

Research Article

Fault Diagnosis of Rotating Machinery Based on One-Dimensional Deep Residual Shrinkage Network with a Wide Convolution Layer

Jingli Yang ¹, Tianyu Gao ¹, Shouda Jiang ¹, Shijie Li ² and Qing Tang ³

¹Department of Automatic Test and Control, Harbin Institute of Technology, Harbin 150001, China

²China Tobacco Henan Industrial Co., Ltd., Zhengzhou 450000, China

³China Institute of Marine Technology and Economy, Beijing 10001, China

Correspondence should be addressed to Jingli Yang; jinglidg@hit.edu.cn

Received 17 September 2020; Revised 3 November 2020; Accepted 26 November 2020; Published 9 December 2020

Academic Editor: Changqing Shen

Copyright © 2020 Jingli Yang 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.

In actual engineering applications, inevitable noise seriously affects the accuracy of fault diagnosis for rotating machinery. To effectively identify the fault classes of rotating machinery under noise interference, an efficient fault diagnosis method without additional denoising procedures is proposed. First, a one-dimensional deep residual shrinkage network, which directly takes the raw vibration signals contaminated by noise as input, is developed to realize end-to-end fault diagnosis. Then, to further enhance the noise immunity of the diagnosis model, the first layer of the model is set to a wide convolution layer to extract short time features. Moreover, an adaptive batch normalization algorithm (AdaBN) is introduced into the diagnosis model to enhance the adaptability to noise. Experimental results illustrate that the fault diagnosis model for rotating machinery based on one-dimensional deep residual shrinkage network with a wide convolution layer (1D-WDRSN) can accurately identify the fault classes even under noise interference.

1. Introduction

With the rapid development of integrated technology, complex industrial machinery systems are facing huge safety challenges [1, 2]. At present, as the core component of the mechanical system, rotating machinery always works in a complex environment. The operation failure of rotating machinery will affect the function of the entire mechanical system and even cause catastrophic accidents [3, 4]. Therefore, to ensure the continuous and effective operation of industrial machinery systems, research on intelligent diagnosis methods has turned into a trend in machinery health monitoring.

Traditional fault diagnosis of rotating machinery relied on the professional experience of engineers; thus, the diagnosis accuracy inevitably has a certain degree of randomness. In actual engineering applications, it is more desirable to adopt an automatic method to ensure the reliability of rotating machinery. The functional framework of the automatic fault diagnosis process, which consists of data

preprocessing, feature extraction, fault classification, and result analysis, is shown in Figure 1. Its main functions are as follows: the vibration sensors are used to collect the vibration signal data of the rotating machinery; the normalization and other preprocessing operations are performed to construct the original sample set; the appropriate feature extraction methods are adopted to find feature vectors that reflect the difference of sample features; an efficient fault classification model is trained to identify the fault classes of unknown samples; the diagnosis results are output and evaluated.

The development of artificial intelligence and machine learning has promoted the continuous development of fault diagnosis methods for rotating machinery in the direction of intelligence, thereby gradually forming a rotating machinery fault diagnosis mode that combines artificial feature extraction and shallow learning classification. In the above diagnosis mode, traditional machine learning classifiers, including sparse representation classifier (SRC) [5, 6], support vector machine (SVM) [7, 8], artificial neural network (ANN) [9], and nearest mean (NM) [10–12], are

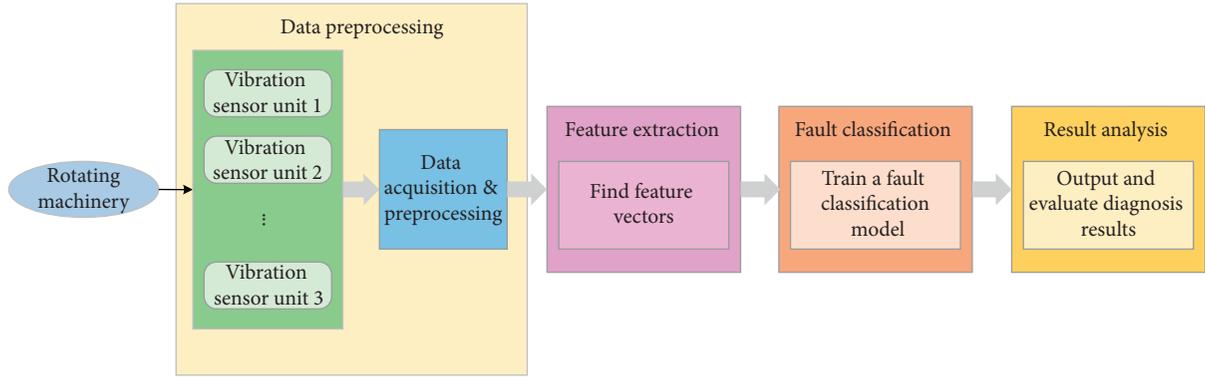


FIGURE 1: Functional framework of the automatic fault diagnosis process.

trained using sensitive features extracted from raw data to achieve automatic identification of fault classes. However, the traditional feature extraction methods based on manual intervention have a strong dependence on empirical factors, which makes it easy to lose useful information of samples. In addition, the classifier based on shallow learning has the problems of less robustness and poor generalization ability.

Recently, deep learning methods, including stacked autoencoder (SAE), deep neural network (DNN), convolution neural network (CNN), and deep belief network (DBN), are widely applied in various data mining and pattern recognition scenarios. With the above methods, the feature representation of samples is transformed from the raw space into a new feature space by multilayer feature transformation, thus improving the classification accuracy. Compared with the methods of manual intervention, deep learning methods can obtain rich internal information of the original data. Therefore, the application of deep learning methods has become a research hotspot in the field of fault diagnosis for rotating machinery. Chen et al. [13] extracted the time-domain and frequency-domain features of the raw vibration signals to train the CNN model, thereby achieving the fault diagnosis for the gearbox. Guo et al. [14] proposed a multitask convolutional neural network with information fusion, which solved the problem of fault diagnosis for rotating machinery under noise interference. Sun et al. [15] designed a fault diagnosis model based on a sparse autoencoder to identify the fault classes of the induction motor. First, the damage noise was added into the input layer of the sparse autoencoder to improve its robustness, and then the Dropout algorithm was introduced to avoid overfitting of the training process. Gai et al. [16] used the hybrid grey wolf optimizer to determine the penalty parameters and the number of the variational modal decomposition (VMD) components, and they performed the singular value decomposition (SVD) to construct the feature matrix of the vibration signals. The feature matrix is input into the DBN fault diagnosis model for training to realize the fault diagnosis for the rolling bearing. Pan et al. [17] constructed a hybrid fault diagnosis model based on CNN and the long-short term memory (LSTM) neural network, and they utilized the features learned by CNN to train the LSTM model, thereby accurately identifying the fault classes of the

bearing. To improve the feature learning ability of gearbox signals, Shao et al. [18] adopted the maximum entropy to construct the loss function of the deep autoencoder and used the artificial fish school algorithm to optimize the network parameters.

