, TF_{16}, FF_1, \dots, FF_{13}\}$, a one-by-one corresponding weighting vector $W = \{w_1, w_2, \dots, w_{39}\}$ is employed, which is expressed in equation (10). For the 64-dimensional LSTM output, due to the facts that no general consensus has been reached about the specific meaning of each dimension of LSTM output and there already exists a large amount of intrinsic weights and biases in LSTM, we assign a unique weight w_{40} to the 64-dimensional vector, which is expressed in equation (11).

$$F_i^l = w_i^l F_i^{l-1}, \quad i = (1, 2, \dots, 39), \quad (10)$$

$$L_j^l = w_{40}^l L_j^{l-1}, \quad j = (1, 2, \dots, 64), \quad (11)$$

where F_i is the i th term in the combined feature set $\{R_1, R_2, \dots, R_{10}, TF_1, \dots, TF_{16}, FF_1, \dots, FF_{13}\}$, L_j is the j th number of the LSTM output, and l denotes the l th iteration. Note that the update of weights in the weighting layer is a

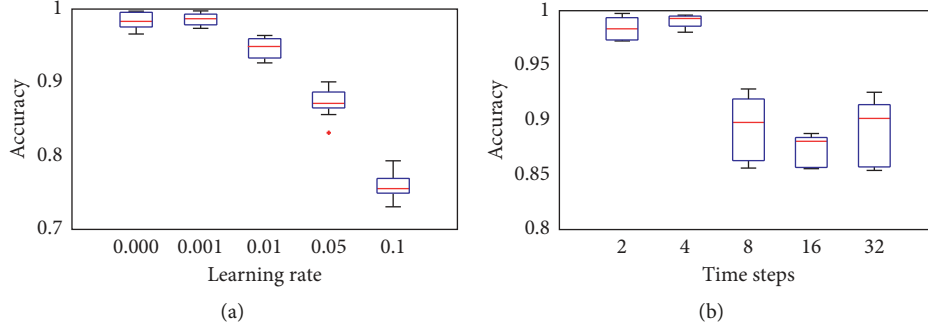


FIGURE 7: Diagnostic accuracy at different optional values. (a) Learning rate. (b) Time steps.

TABLE 2: Statistical features.

Domain	Features	Expression
Time	TF ₁	$(\sum_{n=1}^N t(n))/N$
	TF ₂	$\sqrt{\sum_{n=1}^N t^2(n)/N}$
	TF ₃	$(\sum_{n=1}^N \sqrt{ t(n) }/N)^2$
	TF ₄	$(\sum_{n=1}^N t(n))/N$
	TF ₅	$(\sum_{n=1}^N (t(n) - TF_1)^3)/N$
	TF ₆	$(\sum_{n=1}^N (t(n) - TF_1)^4)/N$
	TF ₇	$\max(t(n))$
	TF ₈	$\min(t(n))$
	TF ₉	$(\sum_{n=1}^N (t(n) - TF_1)^2)/(N - 1)$
	TF ₁₀	$TF_7 - TF_8$
	TF ₁₁	TF_2/TF_4
	TF ₁₂	TF_7/TF_2
	TF ₁₃	TF_7/TF_3
	TF ₁₄	TF_7/TF_4
	TF ₁₅	$TF_5/(TF_2)^3$
	TF ₁₆	$TF_6/(TF_2)^4$
Frequency	FF ₁	$(\sum_{m=1}^M F(m))/M$
	FF ₂	$(\sum_{m=1}^M (F(m) - FF_1)^2)/(M - 1)$
	FF ₃	$(\sum_{m=1}^M (F(m) - FF_1)^3)/(M(\sqrt{FF_2})^3)$
	FF ₄	$(\sum_{m=1}^M (F(m) - FF_1)^4)/(M(FF_2)^3)$
	FF ₅	$\sqrt{(\sum_{m=1}^M (f_m - FF_1)^2 F(m))/M}$
	FF ₆	$(\sum_{m=1}^M f_m F(m))/(\sum_{m=1}^M F(m))$
	FF ₇	$\sqrt{(\sum_{m=1}^M f_m^2 F(m))/(\sum_{m=1}^M F(m))}$
	FF ₈	$\sqrt{(\sum_{m=1}^M f_m^4 F(m))/(\sum_{m=1}^M f_m^2 F(m))}$
	FF ₉	$(\sum_{m=1}^M f_m^2 F(m))/(\sqrt{\sum_{m=1}^M F(m) \sum_{m=1}^M f_m^2 F(m)})$
	FF ₁₀	$(\sum_{m=1}^M (F(m) - FF_5)^3 F(m))/(m(FF_6)^3)$
	FF ₁₁	$(\sum_{m=1}^M (F(m) - FF_5)^3 F(m))/(m(FF_6)^4)$
	FF ₁₂	$(\sum_{m=1}^M (F(m) - FF_5)^{1/2} F(m))/(m\sqrt{FF_6})$
	FF ₁₃	FF_5/FF_6

part of the optimization in the training process based on the batch gradient descent algorithm.

