

Research Article

Translation Invariance-Based Deep Learning for Rotating Machinery Diagnosis

Wenliao Du ¹, Shuangyuan Wang ², Xiaoyun Gong,¹ Hongchao Wang,¹ Xingyan Yao,³ and Michael Pecht ⁴

¹Henan Provincial Key Laboratory of Intelligent Manufacturing of Mechanical Equipment, Zhengzhou University of Light Industry, 5 Dongfeng Road, Zhengzhou 450002, China

²School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

³School of Computer Science and Information Engineering, Chongqing Technology and Business University, Chongqing 400067, China

⁴Center for Advanced Life Cycle Engineering (CALCE), University of Maryland, College Park, MD 20742, USA

Correspondence should be addressed to Wenliao Du; dwenliao@zzuli.edu.cn, Shuangyuan Wang; wswywy86@163.com, and Michael Pecht; pecht@calce.umd.edu

Received 11 February 2020; Accepted 15 June 2020; Published 11 August 2020

Academic Editor: Jiawei Xiang

Copyright © 2020 Wenliao Du 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.

Discriminative feature extraction is a challenge for data-driven fault diagnosis. Although deep learning algorithms can automatically learn a good set of features without manual intervention, the lack of domain knowledge greatly limits the performance improvement, especially for nonstationary and nonlinear signals. This paper develops a multiscale information fusion-based stacked sparse autoencoder fault diagnosis method. The autoencoder takes advantage of the multiscale normalized frequency spectrum information obtained by dual-tree complex wavelet transform as input. Accordingly, the multiscale normalized features guarantee the translational invariance for signal characteristics, and the stacked sparse autoencoder benefits the unsupervised feature learning and ensures accurate and stable diagnosis performance. The developed method is performed on motor bearing vibration signals and worm gearbox vibration signals, respectively. The results confirm that the developed method can accommodate changing working conditions, be free of manual feature extraction, and perform better than the existing intelligent diagnosis methods.

1. Introduction

Rotating machinery plays an important role in modern industries and has become more automatic, precise, and efficient [1]. On the one hand, higher requirements for the quality and performance of products means the machinery must be reliable and stable; on the other hand, the severe operating environments always lead to unplanned downtime, and failures can incur economic loss and endanger human safety. Therefore, an accurate and robust fault diagnosis tool for rotating machinery needs to be developed [2].

In general, fault diagnosis methods can be classified into either model-based methods or data-driven methods [3]. Model-based approaches need precise physical models of the

system, which is a challenging task in most cases due to the system structure complexity [4], whereas the data-driven methods always combine artificial intelligence with signal processing method, and these methods specifically identify different faults by a series of steps, including data collection, feature extraction, and classifier training [3]. Data-driven methods can be used in complex systems and do not need to build an accurate mechanical failure physical model. Thus, they have become a promising tool in the field of mechanical condition monitoring [5].

In the traditional intelligent diagnosis method, the quality of the extracted features directly affects the classifier training performance [6]. Since mechanical systems often work in complex and variable environments, including load

changes and unstable speeds, the collected vibration signals usually exhibit typical nonlinear and nonstationary characteristics. Traditionally, statistical features including mean, variance, kurtosis, root mean square, and so forth are collected in the time domain as input to the classifier. However, if the distributions of the derived features are not separable enough for different conditions, it is hard to get high diagnostic accuracy. In fact, due to the complex structure and transmission path, and variable working conditions, the distributions of features are easily overlapped. Although some researchers use their domain knowledge to append a features selection step to find a reliable set of statistical parameters for fault diagnosis [3], there is still no guarantee that the remaining features can fully represent the dynamic characteristics under such complex operating conditions. Furthermore, if the features themselves do not adequately express the mechanical failure characteristics, the performance improvement is limited.

Recently, some scholars have proposed to directly use the collected vibration signal as the input of the deep learning model with the help of the classifier's powerful learning ability to automatically learn the fault characteristics and obtained a better classification effect. Lei et al. [1] used sparse filtering combined with softmax regression to diagnose bearing faults. Shao et al. [5] proposed an integrated self-encoding neural network to diagnose bearing faults. A sparse self-encoding neural network was employed to diagnose the motor rotor faults [7]. Furthermore, Jiang et al. [8] substituted the Fourier spectral features from the original vibration signal as the input of the denoising autoencoders to identify gearbox breakage, pitting, peeling, and other faults. In [6], wavelet time-frequency spectrum on the acquired gearbox vibration signal was combined with a residual neural network to diagnose gearbox broken teeth, pitting, missing teeth, and other faults.

However, mechanical equipment often works under complicated conditions, and the collected signals are nonstationary and nonlinear [9]. When the collected signals are segmented to train the classification model, the nonstationary and nonlinear characteristics of the signal often limit the learning ability of the deep learning network. While the Fourier transform is just a powerful tool for stationary signal analysis [10], and although the traditional wavelet transform benefits from its adaptive and multiresolution capability, it is also difficult to guarantee the time-invariant characteristics of the signal [11]. The dual-tree complex wavelet transform first proposed by Kingsbury [12] was verified to enjoy super shift invariance and reducing spectral aliasing to traditional wavelet transform. Luo et al. utilized the dual-tree complex wavelet transform to extract features from the vibration signals and strain signals to monitor the damage to an automotive suspension component [13]. A dual-tree complex wavelet packet transform based Bayesian belief method was proposed to diagnose gearbox and locomotive roller bearing faults [14]. Kumar et al. [15] used the dual-tree complex wavelet transform to decompose the load current to extract the fundamental component of distorted load current and developed a control algorithm for power quality improvement in a distribution system.

In this paper, a multiscale information fusion-based stacked sparse autoencoder (dual-tree complex wavelet transform based stacked sparse autoencoder, DCWT-SSAE) was developed to further improve rotating machinery diagnostic performance. Likewise, the developed DCWT-SSAE employs dual-tree complex wavelet transform and fast Fourier transform (FFT) to avoid shift-variance and spectral aliasing caused by the nonstationary and nonlinear character, and the stacked sparse autoencoder benefits the unsupervised feature learning and ensures accurate and stable diagnosis performance.

The rest of the paper is organized as follows. The theoretical background is offered in Section 2, and Section 3 elucidates the developed DCWT-SSAE fault diagnosis method. In Section 4, the developed DCWT-SSAE is applied to inductor motor bearing for the sake of finding discriminative features, and its effectiveness is verified by comparing with other state-of-the-art intelligent fault diagnosis methods. In Section 5, the DCWT-SSAE is further applied to worm gearbox faults diagnosis, and the effectiveness is also analyzed. Section 6 presents the conclusions.

2. Theoretical Background

2.1. Dual-Tree Complex Wavelet Transform. The dual-tree complex wavelet transform (DTCWT) has nearly shift-invariant properties, is free from frequency aliasing, and has perfect reconstruction and good directional selectivity [12]. The dual-tree complex wavelet transform applies a real wavelet transform with two different low-pass filters and high-pass filters to decompose and reconstruct the signal, and this pair of wavelets is called real-tree and imaginary-tree, respectively. Each filter satisfies the perfect reconstruction conditions, and together they form a Hilbert transform pair (90° out of phase with each other). Because the filters in the real-tree wavelet have a half sample delay compared to those in the imaginary-tree wavelet, the sampling points of the real-tree wavelet always locate in the middle of the imaginary-tree wavelet in the decomposition and reconstruction, which ensures the information complementarity between the two trees and realizes the approximate shift invariance. In the decomposition process in each scale, for the real-tree and the imaginary-tree wavelets, because the two independent wavelet transforms are implemented independently with the pyramid algorithm, the DTCWT can be achieved using the existing discrete wavelet transform (DWT) algorithm, and the computational cost is dramatically decreased (only 2 times that of the basic DWT).

Let $\psi_h(t)$ and $\psi_g(t)$ denote the real-valued wavelet in the dual-tree transform, respectively, and $\phi_h(t)$ and $\phi_g(t)$ are the corresponding scaling functions. These two real wavelets constitute a complex analytical wavelet, which is only supported on the positive frequency. The dual-tree wavelet transform is implemented by two independently parallel wavelet transforms. Based on the wavelet theory, the wavelet coefficients $d_l^{\text{re}}(k)$ and the scaling coefficients $c_j^{\text{re}}(k)$ of the

real-valued wavelet transform can be calculated through the following formula:

$$d_l^{\Re e}(k) = 2^{l/2} \int_{-\infty}^{+\infty} x(t) \psi_h(2^l t - k) dt, \quad l = 1, \dots, J, \quad (1)$$

$$c_J^{\Re e}(k) = 2^{J/2} \int_{-\infty}^{+\infty} x(t) \varphi_h(2^J t - k) dt. \quad (2)$$