At present, it is worth noting that researchers usually carry out fault diagnosis based on ideal vibration signals of rotating machinery. In actual engineering applications, key information which can reflect the fault status of rotating machinery is always hidden in contaminated signals with noise [19, 20]. Therefore, the performance of existing fault diagnosis methods degrades seriously due to the noise interference [21, 22]. To realize health monitoring under actual engineering conditions, there is a certain practical value to study the fault diagnosis model for rotating machinery with noise immunity. Ou et al. [23] introduced the improved particle swarm optimization (IPSO) to optimize the majorization-minimization-based total variation (TV-MM) algorithm, thereby effectively removing the noise interference in signals. Zhang et al. [24] first adopted the raw signals to construct the Hankel matrix and restricted the weight vector to a unit vector. Then, to eliminate the random noise in the raw vibration signals, the $L_{3/2}$ -norm sparse filter was adopted to extract the different frequency components among the sample signals. Jiang et al. [25] found the converging U-shape phenomenon by investigating the variation features of the centre frequency (CF) in VMD, thereby constructing a CF-guided VMD optimization strategy to obtain adaptively the weak damage features of rotating machinery. Shen et al. [26] designed a fault diagnosis model based on stacked contractive autoencoder (CAE) for rotating machinery, which can obtain the hidden robust features by penalizing the Frobenius norm of the Jacobian matrix of the hidden features. Combining VMD and multiscale convolution neural networks, Wu et al. [27] established a multi-perspective fault diagnosis architecture, which provided the features of the raw signals in multiple perspectives including channel, component, and time scale. The noise robustness of the architecture is verified in the high-speed train dataset and bearing dataset. Yao et al. [28] replanned the positions of the bottleneck layers, convolution layers, and linear bottleneck layers of CNN to propose a stacked inverted residual CNN structure, and the proposed method can identify

bearing fault classes in different noise environments. Jin et al. [29] designed a hybrid model based on CNN and gated recurrent unit (GRU) neural network to solve the issue of fault diagnosis for bearing under noise interference. To improve the adaptive feature learning ability, the exponential linear unit (ELU) was introduced into CNN as an activation function, while the attention mechanism was added to the GRU. After the time-frequency decomposition of the vibration signals, Qiao et al. [30] also combined CNN and LSTM to identify the fault classes of bearings under noise interference. Zhang et al. [31] introduced the integrated learning mechanism into the deep shrinkage autoencoder and constructed a fault diagnosis model based on the integrated deep shrinkage autoencoder, which possesses a strong antinoise ability. Furthermore, LiftingNet [32], Bayesian network [33], concurrent convolution neural network [34], and capsule network [35] are also adopted to identify the fault classes of rotating machinery under noise interference.

To address the problem of low accuracy of fault diagnosis for rotating machinery caused by contaminated signals with noise, a novel method based on one-dimensional deep residual shrinkage network with a wide convolution layer (1D-WDRSN) is proposed in this paper. The innovative contributions of this paper are summarized as follows.

- (1) An end-to-end fault diagnosis method is provided; that is, the raw vibration signals contaminated by noise are directly input to obtain the diagnosis results. The fault identification for rotating machinery under different noise interference degrees can be realized by using the fault diagnosis model based on 1D-WDRSN with soft thresholding and attention mechanism.
- (2) The one-dimensional deep residual shrinkage network is developed to realize end-to-end fault diagnosis under noise interference. By introducing a suitable wide convolution layer to set the first layer of deep residual shrinkage network (DRSN), the short-term features of the raw vibration signals contaminated by noise can be obtained, thereby further enhancing the noise immunity of the network.
- (3) The AdaBN algorithm is used to enhance the domain adaptability of the network. By introducing the AdaBN algorithm, the training samples and test samples are adjusted to a new distribution space. Because the above two distributions are approximately the same in this distribution space, the influence of noise interference on the fault diagnosis model is reduced.

The paper is organized as follows. After the introduction, Section 2 illustrates the basic concept and procedures of the fault diagnosis method. To illustrate the performance of the proposed method for fault diagnosis of rotating machinery, experiments for various situations are conducted in Section 3. Finally, the conclusions are given in Section 4.

2. Methodology

2.1. Overview. To confront the challenge of fault diagnosis caused by noise interference, a fault diagnosis model based on 1D-WDRSN is proposed. Figure 2 shows the framework of the 1D-WDRSN model, which consists of a wide convolutional layer and multiple residual building units (RBUs). The fault diagnosis model based on 1D-WDRSN can realize adaptive feature learning, which extracts essential features from one-dimensional vibration signals contaminated by the noise and outputs the diagnosis results. Firstly, a larger size convolution kernel is adopted in the wide convolution layer to extract the short-term features of the signals, which can reduce noise interference to a certain extent. After the wide convolution layer, the Dropout technology is applied to avoid the overfitting problem in network training. A zero-filling technique is introduced in the RBUs to ensure that the features of different convolution kernels are of the same size. Moreover, the AdaBN algorithm is further introduced to accelerate the network training process and enhance the adaptability of the network under noise interference. Finally, the Softmax classifier is utilized to classify the extracted features, wherein cross entropy is constructed as a loss function to represent the error between the predicted value and the true value.

2.2. DRSN Model. By introducing soft thresholding and attention mechanism into ResNet, DRSN [36] is constructed to achieve accurate classification of noise-contaminated samples. The working principle of DRSN is to find the interference characteristics of the input samples according to the attention mechanism and use the soft threshold function to set them to zero, thereby reducing the influence of noise interference on the pattern recognition effect.

ResNet is a novel deep learning method that can solve the degradation problem caused by the increase of network depth. ResNet is composed of a series of RBUs, which can be expressed as follows.

$$x_{l+1} = x_l + F(x_l, W_l), \quad (1)$$

where x_l represents the identity shortcut part and $F(x_l, W_l)$ is the residual part. The residual part $F(x_l, W_l)$ is composed of basic components, including batch normalization (BN), rectified linear units (Relu), and convolutional layers (Conv).