3.5. Fully Connected Neural Networks. Three fully connected layers with the size of {64, 32, 8} is added. In these layers, the neurons at different layers are all connected to each other;

activation functions of ReLU are used at the first two layers, and SoftMax is used at the last layer.

The whole architecture is constructed through minimizing the following cross-entropy loss function between the predicted values and real labels [30]:

$$L = -\frac{1}{n} \sum_x \hat{y} \ln y + (1 - \hat{y}) \ln (1 - y), \quad (12)$$

where y is the model output and \hat{y} is the real-label value.

The batch gradient descent and backpropagation algorithm are used for optimization and to minimize the cost function L . Thus, the weights and biases can be updated through the following equations:

$$w = w - \lambda \frac{\partial L}{\partial w},$$

$$b = b - \lambda \frac{\partial L}{\partial b}, \quad (13)$$

where w denotes the weight, b represents the bias, and λ is the learning rate.

In addition, the programs are performed on a GeForce GTX TITAN X graphics card and a E5-2630 processor with 126 GB memory, using TensorFlow as backend.

4. Results and Discussion

In this section, the performance of the proposed method is illustrated and discussed. Several other methods are tested and contrasted.

4.1. The Diagnostic Accuracy of the Proposed Method.

Figure 8 displays the training process during 100 epochs with the training and test accuracy and losses from the original dataset 1 and the noised dataset 2. It can be seen that a significant convergence trend occurs in both datasets. Besides, there are less fluctuations in the accuracy and loss curves of dataset 1 compared with the curves of dataset 2. The distinction can be explained by the fact that the difficulty level of fault diagnosis for dataset 2 with 5 dB noise is obviously higher than that of dataset 1. Figure 9 illustrates the visualized outputs of every step using the established model with 8 test samples of different fault types.

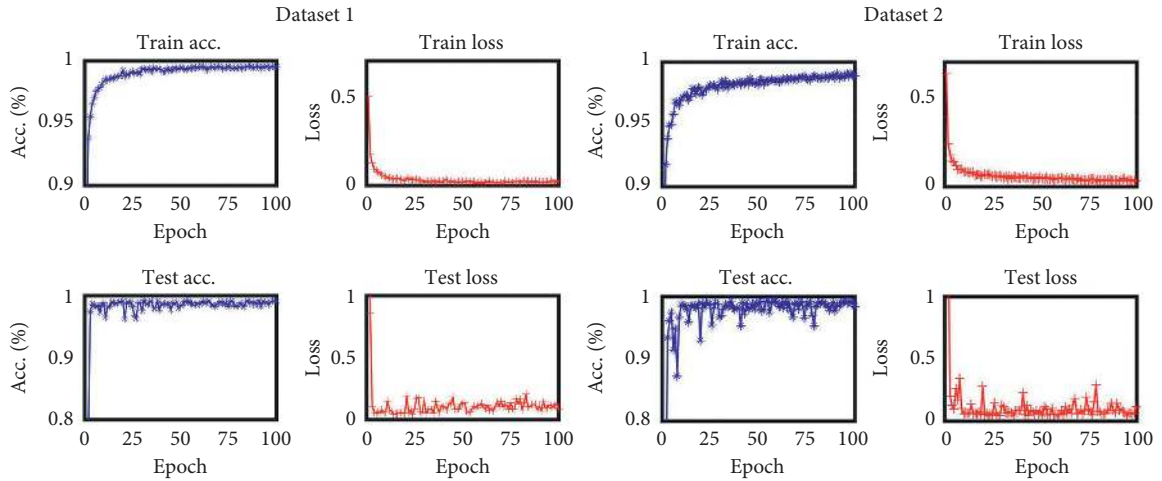


FIGURE 8: The training process of dataset 1 and dataset 2. A significant convergence trend occurs in both datasets with bigger fluctuations in dataset 2.