Similarly, the wavelet coefficients $d_l^{\Im m}(k)$ and the scaling coefficients $c_J^{\Im m}(k)$ of the imaginary-tree can be calculated. Combining the output of the two trees, the wavelet

coefficients and the scaling coefficients of the dual-tree complex wavelet transform can be obtained as follows:

$$d_l^C(k) = d_l^{\Re e}(k) + j d_l^{\Im m}(k), \quad l = 1, \dots, J, \quad (3)$$

$$c_J^C(k) = c_J^{\Re e}(k) + j c_J^{\Im m}(k). \quad (4)$$

Furthermore, using the wavelet coefficients and the scaling coefficients, the detail components at all levels and the approximation component at the last level can be individually reconstructed using the following equations:

$$d_l(t) = 2^{(l-1)/2} \left[\sum_{n=-\infty}^{\infty} d_l^{\Re e}(k) \psi_h(2^l t - n) + \sum_{m=-\infty}^{\infty} d_l^{\Im m}(k) \psi_g(2^l t - m) \right], \quad l = 1, \dots, J, \quad (5)$$

$$c_J(t) = 2^{(J-1)/2} \left[\sum_{n=-\infty}^{\infty} c_J^{\Re e}(k) \varphi_h(2^J t - n) + \sum_{m=-\infty}^{\infty} c_J^{\Im m}(k) \varphi_g(2^J t - m) \right]. \quad (6)$$

The reconstructed signal $\hat{x}(t)$ is obtained by summing all the detail components and the approximation component as follows:

$$\hat{x}(t) = \sum_{l=1}^J d_l(t) + c_J(t). \quad (7)$$

2.2. Stacked Sparse Autoencoders. An autoencoder can be regarded as an unsupervised neural network [16]. The network consists of three layers, including an input layer, a hidden layer, and an output layer. The input layer has the same number of nodes as the output layer. The input layer and the hidden layer constitute the encoder, and the hidden layer and the output layer constitute the decoder. Generally, the training principle of the network is similar to that of a BP neural network, including forward calculation and back-propagation of errors. In the training process, the data in the previous layer is reconstructed, and the hidden layer can be regarded as the abstract of the previous layer. In fault diagnosis, the output of the hidden layer is the extracted features. For input samples $X = \{x_i^p\}$, where $p = 1, 2, \dots, N$, N is the number of samples, and for each sample $x_i^p \in R^M$, M is the length of the sample. The output of the hidden layer can be expressed as follows:

$$a_j^p = \sum_{i=1}^M \omega_{ij} x_i^p + b_1, \quad (8)$$

$$H_j^p = f(a_j^p) = \frac{1}{1 + e^{(-a_j^p)}},$$

where f is the activation function, which is the sigmoid function, and ω_{ij} is the weight between the i th node in the input layer and the j th node in the hidden layer, and b_1 is the bias vector. Similarly, the output of the last layer can be expressed by the softmax function as follows:

$$u_k^p = \sum_{j=1}^N \omega_{jk} H_j^p + b_2, \quad (9)$$

$$\hat{x}_k^p = g(u_k^p) = \frac{u_k^p}{\sum_{k=1}^M (e^{u_k^p})},$$

where $g(\cdot)$ is the softmax function, ω_{jk} is the weight between the j th node in the hidden layer and the k th node in the output layer, and b_2 is the corresponding bias vector. The input layer to the hidden layer is regarded as a coding process, and the hidden layer to the output layer can be a decoding process. The network training process uses the gradient descent method to adjust the weight iteratively and makes $\hat{x} \approx x$. For a network including m number of samples, the cost function can be defined by

$$J(\omega, b) = \frac{1}{m} \sum_{i=1}^m \frac{1}{2} \|x^i - \hat{x}^i\|^2 + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (\omega_{ji}^{(l)})^2, \quad (10)$$

where the first item on the right side is the error between the input value and the real output, the second item is the regularization term, and λ is the coefficient for the L_2 regularization term. The weight delay parameter balances these two terms. After training the autoencoder, the output of the hidden layer is the abstract expression of the original samples, which are the extracted features. In the traditional training process, the nodes in the hidden layer must fire for all samples. Inspired by the learning process of biological neurons, many neurons have low activations and even do not activate for some excitation. In order to obtain a sparser expression of the input layer, it is better to keep the neurons of the hidden layer "inactive." An improved autoencoder named sparse autoencoder (SAE) was realized by imposing sparsity constraint [17]. Therefore, the trained neural network obtains a more compressed representation of the input data, which is more effective in reducing information redundancy and improving the accuracy of data expression.

The cost function of the sparse autoencoder can be defined by

$$J_{\text{sparse}}(\omega, b) = J(\omega, b) + \beta \sum_{j=1}^{S_2} \text{KL}(\rho \parallel \hat{\rho}_j), \quad (11)$$

where β is the coefficient for the sparsity regularization term and $\text{KL}(\cdot)$ is Kullback–Leibler divergence function, $\text{KL}(\rho \parallel \hat{\rho}_j) = \rho \log(\rho/\hat{\rho}_j) + (1 - \rho) \log(1 - \rho/1 - \hat{\rho}_j)$, which measures how different the two distributions of $\hat{\rho}_j$ and ρ are, $\hat{\rho}_j$ is the average activation value of the neuron j in the hidden layer, and ρ is a desired small value of this neuron; the ratio $(\rho/\hat{\rho}_j)$ presents the sparsity proportion, a smaller ratio corresponding to a higher sparsity. Adding this sparsity proportion term to the cost function that constrains the values of $\hat{\rho}_j$ to be low encourages each neuron in the hidden layer to fire to a small number of training examples. If the sparsity parameter $\rho = \hat{\rho}_j$, the penalty term $\text{KL}(\rho \parallel \hat{\rho}_j) = 0$. Otherwise, the penalty increases monotonically, so $\text{KL}(\rho \parallel \hat{\rho}_j)$ acts as the sparsity constraint.

3. Developed DCWT-SSAE Method

According to the fact that the mechanical vibration signal has typical nonstationary and nonlinear characteristics, the developed DCWT-SSAE method integrates the multiscale analysis ability of the dual-tree complex wavelet transform and the powerful feature learning ability of the sparse self-encoder. As shown in Figure 1, the dual-tree complex wavelet decomposition decomposes the signal into multiple time-frequency planes and achieves a more intensive representation of the signal, while ensuring the translation invariance of the signal. Through a Fourier transform on these multiscale components, they are used as input to the sparse autoencoder, and the encoding decoding process is used to learn the sparse representation of the original input. Multiple sparse autoencoders are stacked, and a softmax network is used as a classifier to identify different fault types.

The training process of the DCWT-SSAE consists of the following five steps:

- (1) Decompose the training samples with double-tree complex wavelet decomposition, and obtain a multiscale translation invariant representation of the original signal. For each scale, the FFT is used to obtain the frequency spectrum and a normalized treatment is followed.
- (2) Set SAE model parameters, including learning rate, sparse rate, and other parameters. Use unsupervised learning with the training samples for the first SAE model, and obtain the weight between the neuron nodes of the input layer and the hidden layer, and the offset parameters of each layer.
- (3) Regard the hidden layer output of the first SAE as the representation layer, which is used as the input of the following SAE model, and train the following SAE; likewise, the layer-wise unsupervised learning of all the individual SAE models is completed in sequence.

- (4) Stack all representation layers to form a deep network. The softmax layer is connected on the top of the DCWT-SSAE. The network is further trained with the supervised training process to get a diagnosis model, and the parameters, including the weights and offset parameters, are fine-tuned using label information, which ensures more discriminative feature representations are obtained. That is, in the fine-tuning stage, for the training samples, the signal and the corresponding known label are used to further optimize the entire weights and biases with the error backpropagation principle.
- (5) Verify the trained model using test samples. If the diagnostic accuracy does not meet the requirements, the initial parameters of the SAE model, including the weight and bias, are initialized randomly, and the maximum epochs, learning rate, and sparsity parameters are set up with a given value, and then the model training is re-executed until the required diagnostic accuracy is achieved.

4. Fault Diagnosis of Motor Bearing

4.1. Dataset. The dataset of 6205-2RS JEM SKF deep-groove ball bearing with different faults was provided by Case Western Reserve University [18] and used to verify the effectiveness of the developed method. An accelerometer was mounted on the motor housing at the drive end of the motor, and single-point faults were introduced to the outer race, inner race, and ball. The bearing was tested with a rotating speed of 1800 rpm under four different loads (0, 1, 2, and 3 hp), four different fault locations (normal condition (N), ball fault (BF), inner race fault (IF), and outer race fault (OF)), and four different severity levels (normal, slight, medium, and serious levels corresponding to 0, 0.18 mm, 0.36 mm, and 0.53 mm fault diameter, respectively), and the vibration data were acquired with a sampling frequency of 12 kHz.

These acquired vibration signals comprise the bearing dataset, which is used to verify the performance of the developed method. The dataset contains 10 bearing health conditions corresponding to different fault severity levels and different fault locations under the 4 loads, where the same fault location but different loads is treated as one class. Each health condition under one load contains 100 samples, and each sample contains 1024 data points. Therefore, the dataset is constructed by 4,000 samples. All these samples are divided into the training and the testing samples randomly, in which 10% of samples are chosen for training and the remaining 90% for testing.

