

Research Article

A Novel Deep Sparse Filtering Method for Intelligent Fault Diagnosis by Acoustic Signal Processing

Guowei Zhang, Jinrui Wang, Baokun Han , Sixiang Jia, Xiaoyu Wang, and Jingtao He

College of Mechanical and Electronic Engineering, Shandong University of Science and Technology, Qingdao, China

Correspondence should be addressed to Baokun Han; bk_han@163.com

Received 2 April 2020; Revised 18 June 2020; Accepted 11 July 2020; Published 26 July 2020

Academic Editor: Changqing Shen

Copyright © 2020 Guowei Zhang 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.

Increased attention has been paid to research on intelligent fault diagnosis under acoustic signals. However, the signal-to-noise ratio of acoustic signals is much lower than vibration signals, which increases the difficulty of signal denoising and feature extraction. To solve the above defect, a novel batch-normalized deep sparse filtering (DSF) method is proposed to diagnose the fault through the acoustic signals of rotating machinery. In the first stage, the collected acoustic signals are prenormalized to eliminate the adverse effects of singular samples, and then the normalized signal is transformed into frequency-domain signal through fast Fourier transform (FFT). In the second stage, the learned features are obtained by training batch-normalized DSF with frequency-domain signals, and then the features are fine-tuned by backpropagation (BP) algorithm. In the third stage, softmax regression is used as a classifier for health condition recognition based on the fine-tuned features. Bearing and planetary gear datasets are used to validate the diagnostic performance of the proposed method. The results show that the proposed DSF model can extract more powerful features and less computing time than other traditional methods.

1. Introduction

Rotating machinery is widely used in automobile engine, wind power equipment, and water turbine generator equipment. A slight mechanical fault may directly affect the operation of machines and cause severe accident. Thus, it is important to ensure the high stability operation of machines [1]. Fault diagnosis technology has been proven to be an effective method to monitor the operating state of equipment in recent years [2–4]. How to extract useful features from massive mechanical data to precisely diagnose the health condition of machines has always been a hot research topic [5–7].

As an effective fault diagnosis method of rotating machinery, increased attention has been paid to research on acoustic signal processing. Given that the acoustic signals are easier to obtain and less costly than vibration signals [8], therefore it has become a trend to apply acoustic signal to fault diagnosis [9]. However, the signal-to-noise ratio of acoustic signals is low, which increases the difficulty of signal denoising and feature extraction. Traditional fault diagnosis

is usually based on the signal processing methods such as short time Fourier transform (STFT) [10], wavelet transform (WT) [11], and empirical mode decomposition (EMD) [12], which overcomes the difficulty of processing acoustic signals and achieves certain results. However, all of the methods mentioned above require empirical knowledge and are time-consuming.

Unsupervised learning may hold potential to overcome the aforementioned weakness in traditional intelligent fault diagnosis method. The basic idea behind unsupervised feature learning is that training artificial intelligence techniques can be viewed as learning a nonlinear function, which transforms the raw data from the original space into a feature space. Hence, unsupervised feature learning is recommended. The purpose of unsupervised feature learning is to adaptively learn effective features from unlabeled data rather than from artificial engineering feature representation [13]. Meanwhile, unsupervised feature learning has been widely applied in the speech recognition [14], face recognition [15], image classification [16], and other fields. Sparse filtering (SF) [17] is an unsupervised two-layer neural

network, which can use its own population sparsity, life sparsity, and high dispersion to learn deep discriminative features automatically. So far, SF has been successfully applied to many scenarios, and its usefulness repeatedly confirmed. Meanwhile, how the implicit hypothesis and constraints of sparse filtering make it suitable for some scenes is confirmed by [18]. Lei et al. [19] first constructively employed SF to bearing fault diagnosis by adopting a two-stage learning method, which greatly reduced human labor and made intelligent fault diagnosis much easier to handle big data. Yang et al. [20] introduced L2-norm regularization to enhance the generalization ability of SF and achieved better classification performance. Qian et al. [21] introduced L1-norm regularization into the cost function of SF and replaced the soft-absolute activation function with logarithm function to prevent overfitting more efficiently. Wang et al. [22] proposed a framework based on sparse filtering which can adaptively extract features from frequency-domain signals.