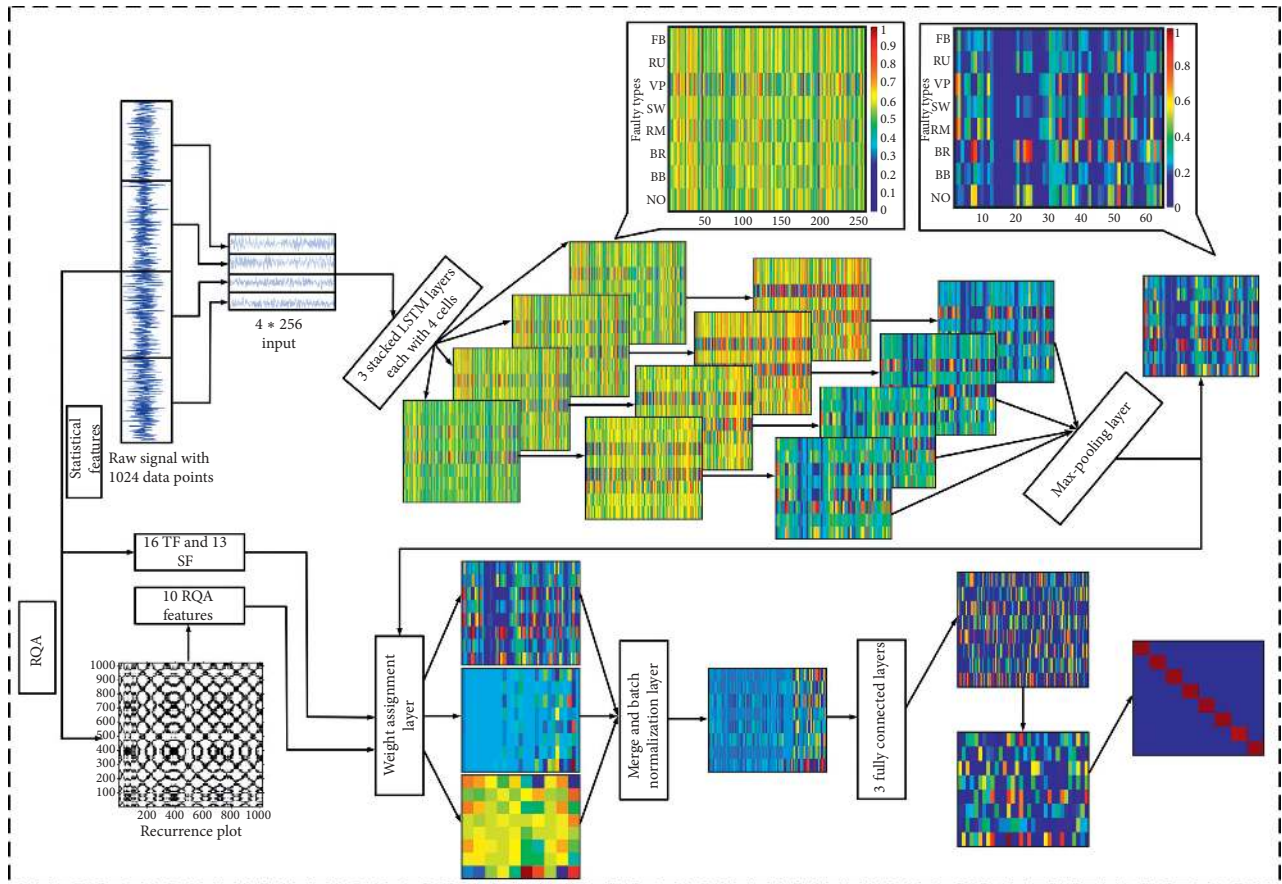


FIGURE 9: Visualization of the outputs of every step of 8 different samples.

Figure 10 shows the diagnostic accuracy of each faulty type using confusion matrix. The diagnostic accuracy is calculated as the rate of the correctly classified test sample number to the total test sample number. It can be seen that average classification accuracy reaches 99.3% in dataset 1 and 98.9% in dataset 2.

Figure 11 displays the proportions of the RQA features' weights to the sum of all weights. These proportions are calculated by