The convolution kernel in the convolution layer can reduce the network training parameters, thereby reducing the probability of the network suffering from overfitting. The convolution mapping is as follows.

$$y_j = \sum_{i \in M_j} x_i * k_{ij} + b_j, \quad (2)$$

where x_i represents the i -th channel of the input feature, y_j represents the j -th channel of the output feature, M_j is the channel set for calculating y_j , k represents the convolution kernel, and b represents the bias.

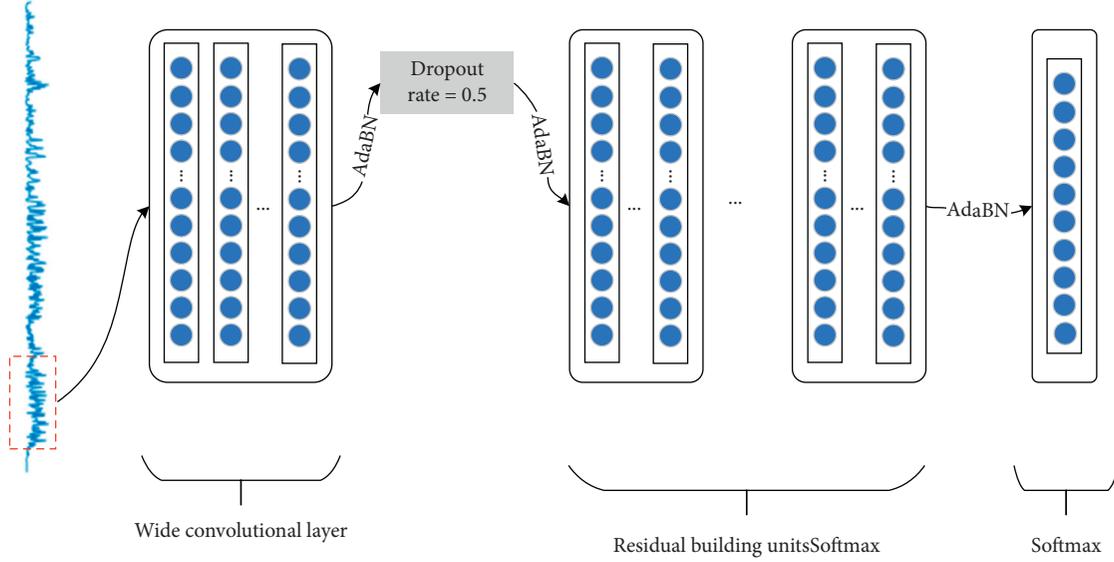


FIGURE 2: Framework of the fault diagnosis model based on 1D-WDRSN.

Adding BN between the neighbour convolutional layers can reduce the transfer of internal covariates, thereby improving the efficiency of network training and enhancing the generalization ability of the network. BN mainly performs feature normalization operations; that is, features are firstly normalized into the standard distribution and are then adjusted to the ideal distributions. The process of BN is expressed as follows:

$$\begin{aligned} \mu &= \frac{1}{N_{\text{batch}}} \sum_{n=1}^N x_n, \\ \sigma^2 &= \frac{1}{N_{\text{batch}}} \sum_{n=1}^N (x_n - \mu)^2, \hat{x}_n = \frac{x_n - \mu}{\sqrt{\sigma^2 + \varepsilon}}, y_n = \gamma \hat{x}_n + \beta, \end{aligned} \quad (3)$$

where N_{batch} represents the size of mini-batch, x_n and y_n are the input and output of the n -th observation value in the mini-batch, ε is a constant close to 0 to ensure numerical stability, γ is a scaling parameter, and β is a bias parameter.

The activation function is the nonlinear transformation part of the neural network. With the ability of addressing the issue of gradient disappearance, Relu is widely used in deep learning methods to help maintain the stability of features. The Relu activation function is expressed as follows.

$$y = \max(x, 0). \quad (4)$$

Compared with the ordinary network, ResNet introduces skip connections, which can make the information of the previous RBU flow into the next RBU without hindrance. Therefore, ResNet can not only accelerate the network convergence speed, but also avoid the problem of gradient disappearance and degradation caused by the increasing depth of network.

At present, soft thresholding is the core step of many denoising algorithms, which can remove the feature whose

absolute value is less than the threshold and can shrink the feature whose absolute value is greater than the threshold toward 0. The soft thresholding function is shown as follows:

$$y = \begin{cases} x - \tau, & x > \tau, \\ 0, & -\tau \leq x \leq \tau, \\ x + \tau, & x < -\tau, \end{cases} \quad (5)$$

where x and y are the input and output, respectively, and τ represents the threshold. The threshold setting must meet two conditions: one is that the threshold is positive and the other is that the threshold cannot be greater than the maximum value of input. In addition, it is better to set a corresponding independent threshold according to the input noise.

The derivative of the soft thresholding function is as follows:

$$\frac{\partial y}{\partial x} = \begin{cases} 1, & x > \tau, \\ 0, & -\tau \leq x \leq \tau, \\ 1, & x < -\tau. \end{cases} \quad (6)$$

It can be seen from the above that the function can only be 1 or 0, which has the same properties as Relu. Therefore, soft thresholding can not only reduce noise interference, but also avoid the issue of model gradient disappearance.

In the field of computer vision, the visual system of animal can quickly scan the entire area to find the target object and then focus on the target object to extract more essential information while suppressing irrelevant information. Squeeze-and-Excitation Network (SENet) [37] is a newer deep learning method using the attention mechanism. The contribution of various feature channels of the input samples to the classification task is often different. A small subnetwork is used to obtain a set of weights in SENet, which is multiplied with the features of each channel to adjust its

size. The above process can be considered as applying different amounts of attention to each feature channel. In the construction of SENet, each input sample will get an independent set of weights. The specific path is “global average pooling (GAP)→full connection layer (FC)→Relu→FC→Sigmoid”.

DRSN learns from the subnetwork structure of SENet to realize the soft thresholding using the attention mechanism. The threshold can be automatically set through the learning of the subnetwork in the red dashed box as shown in Figure 3. In this subnetwork, the absolute values of the input sample features are calculated, and then a feature A is obtained after GAP. In the other path, the features after GAP are input into a small full connection network, where the Sigmoid function is adopted as the last layer, and the output is normalized to obtain a coefficient α . Finally, the threshold corresponding to the input sample can be expressed as $\alpha \times A$, which meets the above two conditions of the soft thresholding.