Although SF can automatically extract useful features from vibration signals, its feature extraction ability remains poor in processing acoustic signals. So we set up the batch-normalized deep sparse filtering (DSF) model to filter the acoustic signal twice to remove the redundant information better, and then the weights are fine-tuned by back propagation (BP) algorithm. Therefore, a novel DSF model is proposed to deal with the acoustic signal in this paper. The main contributions of this paper are summarized as follows:

- (1) Batch normalization is introduced into DSF model, which can reparametrize the hidden layer of DSF, improve the training speed, and accelerate the convergence speed.
- (2) A two-layer batch-normalized DSF is established to filter the acoustic signal twice, and then the weights are adjusted by BP algorithm to get more robust features. The experimental results of gearbox and bearing datasets show that DSF has higher accuracy and faster computing time than other SF models.

The rest of this paper is organized as follows. In Section 2, DSF in deep learning is introduced. In Sections 3 and 4, a deep sparse filtering framework is established, and planetary gear and bearing datasets are investigated using the different SF models, respectively. The conclusion is presented in Section 5.

2. Theoretical Background

2.1. Sparse Filtering. Traditional unsupervised feature learning methods [23] need to adjust multiple parameters to achieve better performance, which is an arduous task. Therefore, Ngiam et al. [17] proposed a method with only one parameter called SF to solve this weakness. As a simple and efficient unsupervised learning method, SF only focuses on the sparse distribution of data features. The structure of SF is shown in Figure 1. Its input and output are collected dataset and the learned features, respectively.

First, acoustic signals are collected from each health condition and combined into training set $\{x^i\}_{i=1}^K$, where $x^i =$

$\mathfrak{R}^{L \times 1}$ is a sample containing L data points and K is the sample number. Second, weight matrix $W \in \mathfrak{R}^{L \times P}$ is obtained by $\{x^i\}_{i=1}^K$ training with SF model. Finally, the SF model can learn the corresponding feature set $\{f^i\}_{i=1}^K$ by weight matrix W , where $f^i \in \mathfrak{R}^{P \times 1}$ denotes the feature vector with P learned features. The features of each sample can be calculated as

$$f_j^i = W_j^T x^i, \quad (1)$$

where f_j^i corresponds to the j th feature of the i th sample. Features f_j^i make up a feature matrix; each row of the feature matrix is normalized through its L2 norm

$$\tilde{f}_j = \frac{f_j}{\|f_j\|_2}. \quad (2)$$

Then, each column is normalized by its L2 norm, and the features are mapped into the unit L2-ball:

$$\tilde{f}^i = \frac{\tilde{f}^i}{\|\tilde{f}^i\|_2}. \quad (3)$$

Finally, the optimization features can be obtained by L1 penalty; the cost function of SF is given as

$$\underset{W}{\text{minimize}} \sum_{i=1}^K \|\tilde{f}^i\|_1 = \sum_{i=1}^K \left\| \frac{\tilde{f}^i}{\|\tilde{f}^i\|_2} \right\|_1. \quad (4)$$

2.2. Deep Sparse Filtering. The standard SF model is a simple two-layer neural network. In this paper, DSF model is designed by layer-by-layer unsupervised learning. Specifically, the output features of the first SF are used as the input of the second SF to extract the feature layer by layer, and then the weights are reversely fine-tuned using the BP algorithm. In addition, we use a rectified linear unit function (ReLU) [24] as activation function. The output features of n th hidden layer can be calculated as

$$h_n = \text{ReLU}(W_n^T x_{n-1}), \quad (5)$$

where W_n represents the weight matrix between n th hidden layer and $(n-1)$ th hidden layer. Each SF layer is trained by solving the minimization problem, as

$$\underset{W}{\text{minimize}} \sum_{i=1}^k \|\hat{h}_n\|_1 = \sum_{i=1}^k \left\| \frac{\tilde{h}_n}{\|\tilde{h}_n\|_2} \right\|_1, \quad (6)$$

where k is the number of training samples in n th layer.

At the same time, the batch normalization is introduced to optimize the DSF. Batch normalization can reparametrize almost deep networks in an elegant way. The procedure is able to be used in every activation layer without parameter adjustment. For a layer with n -dimensional input $x = (x_1, \dots, x_k)$, to improve the training and reduce the internal covariate shift, two necessary simplifications are taken by batch normalization.

Firstly, each scalar feature is normalized independently by making its own zero mean and unit variance.

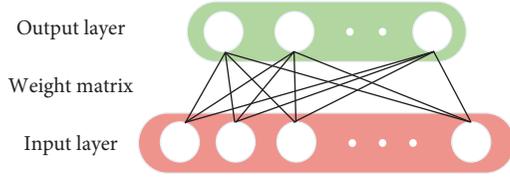


FIGURE 1: SF structure.

$$\widehat{x}_i = \frac{x_i - E[x_i]}{\sqrt{\text{Var}[x_i]}} \quad (7)$$

where $E[x_i]$ is the mean of each unit and $\sqrt{\text{Var}[x_i]}$ denotes the standard deviation.