$$P = \frac{w_i}{\sum_{j=1}^{40} w_j}, \quad (14)$$

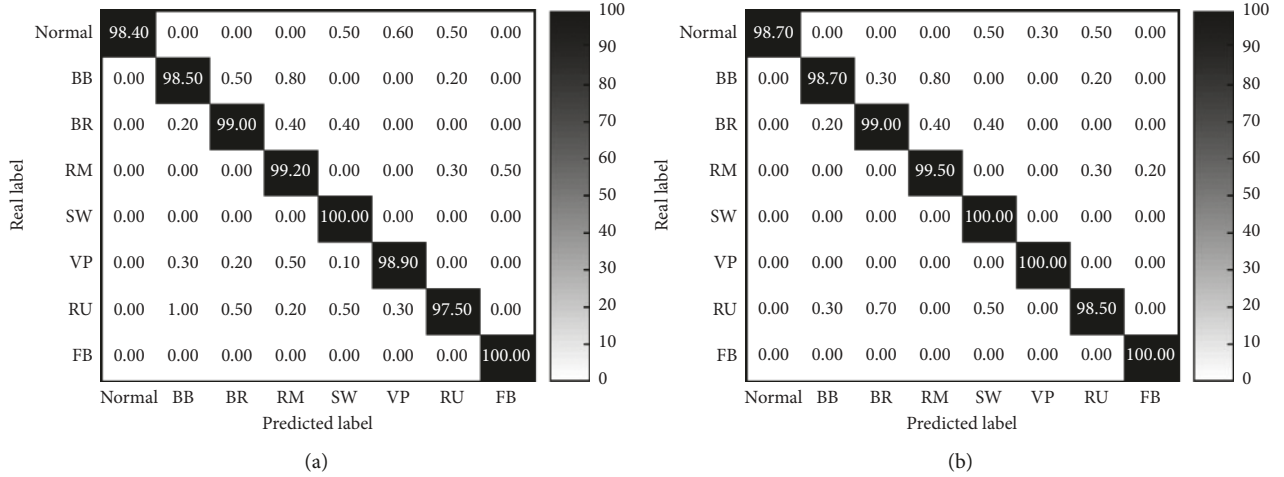


FIGURE 10: Confusion matrix of the classification results of two datasets. (a) Dataset 1. (b) Dataset 2.

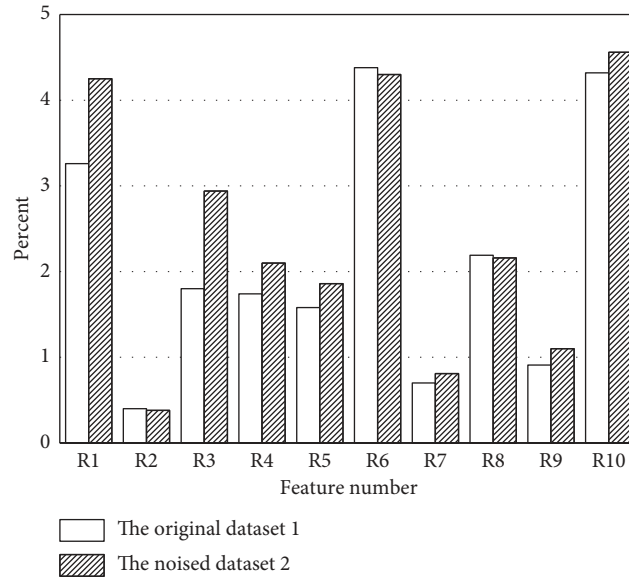


FIGURE 11: The proportions of RQA features.

where w_i is the weight of R_i . It shows that the proportions of RQA features' weights from dataset 2 are generally higher than dataset 1. It proves that RQA plays a bigger role in dealing with the noised data and indicates its potentials in real-world applications which require strong antinoise capability.

4.2. Visualization at Different Layers. In order to illustrate how the proposed method executes the classification task step-by-step, the different layers' outputs are visualized with 3-dimensional sketches using a nonlinear dimension reduction method, t-distributed stochastic neighbour embedding (t-SNE) [31]. t-SNE is employed to convert the high-dimensional outputs at different layers to 3D representations, which are shown in Figures 11-12. It can be observed that the 8 groups of point clouds are heavily

overlapped at the first few layers, and as the layer goes deeper, they become more separable. Although the dimension reduction by t-SNE involves inevitable errors for the loss of information, it demonstrates the effectiveness of the proposal in datasets 1 and 2 in a qualitative manner.

4.3. Performance Comparisons. To verify the superiority of the proposed method, several state-of-the-art intelligent fault diagnosis methods are employed for comparison:

- (1) RQA + SVM [19]: independent use of RQA for feature learning and optimal binary tree SVM for classification
- (2) LSTM: independent use of the proposed LSTM architecture on raw data