4.2. Diagnosis Results. In this paper, for the dual-tree complex wavelet decomposition, the (5, 7)-tap symmetry biorthogonal filters were used at the first level, and at the rest of the levels, the 14-tap linear phase filters produced by Q-shift solution [19] were used. The neural network in the developed DCWT-SSAE has five layers, in which the node number of the input layer is determined by the output

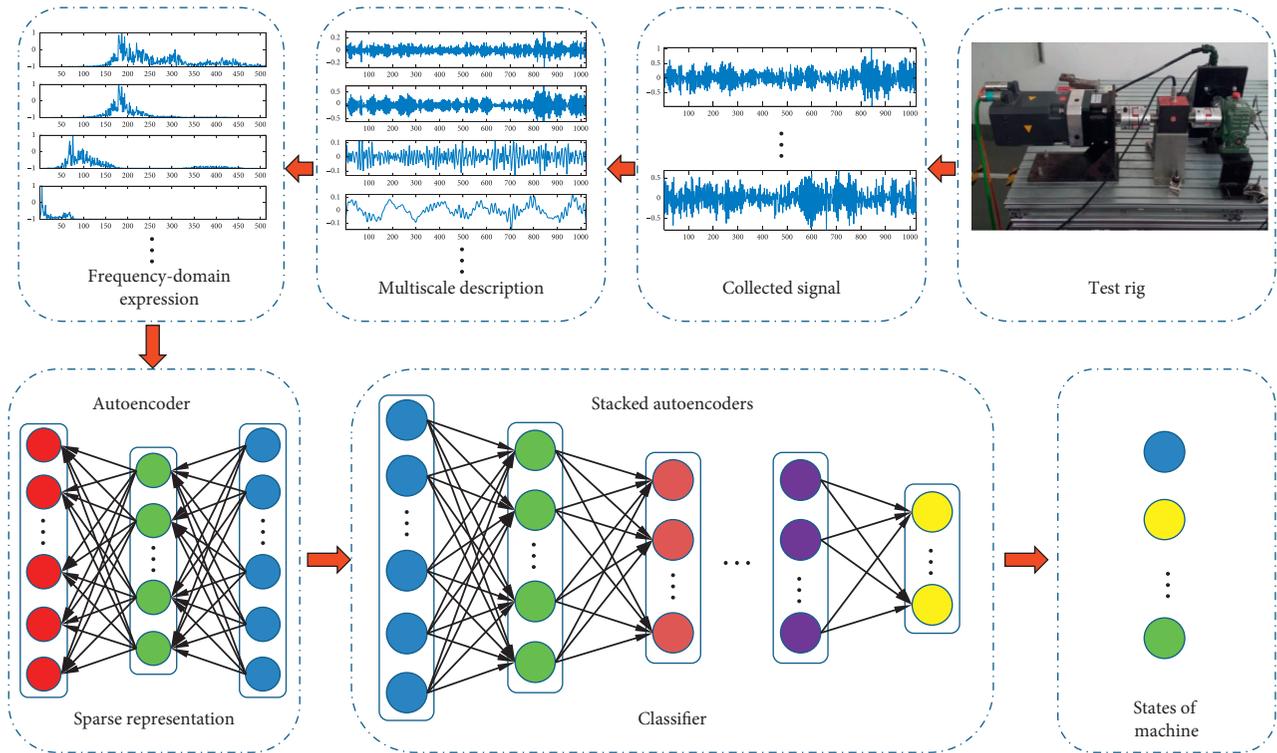


FIGURE 1: Flowchart of fault diagnosis based on the developed DCWT-SSAE approach.

dimension of the dual-tree analysis for the samples. The node number of the first to the third hidden layer is 400, 200, and 50, respectively, and the node number of the last layer is the same as the number of conditions. The coefficients for the L2 regularization term and the sparsity regularization term are 0.0016 and 5, respectively. The desired proportion of training examples a neuron reacts to is 0.5. The training accuracies and testing accuracies were averaged by 10 trials to reduce the effects of the randomness.

Ten trials were implemented for discriminating the mentioned 10 bearing health conditions. The diagnosis results of the developed method are shown in Figure 2(a). In these trials, for the training samples, the diagnosis accuracy of each trial was 100%; for the test samples, all of the diagnostic accuracies were over 99.5%, and the mean accuracy reached 99.71%, and the variance of the accuracy was $2.15e-6$, which means the developed method is effective to distinguish the 10 different conditions of bearings with a high accuracy.

In addition, the t-distributed stochastic neighbor embedding (t-SNE) [20], a nonlinear dimensionality reduction method, was used to provide 3D representations of the learned high-dimensional features at different layers in the DCWT-SSAE. As shown in Figures 3 and 4, the feature maps expressed in a lower-dimensional space involve unavoidable errors due to the loss of information in dimensionality reduction, but the effectiveness for fault diagnosis can be demonstrated qualitatively based on visualization of learned representation.

For comparison, we used the commonly used Db5 wavelet in the multiscale wavelet decomposition [21] to

substitute for the dual-tree complex wavelet in the DCWT-SSAE model for the above bearing dataset (this model is labeled as wavelet-SSAE). In order to keep the same SAE structure, for every scale, the wavelet coefficient was reconstructed. The accuracy of the 10 trials is shown in Figure 2(b) and gets an inferior mean accuracy 98.38%, and the corresponding variance is 0.0011. In addition, we compared the diagnosis results using these two approaches and the model of wavelet transform in tandem with stacked autoencoders (wavelet-SAE) [22] trained by different percentages of samples, as shown in Table 1. The testing accuracy increased for both approaches with the rise of percentage of training samples. At the initial stage, the performance obtained a remarkable improvement with a slightly increasing percentage of training samples. When the percentage reached 5%, both the wavelet-SSAE and the proposed DCWT-SSAE obtained over 96% testing accuracy, whereas the wavelet-SAE only got 84.68% recognition rates. Table 1 shows that the developed DCWT-SSAE method is superior to the wavelet-SSAE method and the wavelet-SAE method with the same percentages of samples, and the proposed method diagnoses the 10 conditions of the bearing dataset with 98.07% accuracy using only 5% of samples for training. The testing accuracy reached 99.71% when the percentage increased to 10%, and the accuracy was 100% when it increased to 40%. This result indicates that the dual-tree method can be trustworthy even when there is a small quantity of training samples.

Compared with the traditional intelligent fault diagnosis framework, such as BP neural networks, and support vector machines (SVM), the developed diagnosis method can

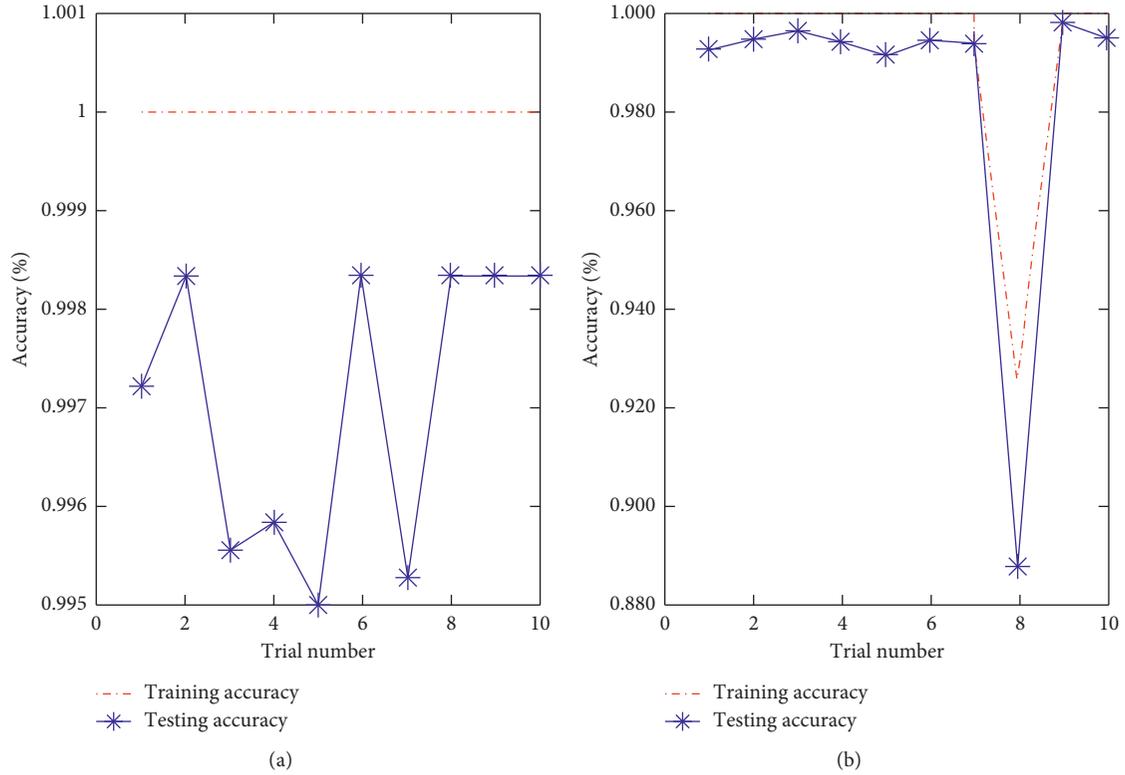


FIGURE 2: Diagnosis results of 10 trials of bearing datasets: (a) DCWT-SSAE; (b) wavelet-SSAE.

directly learn fault features. Both of BP neural network and stacked sparse autoencoder use the forward calculation and error backpropagation principle in training process. But, if the dimensions of the input are particularly large, model training is very difficult. So, the features extraction step is essential for the BP neural network. While, for the sparse autoencoder, the hidden layer can be regarded as the abstract of the input layer, even if the input data has a high dimensionality, the stacked autoencoder model can achieve automatic feature extraction without manual intervention. The developed method firstly decomposes the signal by time-invariant transform and then converts the obtained multi-scale components into the frequency domain as the input of the stacked sparse autoencoder. To verify the advantages of the developed method compared to the traditional intelligent methods, we compared it with the results in the related work using the same bearing dataset. As shown in Table 2, in our previous study [23], for the 10 health conditions of the motor bearings under free load, the classification accuracy of SVM is only 88.90% after a manual feature extraction and feature selection procedure, whereas, in [24], for the 10 health conditions of the motor under 3 hp, the time-domain features combined with wavelet energy features were used in the trace ratio linear discriminant analysis (TR-LDA), and after being trained with 10% of the samples, the testing accuracy achieved 92.5%.