Since the corresponding threshold can be set according to different samples, DRSN introduces a special attention mechanism, which can discover the interference features of the input samples and set them to zero through the soft thresholding. In other words, it can pay attention to the essential features and reserve them in the network.

2.3. Enhancement of Feature Learning Ability Based on the Wide Convolution Layer. Feature extraction is the basis of fault diagnosis for rotating machinery, the purpose of which is to mine the essential information in the raw signals. However, due to the complex working environment of rotating machinery, the actual measured vibration signals will be contaminated by noise, which increases the difficulty of feature extraction. Therefore, to enhance the feature learning ability, the wide convolutional layer is adopted as the first layer of 1D-WDRSN in this paper.

The wide convolution layer can extract short-term features like the classical short-term Fourier transform, and the difference is that the former can adaptively learn the effective features of vibration signals and remove interference features that affect the diagnosis results. The core idea of the wide convolution layer is to improve the noise immunity of the model by using a larger convolution kernel. In general, a large convolution kernel has a larger reception field, thereby ignoring noise interference. Zhang et al. [38] prove that the large convolution kernel can be regarded as a low-pass filter, which can effectively obtain the low-frequency features of the raw vibration signals and suppress high-frequency noise interference. In the proposed fault diagnosis model based on 1D-WDRSN, the wide convolutional layer contains a large convolution kernel, and the appropriate convolution kernel size will be determined through experiments in the next section.

2.4. Improvement of Domain Adaptability Based on AdaBN. In the field of image recognition, AdaBN is an improved BN algorithm with the domain adaptability. In AdaBN, the mean and variance of the BN layer achieved from the source domain samples are replaced by the mean and variance of

the target domain samples. By performing BN on the source domain samples and AdaBN on the target domain samples, the two domains can be adjusted into a new distribution space, thereby achieving the purpose of field adaptation. Therefore, AdaBN can not only ensure that the network parameters are shared between the source domain and the target domain, but also make the statistical information of the two domains be independent at the BN layer.

In this paper, AdaBN algorithm is introduced into 1D-WDRSN to accelerate network training process and improve the adaptability of the network to noise interference. Firstly, training samples are employed to optimize and train the 1D-WDRSN model. If the distributions of the training samples and the test samples are inconsistent, the mean and variance of all BN layers in the network are replaced with those of the test samples while keeping other parameters unchanged. Finally, the 1D-WDRSN model with domain adaptability is used to implement fault diagnosis for rotating machinery. The specific description of 1D-WDRSN based on AdaBN is shown in Algorithm 1.

2.5. Fault Diagnosis Process. The process of the fault diagnosis model based on 1D-WDRSN for rotating machinery proposed in this paper is shown in Figure 4. The 1D-WDRSN model is introduced for adaptive feature learning, which can extract key information from signals contaminated by the noise and accurately identify fault classes of rotating machinery. The specific steps of the fault diagnosis process are as follows:

- (1) Randomly divide the original vibration signal samples into a training sample set which is employed to train the 1D-WDRSN model and a test sample set that is employed to verify the effectiveness of the 1D-WDRSN model.
- (2) Set training parameters, including the number of iterations, learning rate, and training batch.
- (3) Initialize network parameters, including weights and offsets.
- (4) Perform layer-by-layer training in a supervised learning method, and use backpropagation algorithm to fine-tune network parameters. When the number of iterations is met, the training of the 1D-WDRSN model is completed.
- (5) Input the test sample set into the 1D-WDRSN model to obtain diagnosis results.

3. Experiments and Results

To verify the effectiveness of the 1D-WDRSN model on addressing the problem of fault diagnosis under noise interference, the vibration signals obtained from the deep groove ball bearing in the bearing laboratory of Case Western Reserve University (CWRU) are chosen as original experimental data. All fault diagnosis experiments are performed on a computer with 8 GB RAM, and the 1D-WDRSN model is built in the TensorFlow environment using Python.

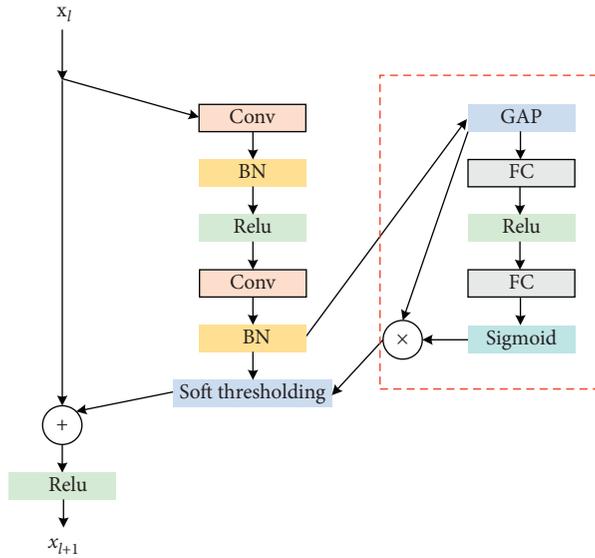


FIGURE 3: Deep residual shrinkage network.