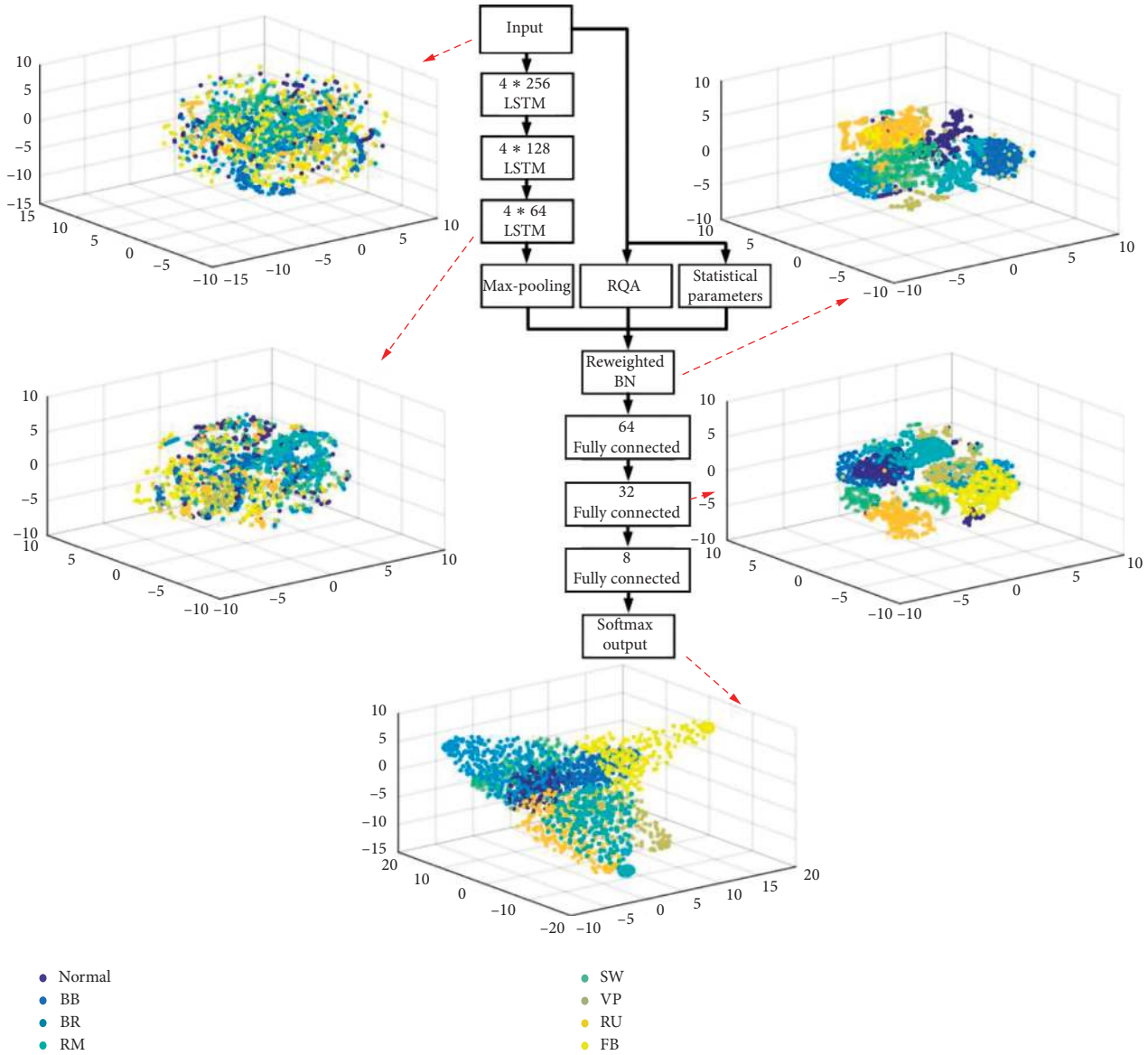


FIGURE 12: 3D representations of high-dimensional outputs at different layers with dataset 1 by t-SNE.

- (3) MLP [7]: multilayer neural network on statistical features
- (4) CNN [32]: one-dimension CNN on raw data
- (5) SIFT + CNN [13]: short-time Fourier transform for feature learning and CNN for classification
- (6) CDFL [14]: convolutional discriminative learning of a BP network

In (1) and (2), the separate use of RQA and LSTM aims at proving its indispensability in the hybrid method. In (1), 10 RQA features mentioned in Section 3 are extracted and 7 optimal binary tree SVMs with RBF kernels are employed as the binary classifiers. In (2), the same three-layer stacked LSTM architecture is adopted. In (3), since MLP is incapable of addressing sequential raw data, the 29 statistical features mentioned in Table 2 are taken as the input. The hidden layers with layer size {64-128-64} are adopted. In (4), a

1D-CNN architecture, comprised of 3 convolutional and pooling layers, one flatten layer, and 3 fully connected layers, is used (Figure 13).

The kernel length of the convolutional layers is set to 64, and the max-pooling function is utilized in the pooling layers. In (5), 1D data are converted to a 2D time-frequency maps through short-time Fourier transform. Then, the maps are compressed into 100 * 100 squares and fed into a deep CNN. In (6), discriminative learning based on BPNN is incorporated into the unsupervised CNN. The training of a BPNN takes the place of the training of a CNN, with the input size equal to filter number 32 and the hidden layer size equal to filter size 64. The classification results of all methods on datasets 1 and 2 are shown in Figure 14.

Several conclusions can be drawn from Figure 14. (1) The proposed hybrid method outperforms the other methods with diagnostic accuracies of 99.3% and 98.9% in the two datasets.