In addition, we compared the developed DCWT-SSAE with the recently developed deep learning fault diagnosis model. The developed DCWT-SSAE is based on the characteristics of the mechanical vibration signal itself. In [5],

ensemble deep autoencoders (EDAEs) were constructed with 15 kinds of DAEs using different activation functions for 12 health conditions of the motor under four different loads (0, 1, 2, and 3 hp). The 2/3 raw vibration data was directly trained with the diagnosis model and obtained 97.18% testing accuracy. In [1], based on the sparse filtering learning features from raw vibration signals and the softmax regression determining the health conditions, the authors proposed a two-stage learning method (2S_LM) for mechanical diagnosis, for 10 health conditions of the motor under the mentioned four different loads; the testing accuracy reached 99.66% when 10% of samples were used to train the diagnosis model. In fact, in 2S-LM, sparse filtering and averaging process are used to extract discriminative features from raw vibration signals in the first stage; the softmax regression is employed to classify mechanical health conditions in the second learning stage. In order to get the ideal features, the original signal must be divided into segments alternately, and the averaging process is essential to eliminate the bad effects of the difference of each segment and random features caused by noise, while the developed DCWT-SSAE method uses dual-tree complex wavelet transforms to overcome the time-varying in the time domain, so the processes of dividing signal and averaging local features can be omitted, and the stacked sparse autoencoder benefits the unsupervised feature learning and ensures accurate and stable diagnosis performance.

From the comparison results, the developed method shows better results than the traditional manual feature extraction-based diagnosis methods. The main reason is that

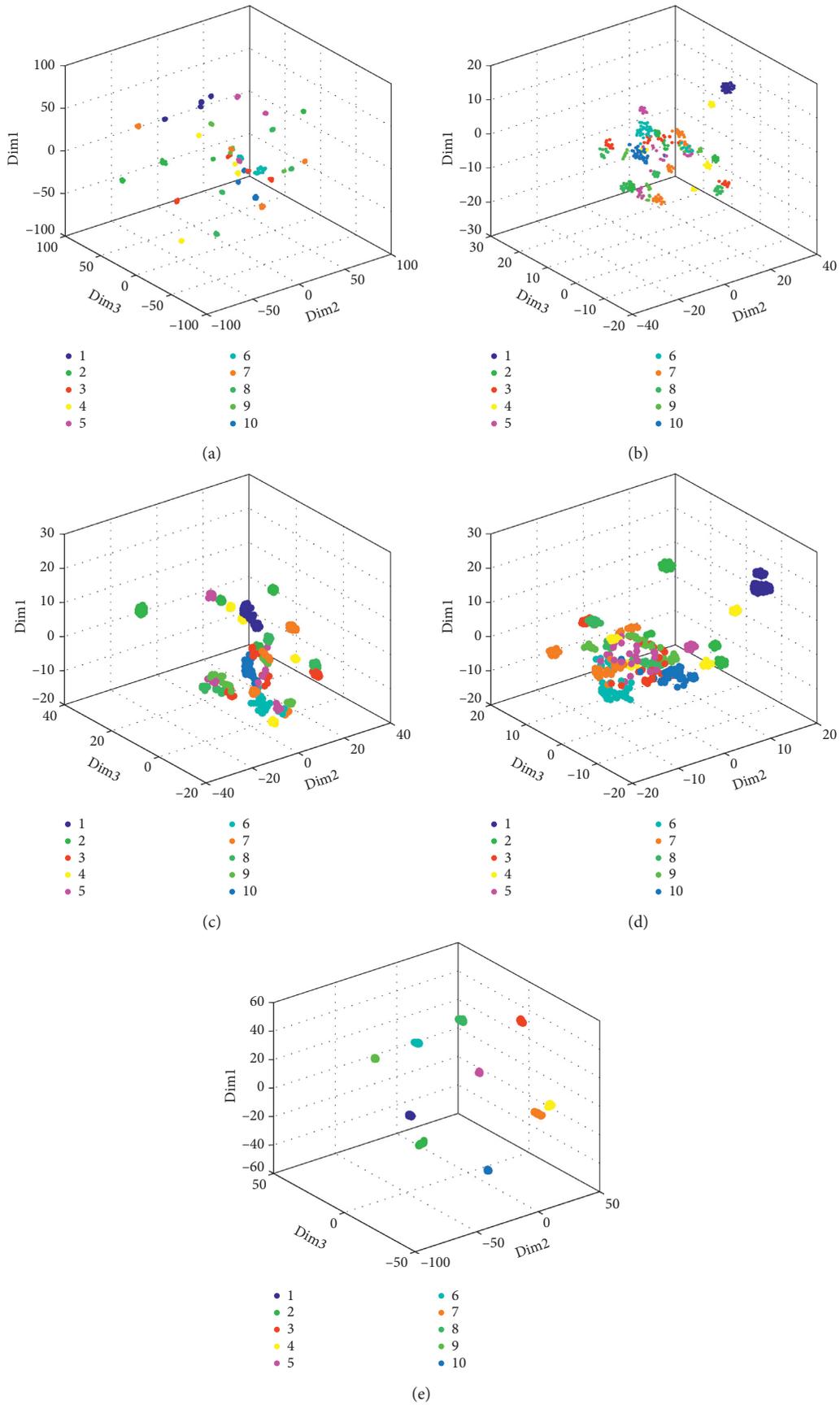


FIGURE 3: t-SNE results of output in different levels for bearing training samples: (a) input level; (b) first hidden level; (c) second hidden level; (d) third hidden level; (e) output level.

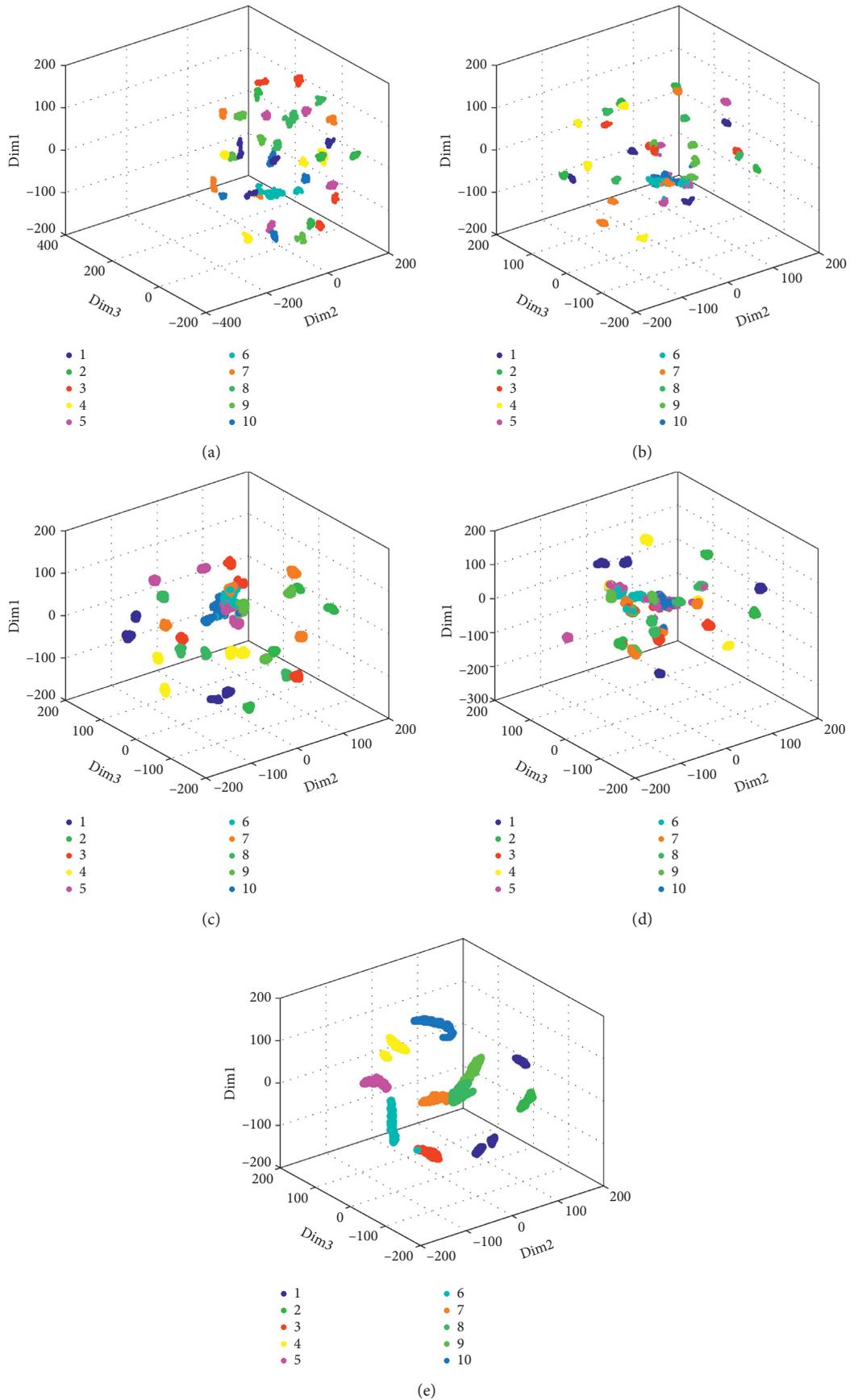


FIGURE 4: t-SNE results of output in different levels for bearing testing samples: (a) input level; (b) first hidden level; (c) second hidden level; (d) third hidden level; (e) output level.