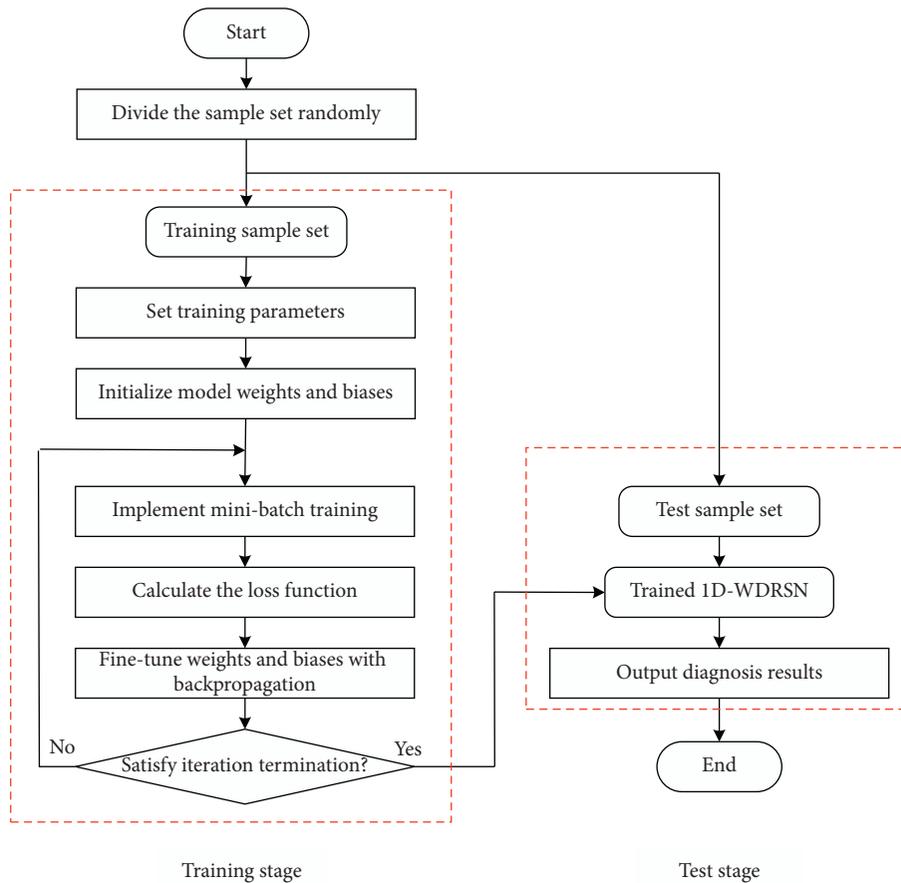


FIGURE 4: Flowchart of the proposed fault diagnosis model.

3.1. Data Description. The data acquisition system of CWRU bearing laboratory is shown in Figure 5. The damage locations of the bearing are outer race, inner race, and ball, wherein the fault diameter at each damage location is 0.007

inches, 0.014 inches, and 0.021 inches, respectively. Therefore, the dataset includes ten bearing fault classes. A dataset under a load of 1 hp is adopted in this paper. Since each vibration signal sample contains 1024 data points, the entire

Input: Sample of the target domain p , the expression of the i -th neuron in the BN layer of 1D-WDRSN $x_t^{(i)}(p) \in x_t^{(i)}$, and the trained network parameters $\gamma_S^{(i)}$ and $\beta_S^{(i)}$

Output: 1D-WDRSN with the domain adaptability

For: Calculate the mean and variance of all samples in the target domain

$$\mu_t^{(i)} \leftarrow E[x_t^{(i)}]$$

$$\sigma_t^{(i)} \leftarrow \text{Var}[x_t^{(i)}]$$

Calculate the output of the BN layer

$$\hat{x}_t^{(i)}(p) = x_t^{(i)}(p) - \mu_t^{(i)}/\sigma_t^{(i)}$$

$$y_t^{(i)}(p) = \gamma_S^{(i)}\hat{x}_t^{(i)}(p) + \beta_S^{(i)}$$

End For

ALGORITHM 1: 1D-WDRSN based on AdaBN.

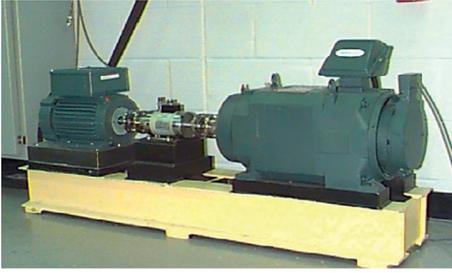


FIGURE 5: Bearing data acquisition system used by CWRU.

dataset consists of 1000 samples, where the ratio of the training sample set to the test sample set is 7 : 3. The detailed information of the dataset is described in Table 1.

3.2. Noise Preparation. Due to the complex working environment of rotating machinery, the noise interference exists in actual engineering. Therefore, Gaussian white noise is introduced to simulate the real noise, and the signal to noise ratio (SNR) is used to measure the noise intensity. SNR can be calculated as follows.

$$\text{SNR}_{dB} = 10 \log \left(\frac{P_{\text{signal}}}{P_{\text{noise}}} \right), \quad (7)$$

where P_{signal} is the signal energy and P_{noise} is the noise energy. Obviously, SNR is inversely proportional to noise intensity.

To evaluate the noise immunity of the fault diagnosis model proposed in this paper, the fault diagnosis experiments for rotating machinery are performed in a noise environment, where the SNR changes from -4 dB to 10 dB. With the purpose of simulating the actual engineering application, the raw vibration signals of the rotating machinery are adopted to establish the training sample set in our experiment, while noise interference is added to samples in the test sample set. Ten training sample sets and test sample sets are randomly generated for each noise intensity, thereby reducing the randomness in the results of fault diagnosis experiments.

3.3. Selection of Model Parameters. Considering that several parameters, i.e., the size and number of convolution kernels in the wide convolution layer, the number of RBUs, and the size of convolution kernels in the RBUs, have a great impact on the diagnosis performance of the proposed 1D-WDRSN model, a series of experiments are implemented to determine the values of these parameters. The vibration signals contaminated by noise of SNR = -4 dB are adopted as the test samples in these experiments. To reduce the random factors of the results, the average fault diagnosis accuracy of 30 experiments is adopted as the final result.

3.3.1. The Size of Convolution Kernels in the Wide Convolution Layer. As shown previously, a larger convolution kernel can enhance noise immunity of the network. Therefore, the size of convolution kernels is set to be 48, 64, 96, 128, 192, 256, 384, and 512 for this experiment. The number of convolution kernels is fixed to 32, the number of RBUs is fixed to 2, the size of convolution kernels in RBUs is fixed to 16, and the influence of the size of convolution kernels in the wide convolution layer on the diagnosis performance is analysed in this experiment. The experimental results are shown in Table 2.

As shown in Table 2, the accuracy increases steadily with the size of convolution kernels. When the size of convolution kernel in the wide convolution layer reaches 256, the fault diagnosis accuracy tends to stabilize. Therefore, the size of convolution kernels is set to 256.

3.3.2. The Number of Convolution Kernels in the Wide Convolution Layer. After determining the size, the influence of the number of convolution kernels in the wide convolution layer on the 1D-WDRSN model is further analysed. In this experiment, the size of convolution kernels is fixed to 256, the number of RBUs is set to 2, and the size of convolution kernels in RBUs is fixed to 16. The experimental results are shown in Table 3.

As shown in Table 3, the fault diagnosis accuracy increases steadily with the number of convolution kernels. When the number of convolution kernels in the wide

TABLE 1: The detailed information of the dataset.

Damage location	None		Outer race		Inner race			Ball		
Label	1	2	3	4	5	6	7	8	9	10
Damage diameter (inch)	0	0.007	0.014	0.021	0.007	0.014	0.021	0.007	0.014	0.021
Training samples	70	70	70	70	70	70	70	70	70	70
Test samples	30	30	30	30	30	30	30	30	30	30

TABLE 2: Influence of the size of convolution kernels in the wide convolution layer on the 1D-WDRSN model.