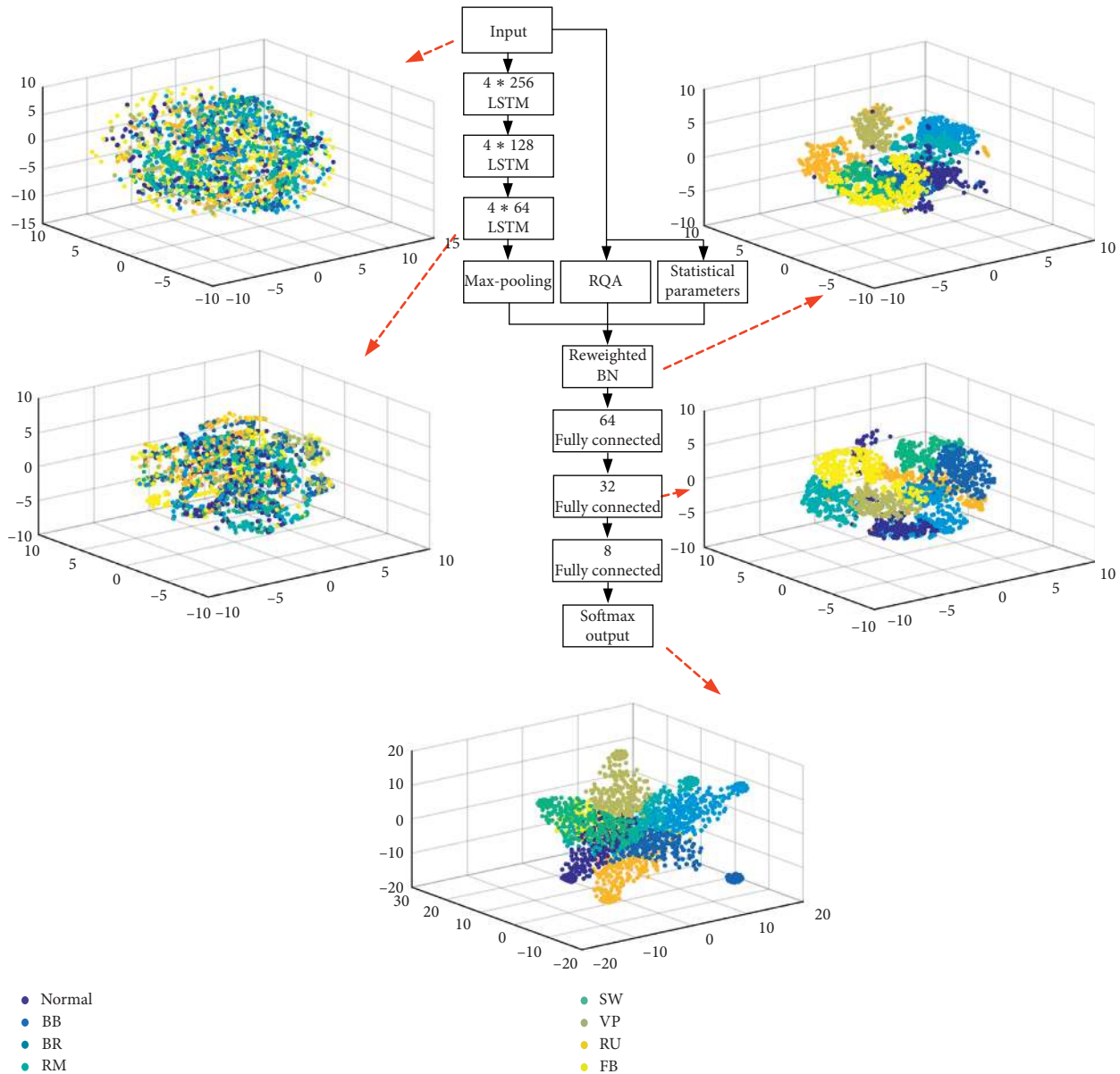


FIGURE 13: 3D representations of high-dimensional outputs at different layers with dataset 2 by t-SNE.

Besides, the difference value 0.04% is the smallest, which indicates its stronger noise resistance. (2) The separate use of RQA and LSTM obtains lower diagnostic accuracy compared with the proposal. It indicates that both RQA and LSTM neural network are indispensable in the hybrid method. In addition, RQA+SVM performs worst, and it can be explained by the limitation of the inadequate number of features and the dependence of manual feature learning. (3) Generally, the deep learning techniques, including CNN, SIFT+CNN, LSTM, and CDFL, perform better than the traditional methods. It can be explained by the fact that the deep learning methods, with stronger feature learning and representation capacity, always present a superior performance to the methods that require manual feature extraction [33]. (4) The major shortcoming of the proposal is the computational cost. The average cost time of 100 epochs is

530 s, only faster than SIFT+CNN's 690 s and slower than most of the comparative methods. It can be interpreted by the hybrid complex architecture which is embedded with three different feature learning approaches. A part of the computational space is occupied by drawing the RPs and extracting various statistical features. Besides, the LSTM neural network contains the deepest layers among those methods.