TABLE 1: Performance comparison of the developed DCWT-SSAE method and wavelet-SSAE trained by different percentages of samples (unit: %).

	1%	2%	3%	5%	10%	25%	40%
Wavelet-SSAE	58.85	79.59	86.68	96.41	98.38	99.95	99.96
Wavelet-SAE	42.83	61.80	71.62	84.68	94.00	98.29	98.44
DCWT-SSAE	64.97	81.85	90.64	98.07	99.71	99.99	100

TABLE 2: Performance comparison of different methods with the motor bearing dataset (unit: %).

Methods	Ratio of training samples (%)	Load (hp)	Number of health conditions	Testing accuracy (%)
SVM [23]	75	0	10	88.90
TR-LDA [24]	10	3	10	92.50
EDAEs [5]	66.7	0	12	97.18
2S_LM [1]	10	0/1/2/3	10	99.66
Developed	10	0/1/2/3	10	99.71

the developed method can effectively learn the distinguished features from the input data, whereas the performance of the traditional methods relies heavily on the quality of the manual extracted features. Compared with directly using the time-domain vibration signal to train the autoencoder diagnosis model, the dual-tree complex wavelet transforms the raw vibration signals to the time-frequency domain and holds the invariant features in the frequency domain, which overcomes the time-varying in the time domain.

Furthermore, for the same vibration signal, compared with the general wavelet transform, the dual-tree wavelet transform is more appropriate to maintain the shift invariance of the signal and keep the impact characteristics of the signal. Figure 5 presents the outer race fault condition (fault diameter is 0.18 mm) as an example. The original signals consist of the original outer race fault signal and its delayed series (delay 2048 and 4096 points, respectively); the three levels of wavelet decomposition are implemented on the original signals; the wavelet functions are the aforementioned dual-tree wavelet and Db5 wavelet. The left panel is the decomposition results of wavelet and scaling components obtained by dual-tree wavelet, and the right panel corresponds to the results obtained by Db5 wavelet.

For dual-tree wavelet decomposition, either the wavelet components in each level or the scaling components hold a more stable translation invariance in the structure, and the delay is clearly shown in the wave form. However, when the time-domain features are extracted from these original signals or the obtained multiscale components, as these time-domain features are calculated with the statistics of the signal waveform, the difference in the external form certainly affects the extracted features. Figures 6 and 7 are the Fourier spectrum of the first three levels of wavelet components of the aforementioned signals with dual-tree wavelet and Db5 wavelet, respectively. In Figure 6, even if there is a time delay, the Fourier spectrum is not affected, and the spectra have coincident components, whereas in Figure 7, the spectra have a different structure for the original signal and its delay series, especially for level 2.

5. Fault Diagnosis for Worm Gearbox

5.1. Dataset. The established worm gearbox experimental setup is shown in Figure 8, and the worm gear dataset [25] was collected from a worm gearbox of WPA40. The gearbox with a 1:10 deceleration ratio had 2 threads and 20 teeth. The reference diameter of the worm gear was 30 mm, and the module, the lead angle, and the pressure angle of the worm gear were 2.5 mm, $9^{\circ}28'$, and 20° , respectively. The two current (AC) servomotors were employed as the driver and the loader (0 and 6 Nm), respectively. The artificial faults (worm gear pitting, worm gear spalling, and broken worm gear) were simulated on the worm gear, as shown in Figure 9. A triaxial acceleration sensor was mounted on the gearbox to measure the vibration at the 1000 rpm of driving rotational speed, and the vibration data were acquired with a sampling frequency of 12.8 kHz. In this study, only the X-axis direction (along the worm axial) vibration signals were used, and the raw vibration signals corresponding to different health conditions are shown in Figure 10.

The collected dataset contained 4 worm gearbox health conditions corresponding to different fault types under 2 loading conditions (free and 6 N m, respectively), where the same fault type but with different loads was treated as one class. Each condition under one load contained 500 samples, and each sample contained 1024 data points. Therefore, the dataset was constructed of 4,000 samples. All these samples were divided into the training and the testing samples randomly, in which 10% of the samples were chosen for training and the remaining 90% for testing.

5.2. Diagnosis Results. As mentioned before, the same dual-tree complex wavelet decomposition is used for the acquired worm gear dataset; that is, the (5, 7)-tap symmetry biorthogonal filters are used at the first level, and the 14-tap linear phase filters produced by Q-shift solution are used at the remaining levels. The neural network in the developed DCWT-SSAE has five layers, in which the node number of the input layer is determined by the output dimension of the dual-tree analysis for the samples, and the number of nodes

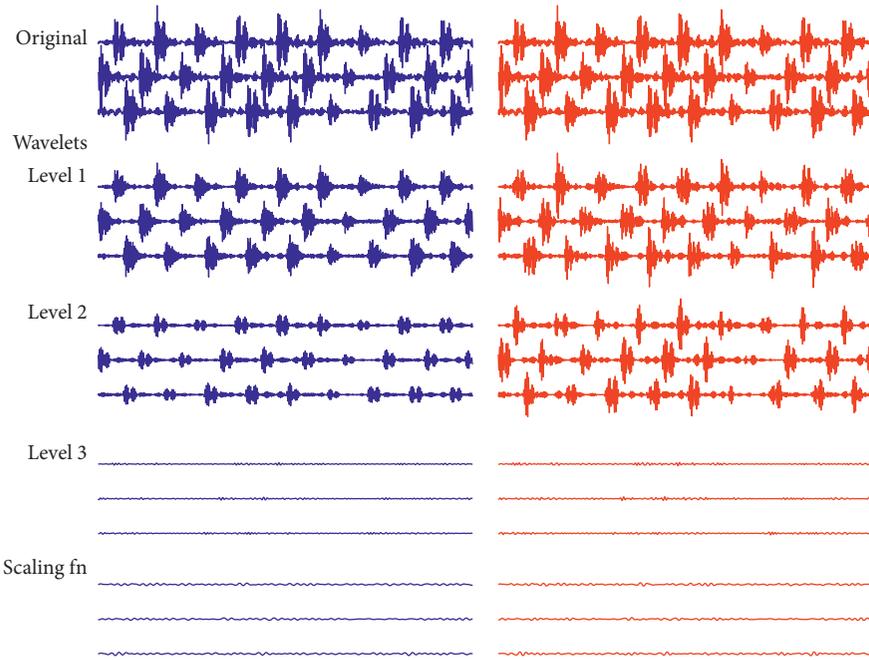


FIGURE 5: Decomposition results of the original signal and its delay series with dual-tree wavelet (left panel) and Db5 wavelet (right panel).

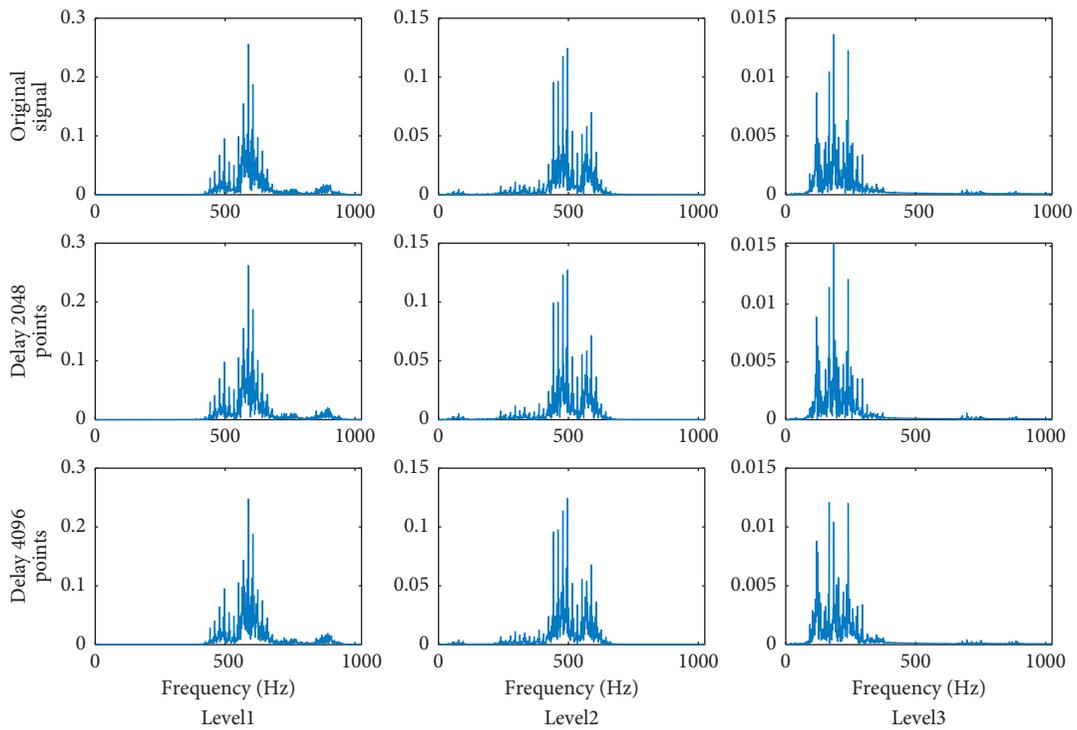


FIGURE 6: The first three levels of Fourier spectrum of the wavelet components of the original signal and its delay series obtained by dual-tree wavelet.

from the first layer to the third hidden layer is 400, 200, and 50, respectively, and the number of nodes in the last layer is 4, which corresponds to the number of conditions. The coefficients for the L2 regularization term and the sparsity regularization term are 0.002 and 5, respectively. The desired proportion of training samples a neuron reacts to is 0.5. The

training accuracies and testing accuracies are averaged by 10 trials to reduce the effects of the randomness.