The size of the convolution kernel	48	64	96	128 (%)	192 (%)	256 (%)	384 (%)	512 (%)
Accuracy	71.44	85.17	83.37	89.84	91.51	91.64	91.64	91.37

convolution layer reaches 48, the fault diagnosis accuracy tends to stabilize. Therefore, the number of convolution kernels is set to 48.

3.3.3. The Number of RBUs. After determining the parameters of the wide convolution layer, the influence of the number of RBUs on the 1D-WDRSN model is further analysed. In this experiment, the size of convolution kernels in RBUs is fixed to 16. The experimental results are shown in Table 4.

As shown in Table 4, the accuracy becomes stable when the number of RBUs is larger than 2. Considering that the increase in the number of RBUs will reduce the network training speed, 2 RBUs are adopted in the 1D-WDRSN model.

3.3.4. The Size of Convolution Kernels in RBUs. Finally, the influence of the size of convolution kernels in RBUs on the diagnosis performance of the 1D-WDRSN model is shown in Table 5.

As shown in Table 5, the fault diagnosis accuracy achieves the highest value when the number of RBUs is 16. Therefore, the size of convolution kernels in RBUs is set to 16.

According to the above results, the network parameters of the proposed 1D-WDRSN model are summarized in Table 6. The model consists of a wide convolutional layer, two residual blocks, and a Softmax layer. The 10 fault classes of the bearing are corresponding to the 10 outputs of the Softmax layer.

3.4. Noise Immunity Evaluation Experiments. The parameters adopted for training the network are set as follows. The Adam optimizer is applied to the 1D-WDRSN model, where the learning rate is set to 0.001 and the number of iterations is set to 200. The Dropout rate after the wide convolution layer is set to 0.5.

To verify the adaptive feature learning ability of the proposed fault diagnosis model based on 1D-WDRSN under noise interference, visualizations of the raw vibration signals and features extracted by the 1D-WDRSN model with noise of SNR = 0 dB are shown in Figures 6 and 7, respectively. Considering the high dimensionality of original signals and extracted features, t -distributed stochastic neighbour

embedding (t -SNE) technology is adopted to visualize the performance of the feature extraction. The extracted features are mapped in a 3-dimensional scatter plot, which provides a way to evaluate the effectiveness of features. The three axes of the scatter plot are named “Feature 1”, “Feature 2”, and “Feature 3”.

As shown in Figure 6, it is observed that most fault classes of original signals have different degrees of overlaps with other fault classes. In contrast, Figure 7 presents that the 10 fault classes of the features extracted by 1D-WDRSN model are clearly separated from each other. From the perspective of fault classification, the higher accuracy of fault diagnosis can be achieved by inputting those features into the classifier. The above results demonstrate that the 1D-WDRSN model can mine the key information of the signals under noise interference.

In addition, the proposed method is compared with other fault diagnosis methods of rotating machinery under noise interference to verify its effectiveness. The comparison methods are sparse filtering (SF) [24], multiperspective CNN (MP-CNN) [27], CNN + LSTM [30], wide convolution and multiscale convolution (WMSCCN) [35], and deep convolutional neural networks with wide first-layer kernels (WDCNN) [38]. In the case of different noise intensities, fault diagnosis results of the above methods are shown in Figure 8.

It can be concluded from the above results that the proposed fault diagnosis method for rotating machinery achieves the best performance under different degrees of noise interference. In addition, this advantage is more obvious in the case of low SNR. The reasons can be explained as follows. Discovering the interference features of the input samples by using the attention mechanism, 1D-WDRSN employs the soft threshold function to zero for those features, thereby reducing the influence of noise interference on the fault diagnosis accuracy. Moreover, a wide convolution layer in the beginning of the 1D-WDRSN model can suppress high-frequency noise interference, which helps to figure out key information of different fault classes from vibration signals contaminated by noise. In addition, the AdaBN algorithm is introduced to adjust the training samples and test samples to a suitable distribution space, thereby strengthening the domain adaptability of the model.

TABLE 3: Influence of the number of convolution kernels in the wide convolution layer on the 1D-WDRSN model.

The number of the convolution kernels	4	8	16	24	32	48	64	96 (%)
Accuracy	31.37	53.54	74.51	84.84	91.64%	94.44	92.17	94.37

TABLE 4: Influence of the number of RBUs on the 1D-WDRSN model.

The number of RBUs	1 (%)	2	3	4	5 (%)
Accuracy	87.84	94.44%	92.37%	93.17	93.84

TABLE 5: Influence of the size of convolution kernels in RBUs on the 1D-WDRSN model.

The number of the convolution kernel	4 (%)	8 (%)	16 (%)	24 (%)	32	48 (%)	64 (%)
Accuracy	87.84	93.51	94.44	93.42	89.44%	91.17	90.93

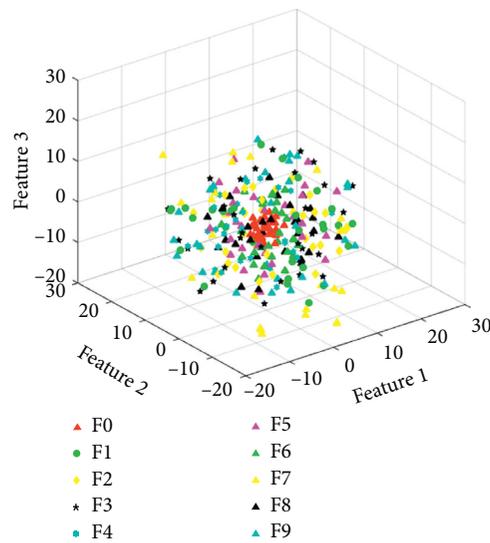


FIGURE 6: Visualization of raw vibration signals.

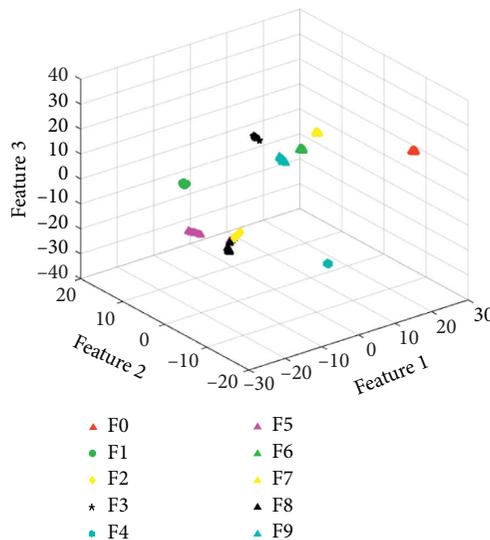


FIGURE 7: Visualization of features extracted by the 1D-WDRSN model.