However, the simply normalization of each input in a layer still can change the representation of the layer. So two parameters γ_i and β_i are employed for each activation x_i , which aim to scale and shift the normalized value:

$$f_i = \gamma_i \widehat{x}_i + \beta_i. \quad (8)$$

γ_i and β_i are learned along with the raw model parameters and restore the representation power of the network. Note that the raw activations can be recovered by setting $\gamma_i = \sqrt{\text{Var}[x_i]}$ and $\beta_i = E[x_i]$. In this case, the steady distribution of activation values can be guaranteed during each training.

Here, we apply the batch normalization immediately before the activation layers of SF. So equation (5) is replaced with

$$h_n = f(\text{BNW}x). \quad (9)$$

Therefore, BN transform introduces normalized activations into the network and ensures the layers can continue learning on input distributions that reduce the influence of internal covariate shift, so that an easy starting condition can be constructed for training and further accelerating the training.

The number of layers can be selected according to different task requirements. It is generally believed that increasing the number of hidden layers can reduce the network error and improve the accuracy, but it also complicates the network, thus increasing the network training time and the tendency of overfitting. So it is actually a trade-off choice in application of the proposed method. In this paper, we choose two layers of DSF to extract the feature of acoustic signal, which can not only ensure less computation but also extract deeper features. Figure 2 shows the schematic of DSF. The acoustic signal datasets are used to train the first SF layer and subsequently the second SF layer. Firstly, the output batch-normalized features of the first SF layer are used as the input features of the second SF layer, and then the softmax regression is connected to the last layer of DSF as the classification layer. Finally, BP algorithm is used for the reverse weight fine-tuning.

3. Intelligent Fault Diagnosis Framework Based on DSF

The proposed fault diagnosis method mainly consists of three stages as shown in Figure 3. In the first stage, the collected time-domain acoustic signals are pre-normalized to eliminate the adverse effects of singular samples. Then, the normalized time-domain signal is transformed into frequency-domain signal through FFT. In the second stage, the weight matrix W is obtained by training batch-normalized DSF with frequency-domain signals, and then the W is fine-tuned by BP algorithm. Finally, the optimized W is used to learn the deep discriminative features from the original frequency-domain signals. In the third stage, softmax regression is used as a classifier for health condition recognition through the learned features:

- (1) Training data collection: the acoustic time-domain signals collected from rotating machinery under different health conditions are divided into K samples to form the training dataset $\{x^i, y^i\}_{i=1}^K$, where $x^i \in \mathfrak{R}^{N \times 1}$ denotes each sample containing N time-domain points and y^i denotes the health condition label of the i th sample.
- (2) Training data processing: training set is rewritten as a matrix form $\mathbf{X} \in \mathfrak{R}^{N \times K}$. Before training the DSF model, each column of training set \mathbf{X} is first normalized by its l_2 -norm as follows:

$$\widetilde{x}^j = \frac{x^j}{\|x^j\|_2}. \quad (10)$$

Then, prenormalization training dataset $\{x_n^i\}_{i=1}^K$ is transformed into training dataset $\{t^i\}_{i=1}^K$ by FFT, where $t^i \in \mathfrak{R}^{N_{in} \times 1}$ denotes each sample containing N_{in} Fourier coefficients. N_{in} represents the input dimension of DSF, and N_{out} is the output dimension. The training set $\{t^i\}_{i=1}^K$ can be further written as a matrix $\mathbf{S} \in \mathfrak{R}^{N_{in} \times K}$ for simplicity.

- (3) DSF model training: firstly, the obtained \mathbf{S} is inputted to the batch-normalized DSF model for the training of weight matrix \mathbf{W} . Then, the BP algorithm is combined with the corresponding sample labels to fine-tune the \mathbf{W} of DSF.
- (4) Model testing: remaining samples are used as testing samples to test the accuracy of the trained DSF model.