The diagnosis results of the developed method are shown in Figure 11. In these trials, the mean recognition rate reached 100% for the training samples and 99.92% for the test samples. The confusion matrix figures of one trail for the

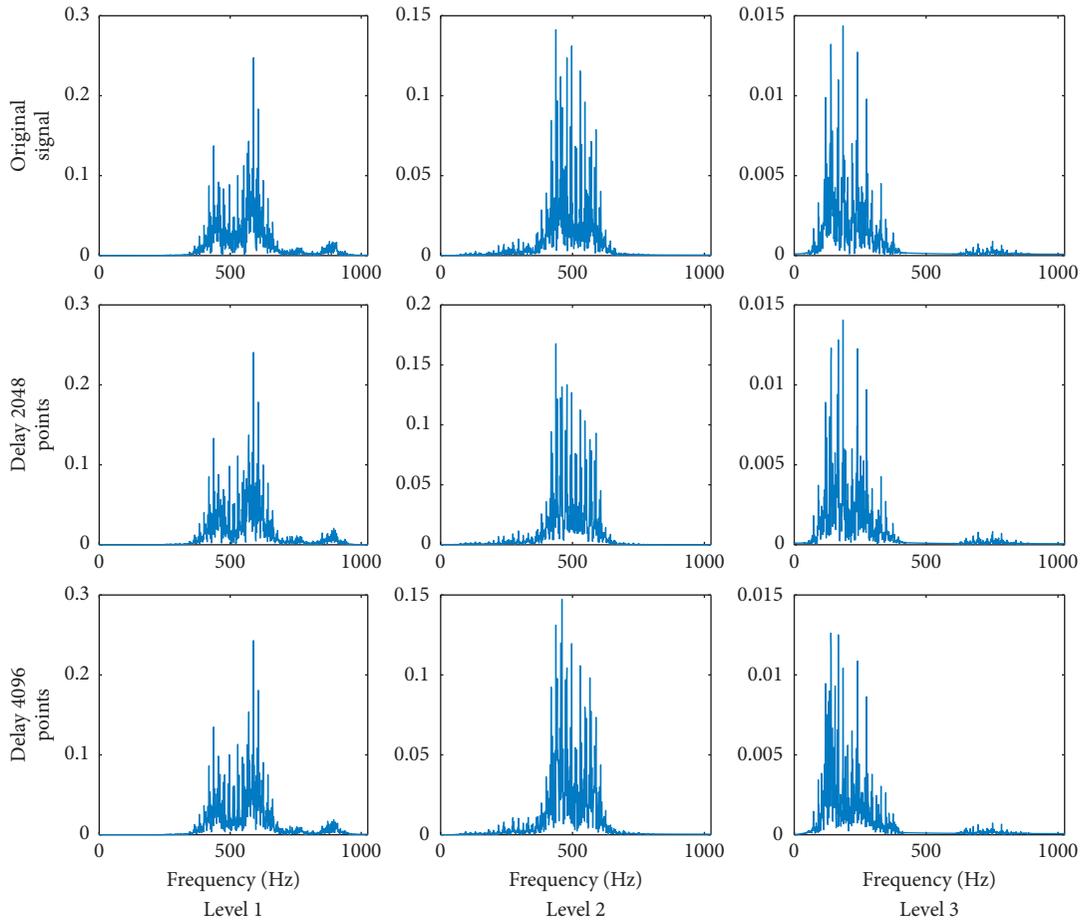


FIGURE 7: The first three levels of Fourier spectrum of the wavelet components of original signal and its delay series obtained by Db5 wavelet.

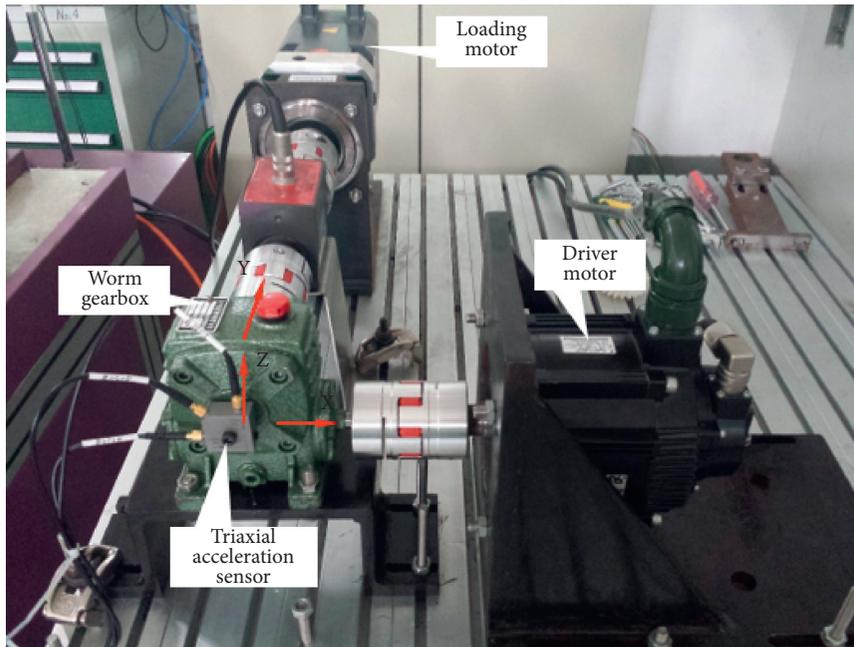


FIGURE 8: Worm gearbox test rig [25].

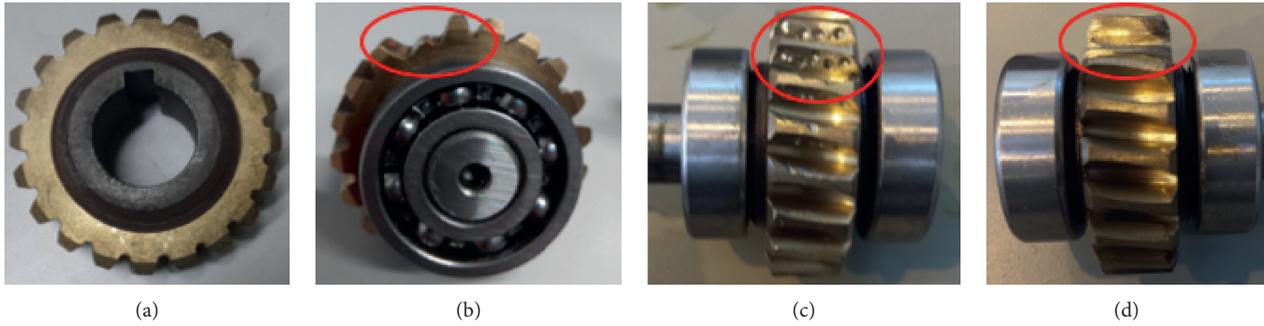


FIGURE 9: Normal worm gear and worm gears with artificial faults [25]: (a) normal; (b) broken; (c) pitting; (d) spalling.

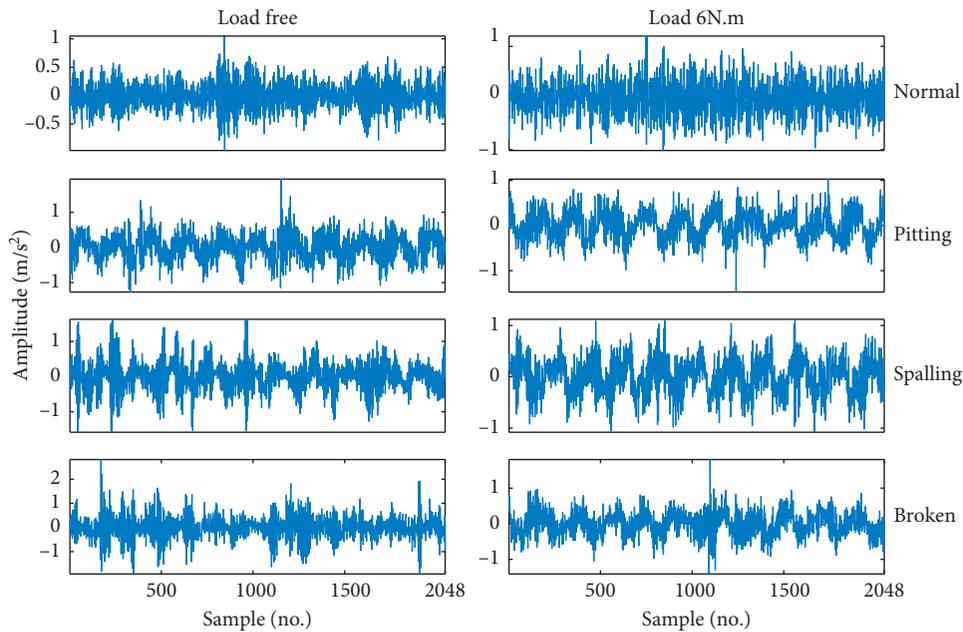


FIGURE 10: Raw vibration signals of four worm gear conditions.

training samples and test samples are shown in Figure 12. For 3600 test samples, only two samples with a broken fault are determined as pitting fault and spalling fault, respectively, and all other samples are classified correctly. When the percentage of training samples is selected as 25%, the classification accuracy increases to 99.98%.