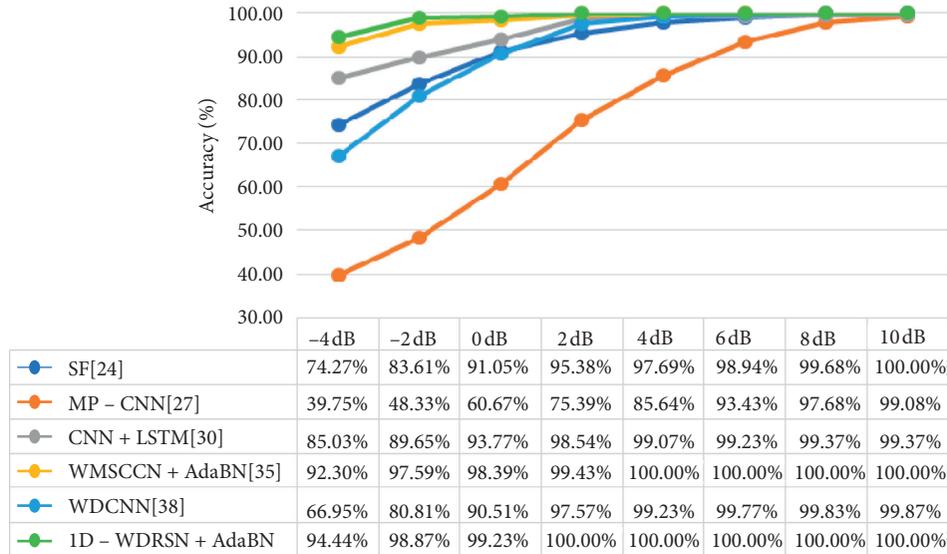


FIGURE 8: Accuracy of all methods under different noise intensities.

TABLE 6: The network parameters of the 1D-WDRSN model.

No.	Layer type	Kernel size	Kernel number	Output size (Width×Depth)
1	Conv1d_1	256 × 1	48	512 × 48
2	RBU_1_Conv_1	16 × 1	48	256 × 48
3	RBU_1_Conv_2	16 × 1	48	256 × 48
4	RBU_2_Conv_1	16 × 1	48	128 × 48
5	RBU_2_Conv_2	16 × 1	48	128 × 48
6	Softmax	10	1	10

4. Conclusions

To solve the problem of poor noise immunity of the fault diagnosis model for rotating machinery, an end-to-end method based on 1D-WDRSN is proposed in this paper. Firstly, to remove the interference features of the input samples achieved by the attention mechanism, the soft threshold function is introduced into the 1D-WDRSN model, thereby reducing the influence of noise interference on the fault diagnosis accuracy. Then, a wide convolution layer is adopted to suppress high-frequency noise interference, which helps to figure out key information of different fault classes from the vibration signals contaminated by noise. After the wide convolution layer, the Dropout technology is applied to avoid the overfitting problem in network training. Finally, to further improve the adaptability of the network under noise interference, the AdaBN algorithm is employed to adjust the training samples and test samples to a suitable distribution space in the fault diagnosis process. The vibration signals obtained from the deep groove ball bearing in the bearing laboratory of CWRU are chosen as original experimental data. Different degrees of noise are added into the original vibration signal to implement simulation experiments of fault diagnosis under noise interference. Experimental results show that the fault diagnosis method based on 1D-WDRSN can effectively identify the fault classes of rotating machinery under noise interference. In brief, the fault diagnosis based on

1D-WDRSN can improve the reliability and maintainability of rotating machinery and provide a good solution for the field of machinery health monitoring. Future work will be devoted to applying the proposed fault diagnosis model to the mechanical system of the water jet propulsion device, which consists of various rotating machinery.

Data Availability

The Case Western Reserve University (CWRU) dataset can be obtained by visiting <https://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website>.

Conflicts of Interest

The authors declare that they do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Acknowledgments

This research was supported by the Aerospace Science and Technology CAST-BISEE Foundation of China (nos. 2019–020). The authors would like to gratefully acknowledge this support.