4. Experiments

4.1. Case Study 1: Rolling Bearing Fault Diagnosis

4.1.1. Data Description. In this section, the acoustic signals of bearing are collected from the specially designed test bench to validate the diagnosis performance of proposed DSF method. As shown in Figure 4(a), the test bench includes a motor, three shaft couplings, a bearing seat, a gearbox, and a brake. As shown in Figure 4(b), the collected dataset includes 9 health conditions: normal condition (NC),

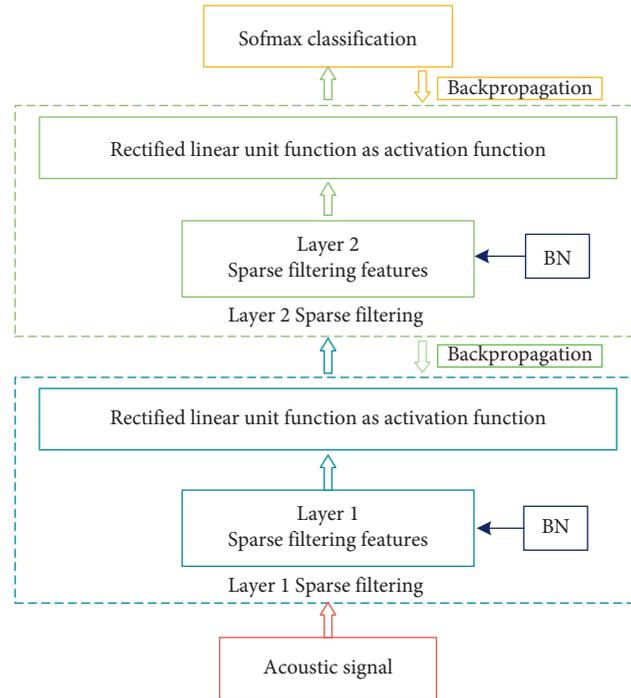


FIGURE 2: Schematic of DSF.

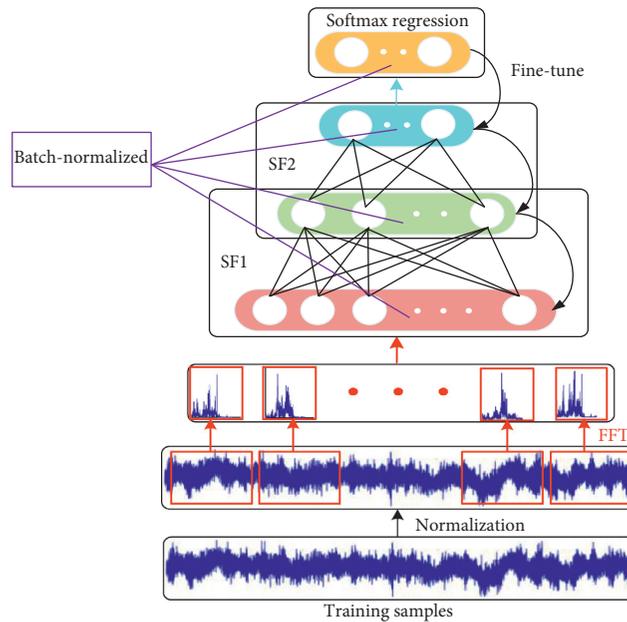


FIGURE 3: Intelligent fault diagnosis framework based on DSF.

outer race fault 0.2 mm (OF0.2), outer race fault 0.4 mm (OF0.4), inner race fault 0.2 mm (IF0.2), inner race fault 0.4 mm (IF0.4), roller fault 0.2 mm (RF0.2), roller fault 0.4 mm (RF0.4), roller fault 0.2 mm and outer race fault 0.2 mm (ROF0.2), and roller fault 0.4 mm and outer race fault 0.4 mm (ROF0.4). The sampling frequency of the acoustic sensor is 12.8 kHz and the rotating speed is 1300 r/min. 200 samples are collected from each health condition, and a total of 1800 samples are obtained. Each sample

contains 2400 time-domain points, and 1200 frequency-domain points are obtained by FFT.

One sample is randomly selected from each health condition to show the acoustic signal details. The time-domain and corresponding frequency-domain waveforms of the samples are shown in Figure 5. It can be seen that it is arduous to distinguish different health conditions artificially, and the huge amount of data also increases the difficulty of feature extraction. Therefore, the DSF model is proposed to