Furthermore, for comparison, as shown in Table 3, we used the previously mentioned wavelet-SSAE model for the worm gear dataset, which has the same structure as the DCWT-SSAE. As shown in Figure 11, for the training samples, the classification accuracy is 100%, while for the test samples it gets an inferior accuracy; the mean accuracy is only 96.94%. When the percentage of training samples is selected as 25%, the classification accuracy goes up to 98.16%. In addition, Wang et al. [25] proposed a fault diagnosis scheme combining structured Fisher discrimination sparse coding with support vector machine (SFDSC). When their diagnosis scheme was used to classify the health conditions of the worm gear, 50% of the samples were used to train the model, and the other 50% were employed for the test. For free loading, the accuracy was 96.67%, and for

6 N m loading, the accuracy was 88.57%. Compared with these methods, the developed DCWT-SSAE obtains a higher accuracy even using a lower percentage of samples for training.

In this study, the main parameter selection of the developed DCWT-SSAE is a serious challenge. For the filters of double-tree complex wavelet and the structure of SSAE, there is still a lack of mature theoretical solutions. In this paper, the determination of these parameters depends on the experimentation and the practical problems to be solved whereas, for the parameters of the SSAE, the selection of sparsity parameters has a quite crucial effect on the classification performance. Here, the grid search strategy [26] is adopted to determine the sparsity regularization parameter and the sparsity proportion term. In this strategy, the L2 regularization term is selected as a constant 0.0016, the sparsity regularization term changes within the range [1, 10], and the sparsity proportion term changes within the range [0.1, 1]. As shown in Figure 13(a), the classification accuracy varies with the combination of parameters. It seems that the sparsity proportion term has a greater impact on the

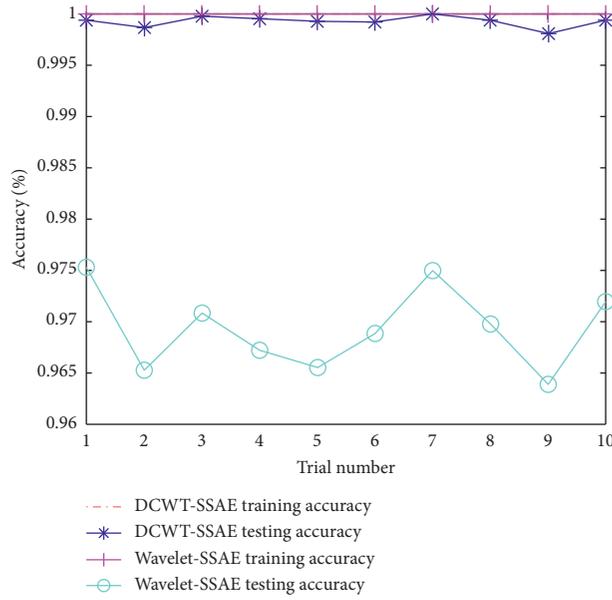


FIGURE 11: Diagnosis results comparison of 10 trials of worm gear dataset.

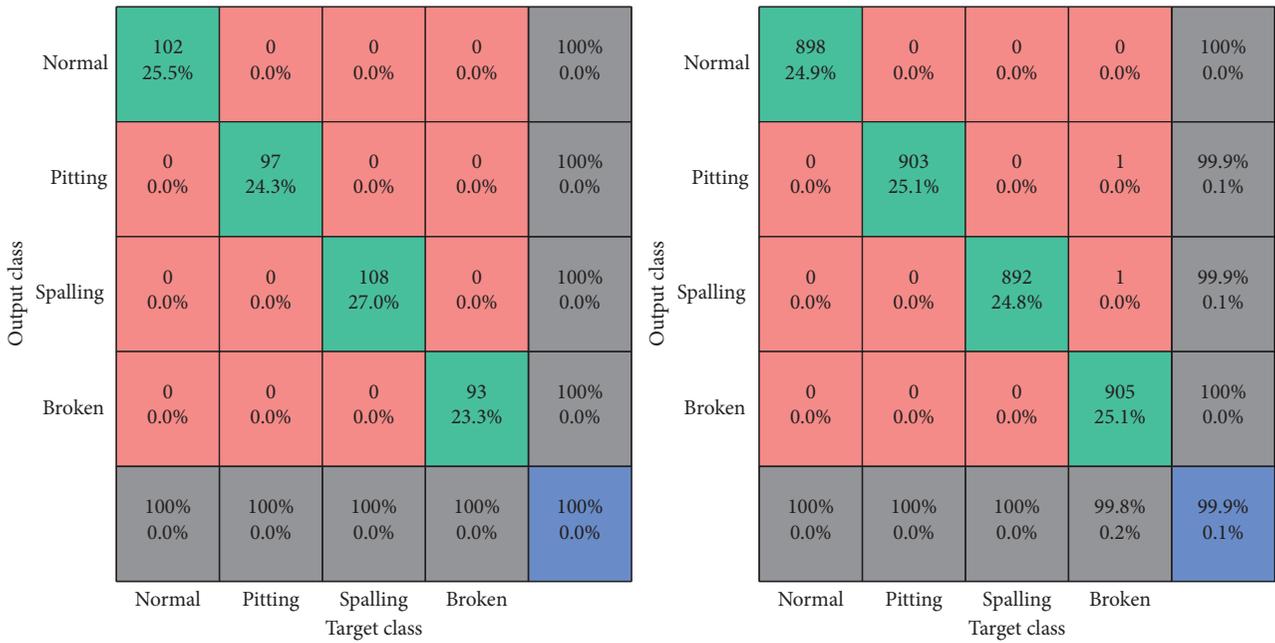


FIGURE 12: Confusion matrix of training samples (left panel) and test samples (right panel) of worm gear dataset.

TABLE 3: Classification comparison of worm gear dataset.

Methods	Percentage of training samples (%)	Load (N · m)	Number of health conditions	Test accuracy (%)
SFDSC [25]	50	0	4	96.67
SFDSC [25]	50	6	4	88.57
Wavelet-SSAE	10	0/6	4	96.94
Wavelet-SSAE	25	0/6	4	98.16
DCWT-SSAE	10	0/6	4	99.92
DCWT-SSAE	25	0/6	4	99.98

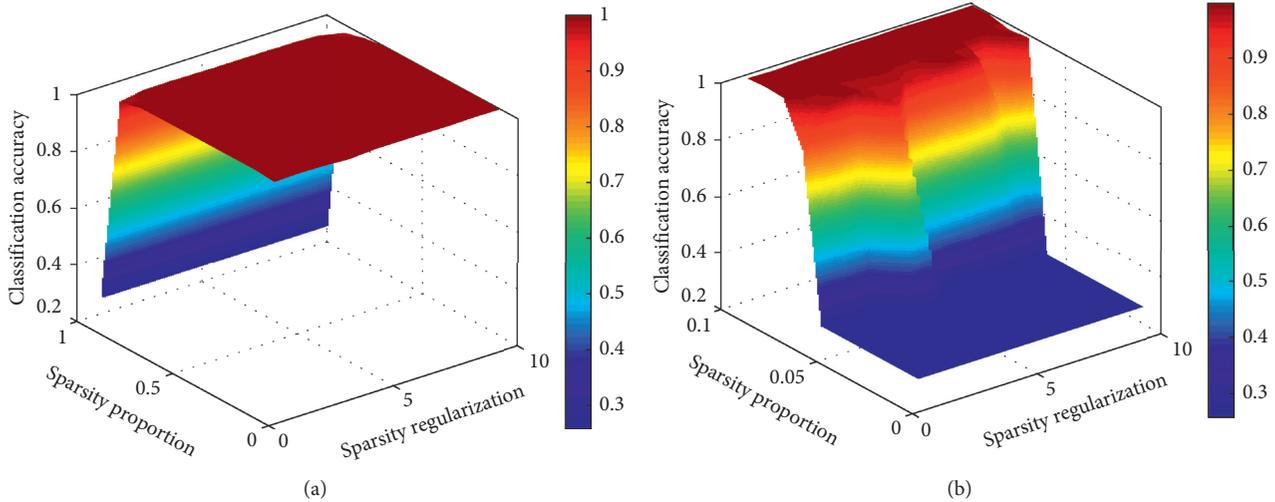


FIGURE 13: The relationship of classification accuracy and parameters for testing samples of worm gear dataset: (a) sparsity proportion term in $[0.1, 1]$ and the sparsity regularization term $[1, 10]$; (b) sparsity proportion term in $[0.01, 0.1]$; the sparsity regularization term $[1, 10]$.

classification performance for this case; the higher sparsity results in a superior recognition accuracy. If we further limit the range of the sparsity proportion term in $[0.01, 0.1]$, and the range of the sparsity regularization term keeps in $[1, 10]$, the distribution of the classification accuracy is shown in Figure 13(b). The classification performance fluctuates with the sparsity regularization term. We further set the L2 regularization term, the sparsity regularization term, and the sparsity proportion term as 0, 0, 1, and other parameters are consistent with the abovementioned DCWT-SSAE, which is a general stacked autoencoder model; the classification accuracy is only 97.86% when the percentage of training samples is selected as 10%, which is far inferior to the developed DCWT-SSAE method.