References

- [1] D. Liu, W. Cheng, and W. Wen, "An online bearing fault diagnosis technique via improved demodulation spectrum analysis under variable speed conditions," *IEEE Systems Journal*, vol. 14, no. 2, pp. 2323–2334, 2020.
- [2] Y. Jiang, Z. Li, C. Zhang et al., "On the bi-dimensional variational decomposition applied to nonstationary vibration signals for rolling bearing crack detection in coal cutters," *Measurement Science and Technology*, vol. 27, no. 6, Article ID 065103, 2016.
- [3] L. Cui, J. Huang, and F. Zhang, "Quantitative and localization diagnosis of a defective ball bearing based on vertical-horizontal synchronization signal analysis," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 11, pp. 8695–8706, 2017.
- [4] A. Abid, M. T. Khan, and C. W. de Silva, "Layered and real-valued negative selection algorithm for fault detection," *IEEE Systems Journal*, vol. 12, no. 3, pp. 2960–2969, 2018.
- [5] C. Wang, M. Gan, and C. a. Zhu, "Intelligent fault diagnosis of rolling element bearings using sparse wavelet energy based on overcomplete DWT and basis pursuit," *Journal of Intelligent Manufacturing*, vol. 28, no. 6, pp. 1377–1391, 2017.
- [6] C. Wang, M. Gan, and C. a. Zhu, "Fault feature extraction of rolling element bearings based on wavelet packet transform and sparse representation theory," *Journal of Intelligent Manufacturing*, vol. 29, no. 4, pp. 937–951, 2018.
- [7] W. Deng, R. Yao, M. Sun et al., "Study on a novel fault diagnosis method based on integrating EMD, fuzzy entropy, improved PSO and SVM," *Journal of Vibroengineering*, vol. 19, no. 4, pp. 2562–2577, 2017.
- [8] Y. Zhang, Q. Gao, Y. Lu, D. Sun, Y. Xia, and X. Peng, "A novel intelligent method for bearing fault diagnosis based on hermitian scale-energy spectrum," *IEEE Sensors Journal*, vol. 18, no. 16, pp. 6743–6755, 2018.
- [9] F. Zhang, J. Yan, P. Fu, L. Wang, and R. Gao, "Ensemble sparse supervised model for bearing fault diagnosis in smart manufacturing," *Robotics and Computer-Integrated Manufacturing*, vol. 65, Article ID 101920, 2020.
- [10] W. Caesarendra, M. Pratama, B. Kosasih et al., "Parsimonious network based on a fuzzy inference system (PANFIS) for time series feature prediction of low speed slew bearing prognosis," *Applied Sciences-Basel*, vol. 8, no. 12, Article ID 122656, 2018.
- [11] A. Glowacz, "Recognition of acoustic signals of commutator motors," *Applied Sciences-Basel*, vol. 8, no. 12, Article ID 122630, 2018.
- [12] A. Glowacz, W. Glowacz, J. Kozik et al., "Detection of deterioration of three-phase induction motor using vibration signals," *Measurement Science Review*, vol. 19, no. 6, pp. 241–249, 2019.
- [13] Z. Chen, C. Li, and R. Sanchez, "Gearbox fault identification and classification with convolutional neural networks," *Shock and Vibration*, vol. 2015, Article ID 390134, 10 pages, 2015.
- [14] S. Guo, B. Zhang, T. Yang, D. Lyu, and W. Gao, "Multitask convolutional neural network with information fusion for bearing fault diagnosis and localization," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 9, pp. 8005–8015, 2020.
- [15] 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.
- [16] J. Gai, J. Shen, Y. Hu, and H. Wang, "An integrated method based on hybrid grey wolf optimizer improved variational mode decomposition and deep neural network for fault diagnosis of rolling bearing," *Measurement*, vol. 162, p. 2020, Article ID 107901.
- [17] H. Pan, X. He, S. Tang, and F. Meng, "An improved bearing fault diagnosis method using one-dimensional CNN and LSTM Strojnski Vestnik," *Journal of Mechanical Engineering*, vol. 64, no. 7-8, pp. 443–452, 2018.
- [18] H. Shao, H. Jiang, H. Zhao, and F. Wang, "A novel deep autoencoder feature learning method for rotating machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 95, pp. 187–204, 2017.
- [19] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mechanical Systems and Signal Processing*, vol. 100, pp. 439–453, 2018.
- [20] D. Peng, H. Wang, Z. Liu, W. Zhang, M. J. Zuo, and J. Chen, "Multibranch and multiscale CNN for fault diagnosis of wheelset bearings under strong noise and variable load condition," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4949–4960, 2020.
- [21] S. Yang, X. Gu, Y. Liu, R. Hao, and S. Li, "A general multi-objective optimized wavelet filter and its applications in fault diagnosis of wheelset bearings," *Mechanical Systems and Signal Processing*, vol. 145, Article ID 106914, 2020.
- [22] Q. He, J. Zhao, G. Jiang, and P. Xie, "An unsupervised multiview sparse filtering approach for current-based wind turbine gearbox fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 8, pp. 5569–5578, 2020.
- [23] Y. Ou, S. He, C. Hu, J. Bao, and W. Li, "Research on rolling bearing fault diagnosis using improved majorization-minimization-based total variation and empirical wavelet transform," *Shock and Vibration*, vol. 2020, Article ID 3218564, 11 pages, 2020.
- [24] Z. Zhang, S. Li, J. Wang, Y. Xin, Z. An, and X. Jiang, "Enhanced sparse filtering with strong noise adaptability and its application on rotating machinery fault diagnosis," *Neurocomputing*, vol. 398, pp. 31–44, 2020.
- [25] X. Jiang, C. Shen, J. Shi, and Z. Zhu, "Initial center frequency-guided VMD for fault diagnosis of rotating machines," *Journal of Sound and Vibration*, vol. 435, pp. 36–55, 2018.
- [26] C. Shen, Y. Qi, J. Wang et al., "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 170–184, 2018.
- [27] Y. Wu, W. Jin, J. Ren, and Z. Sun, "A multi-perspective architecture for high-speed train fault diagnosis based on variational mode decomposition and enhanced multi-scale structure," *Applied Intelligence*, vol. 49, no. 11, pp. 3923–3937, 2019.
- [28] D. Yao, H. Liu, J. Yang, and X. Li, "A lightweight neural network with strong robustness for bearing fault diagnosis," *Measurement*, vol. 159, Article ID 107756, 2020.
- [29] G. Jin, T. Zhu, M. W. Akram, Y. Jin, and C. Zhu, "An adaptive anti-noise neural network for bearing fault diagnosis under noise and varying load conditions," *IEEE Access*, vol. 8, pp. 74793–74807, 2020.
- [30] M. Qiao, S. Yan, X. Tang, and C. Xu, "Deep convolutional and LSTM recurrent neural networks for rolling bearing fault diagnosis under strong noises and variable loads," *IEEE Access*, vol. 8, pp. 66257–66269, 2020.
- [31] Y. Zhang, X. Li, L. Gao, W. Chen, and P. Li, "Ensemble deep contractive auto-encoders for intelligent fault diagnosis of

- machines under noisy environment,” *Knowledge-Based Systems*, vol. vol. 196, Article ID 105764, 2020.
- [32] J. Pan, Y. Zi, J. Chen, Z. Zhou, and B. Wang, “LiftingNet: a novel deep learning network with layerwise feature learning from noisy mechanical data for fault classification,” *IEEE Transactions on Industrial Electronics*, vol. 65, no. 6, pp. 4973–4982, 2018.
- [33] G. Chen, M. Liu, and J. Chen, “Frequency-temporal-logic-based bearing fault diagnosis and fault interpretation using Bayesian optimization with Bayesian neural networks,” *Mechanical Systems and Signal Processing*, vol. 145, Article ID 106951, 2020.
- [34] Y. Chang, J. Chen, C. Qu, and T. Pan, “Intelligent fault diagnosis of wind turbines via a deep learning network using parallel convolution layers with multi-scale kernels,” *Renewable Energy*, vol. 153, pp. 205–213, 2020.
- [35] Y. Wang, D. Ning, and S. Feng, “A novel capsule network based on wide convolution and multi-scale convolution for fault diagnosis,” *Applied Sciences-Basel*, vol. 10, no. 10, Article ID 103659, 2020.
- [36] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, “Deep residual shrinkage networks for fault diagnosis,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4681–4690, 2020.
- [37] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, “Squeeze-and-Excitation networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011–2023, 2020.
- [38] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, “A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals,” *Sensors*, vol. 17, no. 2, Article ID 17020425, 2017.