5. Conclusions

In this paper, a novel fault diagnosis framework with high accuracy for three-phase induction motors is presented. A hybrid feature learning approach that combines empirical statistical parameters, RQA and LSTM, is proposed to integrate the state-of-the-art deep learning techniques with the manual feature learning approaches to gain a superior and

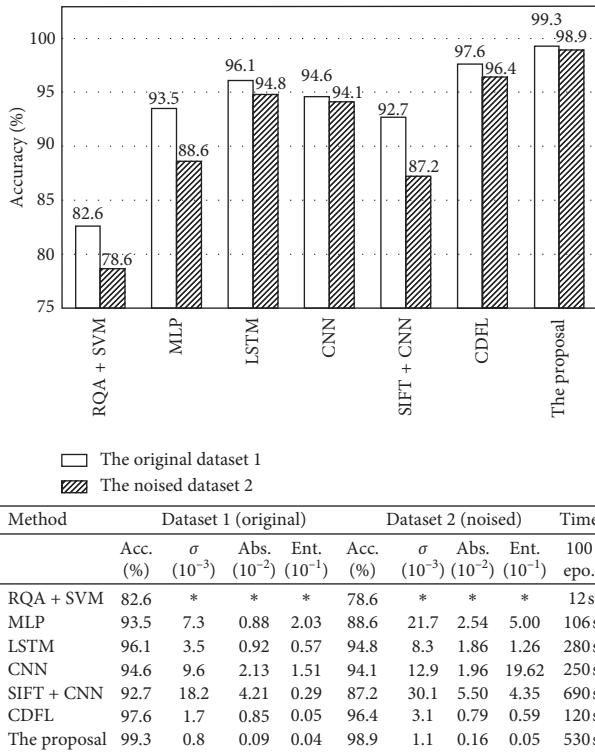


FIGURE 14: The performances of all methods. *Values are non-existent because of the use of multiple classifiers in RQA + SVM.

robust performance. In addition, a modification of BN named weighted BN is designed to evaluate the contributions of each feature learning approach and facilitate validating the noise resistance performance.

The tests of 8 motors (1 healthy and 7 faulty) are carried out to form datasets 1 and 2. It verifies that the proposed method, with the highest accuracies of 99.3% and 98.9% in fault recognition, performs better than other methods and possesses good noise resistance. More specifically, it yields 18.5% and 8.1% average performance improvements compared with RQA + SVM and MLP; it yields 3.6%, 4.7%, 9.2%, and 2.1% average performance improvements compared with the four deep learning methods, LSTM, CNN, SIFT + CNN, and CDFL. The weight distribution of the weighted BN illustrates RQA is more effective in dealing with the real-world noised data.

In future research, our effort will be devoted to two aspects. (1) Make the proposed method reliable for practical use, which demands a large amount of accumulated industry data. (2) Make attempts to figure out the connections between the intermediate representations of deep learning networks and traditional manual features.

Data Availability

The .tdms data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This project was supported by the National Key Technology R&D Program of China (No. 2017YFB1302004) and National Nature Science Foundation of China project (No. 51305258).

References

- [1] E. M. H. Benbouzid, "A review of induction motors signature analysis as a medium for faults detection," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 5, pp. 984–993, 2000.
- [2] S. Nandi, H. A. Toliyat, and X. Li, "Condition monitoring and fault diagnosis of electrical motors—a review," *IEEE Transactions on Energy Conversion*, vol. 20, no. 4, pp. 719–729, 2005.
- [3] H. Su and K. T. Chong, "Induction machine condition monitoring using neural network modeling," *IEEE Transactions on Industrial Electronics*, vol. 54, no. 1, pp. 241–249, 2007.
- [4] R. A. Patel and B. R. Bhalja, "Condition monitoring and fault diagnosis of induction motor using support vector machine," *Electric Power Components and Systems*, vol. 44, no. 6, pp. 683–692, 2016.
- [5] A. M. D. Younus and B.-S. Yang, "Intelligent fault diagnosis of rotating machinery using infrared thermal image," *Expert Systems with Applications*, vol. 39, no. 2, pp. 2082–2091, 2012.
- [6] 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.
- [7] I. Martin-Diaz, D. Morinigo-Sotelo, O. Duque-Perez, and R. J. Romero-Troncoso, "An experimental comparative evaluation of machine learning techniques for motor fault diagnosis under various operating conditions," *IEEE Transactions on Industry Applications*, vol. 54, no. 3, pp. 2215–2224, 2018.
- [8] W. Peng, Z. Jianjing, Y. Ruqiang et al., "Induction motor fault diagnosis and classification through sparse representation," in *Proceedings of ASME 2017 Dynamic Systems and Control Conference*, pp. 1–7, Tysons Corner, VA, October 2018.
- [9] S. Pan, T. Han, A. C. C. Tan, and T. R. Lin, "Fault diagnosis system of induction motors based on multiscale entropy and support vector machine with mutual information algorithm," *Shock and Vibration*, vol. 2016, Article ID 5836717, 12 pages, 2016.
- [10] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 115, no. 1, pp. 213–237, 2019.
- [11] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, and X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, vol. 89, no. 1, pp. 171–178, 2016.
- [12] C. Sun, M. Ma, Z. Zhao, and X. Chen, "Sparse deep stacking network for fault diagnosis of motor," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3261–3270, 2018.
- [13] L. H. Wang, X. P. Zhao, J. X. Wu et al., "Motor fault diagnosis based on short-time Fourier transform and convolutional neural network," *Chinese Journal of Mechanical Engineering*, vol. 30, no. 6, pp. 1–12, 2017.
- [14] W. Sun, R. Zhao, R. Yan, S. Shao, and X. Chen, "Convolutional discriminative feature learning for induction motor fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1350–1359, 2017.