6. Conclusion

Extracting a valuable set of features is a crucial step in intelligent fault diagnosis. Traditional manual statistical features greatly depend on the experience of users, and usually the distributions of the obtained features are not separable enough for different conditions. As a result, it is hard to obtain sufficient diagnostic accuracy. Despite the recent successes of deep learning-based features that automatically learn for fault diagnosis, the feature learning from non-stationary and nonlinear signals has greatly limited further performance improvement. Alternatively, a dual-tree complex wavelet transform based stacked sparse autoencoder (DCWT-SSAE) was developed to learn a discriminative set of features automatically.

To verify the effectiveness of uncovering discriminative features from signals and fault diagnosis, the developed DCWT-SSAE was applied to motor bearings and worm gearbox gears and compared with state-of-the-art intelligent fault diagnosis methods. More specifically, the dual-tree complex wavelet transform and Fourier transform were employed to extract the multiscale features in the frequency domain, which benefits the shift invariance and statistic

stability. The stacked sparse autoencoder benefits from the unsupervised feature learning and receives promising diagnosis accuracy.

The bearing fault diagnosis experimental results indicated that the developed DCWT-SSAE fault diagnosis method improved performance by 10.81% and 7.21% compared with traditional manual time-domain features and wavelet features and led to 2.53% improvement compared with ensemble deep autoencoders. Keeping the other conditions unchanged and only substituting the general wavelet transform (the wavelet function is Db5) for the dual-tree complex wavelet transform, the diagnosis accuracy decreased by 1.33%; furthermore, if only the stacked sparse autoencoders are replaced with stacked autoencoders, the diagnosis accuracy decreased from 99.71% to 94.00%. Further comparing with a two-stage learning method consisting of sparse filtering and softmax regression, a superior diagnosis result of 99.71% to 99.66% was achieved in terms of the averages of 10-trial testing accuracies.

In addition, the worm gear fault diagnosis experiments confirmed that the developed DCWT-SSAE method achieved 99.92% diagnosis accuracy under variable load conditions with 10% of the training samples and 99.98% diagnosis accuracy with 25% of the training samples. If traditional wavelet transform (using Db5 wavelet) is used, the diagnosis accuracies are 96.94% and 98.16%, respectively. Although using a lower percentage of samples for training, when comparing with a combination fault diagnosis method of structured Fisher discrimination sparse coding and support vector machine, which used more training samples, the diagnosis accuracy improved by at least 3.42% and 11.80% for free load and 6 load conditions, respectively.

Based on these mentioned experimental results, the developed DCWT-SSAE can be regarded as a promising candidate for intelligent fault diagnosis of rotating machinery. However, the developed method would be appropriate for general data-driven fault diagnosis with nonstationary and nonlinear signals.

Data Availability

The motor bearing datasets analyzed during the current study are available in the Bearings Vibration Data Set from Case Western Reserve University. The available link is <https://csegroups.case.edu/bearingdatacenter/pages/download-data-file> (last accessed: 22 Feb 2020). The worm gearbox datasets in the current study cannot be shared at this time as the data also form part of an ongoing study. The interested readers can require the raw data from one of the corresponding authors via the e-mail address wswywy86@163.com.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to dedicate this paper to Myeongsu Kang, who unfortunately passed away in July 2018, while he was working as a research scientist with the Center for Advanced Life Cycle Engineering (CALCE). Dr. Kang was a great friend and scholar and played a significant role in this research and he is greatly missed. This work was supported by the National Nature Science Foundation of China (Grant nos. U1804141 and 51605061), the Program for Science and Technology Innovation Talents in Universities of Henan Province (Grant no. 17HASTIT028), the Program for Innovative Research Team in the University of Henan Province (Grant no. 20IRTSTHN015), and the Key Science and Technology Research Project of the Henan Province (Grant no. 202102210086).

References

- [1] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016.
- [2] J. Wang, S. Li, Z. An, X. Jiang, W. Qian, and S. Ji, "Batch-normalized deep neural networks for achieving fast intelligent fault diagnosis of machines," *Neurocomputing*, vol. 329, pp. 53–65, 2019.
- [3] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 11, pp. 6418–6428, 2014.
- [4] Y. Lei, N. Li, S. Gontarz, J. Lin, S. Radkowski, and J. Dybala, "A model-based method for remaining useful life prediction of machinery," *IEEE Transactions on Reliability*, vol. 65, no. 3, pp. 1314–1326, 2016.
- [5] H. Shao, H. Jiang, Y. Lin, and X. Li, "A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders," *Mechanical Systems and Signal Processing*, vol. 102, pp. 278–297, 2018.
- [6] M. Zhao, M. Kang, B. Tang et al., "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, 2017.
- [7] 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, pp. 171–178, 2016.
- [8] G. Jiang, H. He, P. Xie, and Y. Tang, "Stacked multilevel-denoising autoencoders: a new representation learning approach for wind turbine gearbox fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 9, pp. 2391–2402, 2017.
- [9] W. Du, M. Kang, and M. Pecht, "Fault diagnosis using adaptive multifractal detrended fluctuation analysis," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 3, pp. 2272–2282, 2020.
- [10] W. Yao, Z. Teng, Y. Gao, and Q. Tang, "Measurement of power system harmonic based on adaptive Kaiser self-convolution window," *IET Generation, Transmission & Distribution*, vol. 10, no. 2, pp. 390–398, 2016.
- [11] H. D. O. Mota, F. H. Vasconcelos, and C. L. de Castro, "A comparison of cycle spinning versus stationary wavelet transform for the extraction of features of partial discharge signals," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 23, no. 2, pp. 1106–1118, 2016.
- [12] N. G. Kingsbury, "The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters," in *Proceedings of the IEEE Digital Signal Process 1998*, University of Cambridge, Cambridge, UK, 1998.
- [13] H. Luo, M. Huang, and Z. Zhou, "A dual-tree complex wavelet enhanced convolutional LSTM neural network for structural health monitoring of automotive suspension," *Measurement*, vol. 137, pp. 14–27, 2019.
- [14] J. Qu, Z. Zhang, and T. Gong, "A novel intelligent method for mechanical fault diagnosis based on dual-tree complex wavelet packet transform and multiple classifier fusion," *Neurocomputing*, vol. 171, pp. 837–853, 2016.
- [15] R. Kumar, B. Singh, D. T. Shahani et al., "Dual-tree complex wavelet transform-based control algorithm for power quality improvement in a distribution system," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 1, pp. 764–772, 2016.
- [16] V. Ravi and M. Krishna, "A new online data imputation method based on general regression auto associative neural network," *Neurocomputing*, vol. 138, pp. 106–113, 2014.
- [17] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 7, pp. 1693–1702, 2017.
- [18] K. A. Loparo, *Bearings Vibration Data Set*, Case Western Reserve University, Cleveland, OH, USA, 2019, <http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website>.
- [19] Y. Wang, Z. He, and Y. Zi, "Enhancement of signal denoising and multiple fault signatures detecting in rotating machinery using dual-tree complex wavelet transform," *Mechanical Systems and Signal Processing*, vol. 24, no. 1, pp. 119–137, 2010.
- [20] L. V. D. Maaten and G. Hinton, "Visualizing high-dimensional data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.
- [21] N. Saravanan and K. I. Ramachandran, "Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN)," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4168–4181, 2010.
- [22] S. Ma, M. Chen, J. Wu, Y. Wang, B. Jia, and Y. Jiang, "High-voltage circuit breaker fault diagnosis using a hybrid feature transformation approach based on random forest and stacked autoencoder," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9777–9788, 2019.

- [23] W. Du, J. Tao, Y. Li, and C. Liu, "Wavelet leaders multifractal features based fault diagnosis of rotating mechanism," *Mechanical Systems and Signal Processing*, vol. 43, no. 1-2, pp. 57-75, 2014.
- [24] X. Jin, M. Zhao, T. W. S. Chow, and M. Pecht, "Motor bearing fault diagnosis using trace ratio linear discriminant analysis," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 5, pp. 2441-2451, 2014.
- [25] S. Wang, Y. Huang, L. Gong et al., "Improved feature extraction using structured Fisher discrimination sparse coding scheme for machinery fault diagnosis," *Advances in Mechanical Engineering*, vol. 8, no. 1, 2016.
- [26] S. K. Laha, "Enhancement of fault diagnosis of rolling element bearing using maximum kurtosis fast nonlocal means denoising," *Measurement*, vol. 100, pp. 157-163, 2017.