- [15] R. Zhao, D. Wang, R. Yan et al., "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1539–1548, 2018.
- [16] P. Malhotra, T. V. Vishnu, A. Ramakrishnan et al., "Multi-Sensor Prognostics using an Unsupervised Health Index based on LSTM Encoder-Decoder," 1st SIGKDD Workshop on Machine Learning for Prognostics and Health Management, 2016, <http://arxiv.org/abs/1608.06154>.
- [17] R. Zhao, R. Yan, J. Wang et al., "Learning to monitor machine health with convolutional Bi-directional LSTM networks," *Sensors*, vol. 17, no. 2, pp. 1–18, 2017.
- [18] 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, no. 3, pp. 98–109, 2017.
- [19] X. Liu and L. Bo, "Identification of resonance states of rotor-bearing system using RQA and optimal binary tree SVM," *Neurocomputing*, vol. 152, no. 1, pp. 36–44, 2015.
- [20] Y. Qian, R. Yan, and S. Hu, "Bearing degradation evaluation using recurrence quantification analysis and Kalman filter," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 11, pp. 2599–2610, 2014.
- [21] C. Chao, S. Fei, and R. Yan, "Enhanced least squares support vector machine based transfer learning strategy for bearing fault diagnosis," *Chinese Journal of Scientific Instrument*, vol. 38, no. 1, pp. 33–40, 2017.
- [22] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," pp. 448–456, 2015, <http://arxiv.org/abs/1502.03167>.
- [23] M. Zhao, M. Kang, B. Tang, and M. Pecht, "Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4290–4300, 2018.
- [24] F. Takens, "Detecting strange attractors in turbulence," *Lecture Notes in Mathematics*, vol. 898, no. 1, pp. 366–381, Springer, Berlin, Heidelberg, 1981.
- [25] Z. Ma, G. Wen, and C. Jiang, "EEMD independent extraction for mixing features of rotating machinery reconstructed in phase space," *Sensors*, vol. 15, no. 4, pp. 8550–8569, 2015.
- [26] J. Schmidhuber, "Deep Learning in neural networks: an overview," *Neural Networks*, vol. 61, no. 1, pp. 85–117, 2014.
- [27] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A critical review of recurrent neural networks for sequence learning," *Computer Science*, 2015, <http://arxiv.org/abs/1506.00019>.
- [28] N. Srivastava, G. Hinton, A. Krizhevsky et al., "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [29] Z. Shen, X. Chen, X. Zhang, and Z. He, "A novel intelligent gear fault diagnosis model based on EMD and multi-class TSVM," *Measurement*, vol. 45, no. 1, pp. 30–40, 2012.
- [30] P. Zhou and J. Austin, "Learning criteria for training neural network classifiers," *Neural Computing and Applications*, vol. 7, no. 4, pp. 334–342, 1998.
- [31] L. J. P. Maaten and G. E. Hinton, "Visualizing high-dimensional data using t-SNE," *Journal of machine learning research*, vol. 9, no. 1, pp. 2579–2605, 2008.
- [32] L. Guo, Y. Lei, N. Li, T. Yan, and N. Li, "Machinery health indicator construction based on convolutional neural networks considering trend burr," *Neurocomputing*, vol. 292, no. 1, pp. 142–150, 2018.
- [33] T. Han, C. Liu, W. Yang et al., "Deep transfer network with joint distribution adaptation: a new intelligent fault diagnosis framework for industry application," 2018, <http://arxiv.org/abs/1804.07265>.