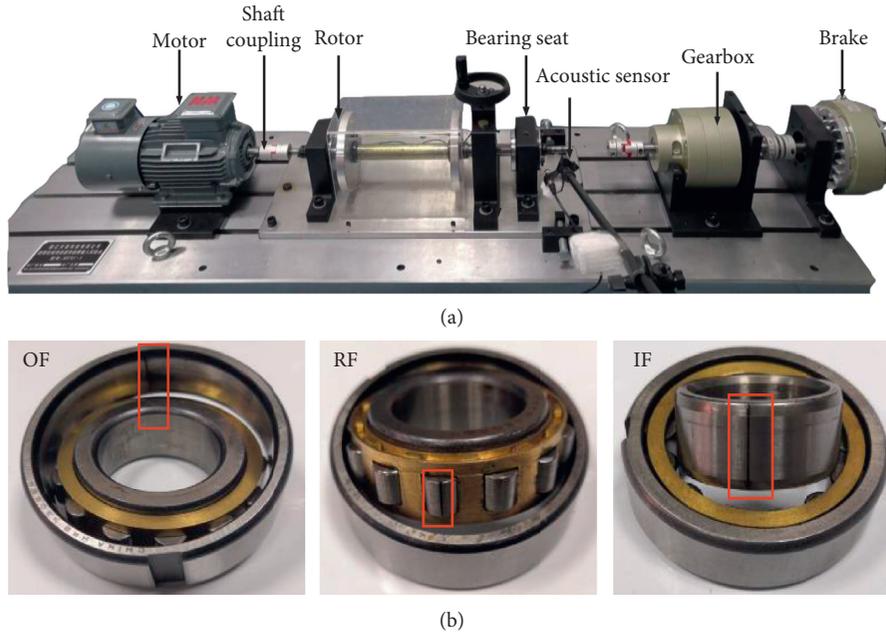


FIGURE 4: (a) Bearing fault test rig and (b) three fault bearings.

automatically extract the feature of acoustic signal and conduct precise fault classification.

4.1.2. Results and Analysis. The frequency-domain signal is used as the input of DSF model, and the output dimensions of the two SF layers are set to 800 and 400, respectively. The number of outputs of softmax classification is 9. Therefore, the structure of the DSF model is 1200-800-400-9. Subsequently, we investigate the effect of iteration number. Randomly select 5% samples for training and the diagnostic accuracies using different iteration number are displayed in Figure 6. Since the increasing of the accuracies is not obvious after number of iterations exceeds 40, we choose 40 as the iteration number of DSF. Meanwhile, the iteration number of the BP algorithm is 50, and the batch size is 30. In order to show the superiority of DSF model, standard SF [19], L1 regularized sparse filtering (L1-SF) [21], and L2 regularized sparse filtering (L2-SF) [20] are used as comparison methods. The output dimension of the three comparison methods is set as 1200, the number of iterations is 100, and the regularization parameter is $1E-5$. 20 trails are conducted for each experiment to reduce the influence of randomness. The computing platform is a PC with an I5-4210M CPU and 8 GB RAM.

The diagnosis results of different numbers of training samples using the proposed DSF model are shown in Figure 7. It is obvious that the accuracy and computing time increase with the rise of the training sample number. It can be seen from the figure that the DSF model with only 5% of the training samples can achieve average testing accuracy of $98.15\% \pm 0.33\%$, indicating that the proposed method can diagnose 9 health conditions in the absence of training samples. When the number of samples increases to 10%, the

average test accuracy reaches $99.92\% \pm 0.027\%$, and the average computing time is 14.9s. Therefore, in the following experiments, 10% of the samples were used for training.

The diagnosis results of the four methods are shown in Figure 8. It is certain that the DSF model has the highest average testing accuracy (99.93%) and the lowest standard deviation (0.027%) among all the methods. It can be seen from the figure that the average accuracy of the standard SF is $89.05\% \pm 1.39\%$, which is the worst among the methods. The testing accuracies of L1-SF and L2-SF are $90.45\% \pm 1.09\%$ and $91.63\% \pm 0.77\%$, respectively, which are slightly higher than those of SF. It is worth mentioning that the proposed DSF model computing time is 14.9s. By contrast, the average computing time of SF, L1-SF, and L2-SF is about 100s. This finding indicates that the DSF method can better overcome the difficulty of extracting the acoustic signal features and achieve the highest accuracy and least computing time among the four methods in terms of diagnosing bearing fault types.

In order to better present the superiority of DSF, here, we make a detailed comparison between our method and other several classical methods by using the same bearing dataset, as summarized in Table 1. In Method 1, ensemble empirical mode decomposition (EMMD) [25] was employed to extract features and then the features were classified by an optimized SVM. It achieved 96.67% testing accuracy on the bearing dataset. In Method 2, Jia et al. [26] constructed SAE based deep network utilizing frequency spectra as inputs to diagnosis and 99.68% testing accuracy is obtained. In Method 3, the frequency spectra are also used as inputs of Back Propagation Neural Networks (BPNN) and the diagnosis accuracy is 73.74%. In Method 4, Xie et al. [27] proposed feature extraction algorithm based on empirical mode decomposition (EMD) and convolutional neural network

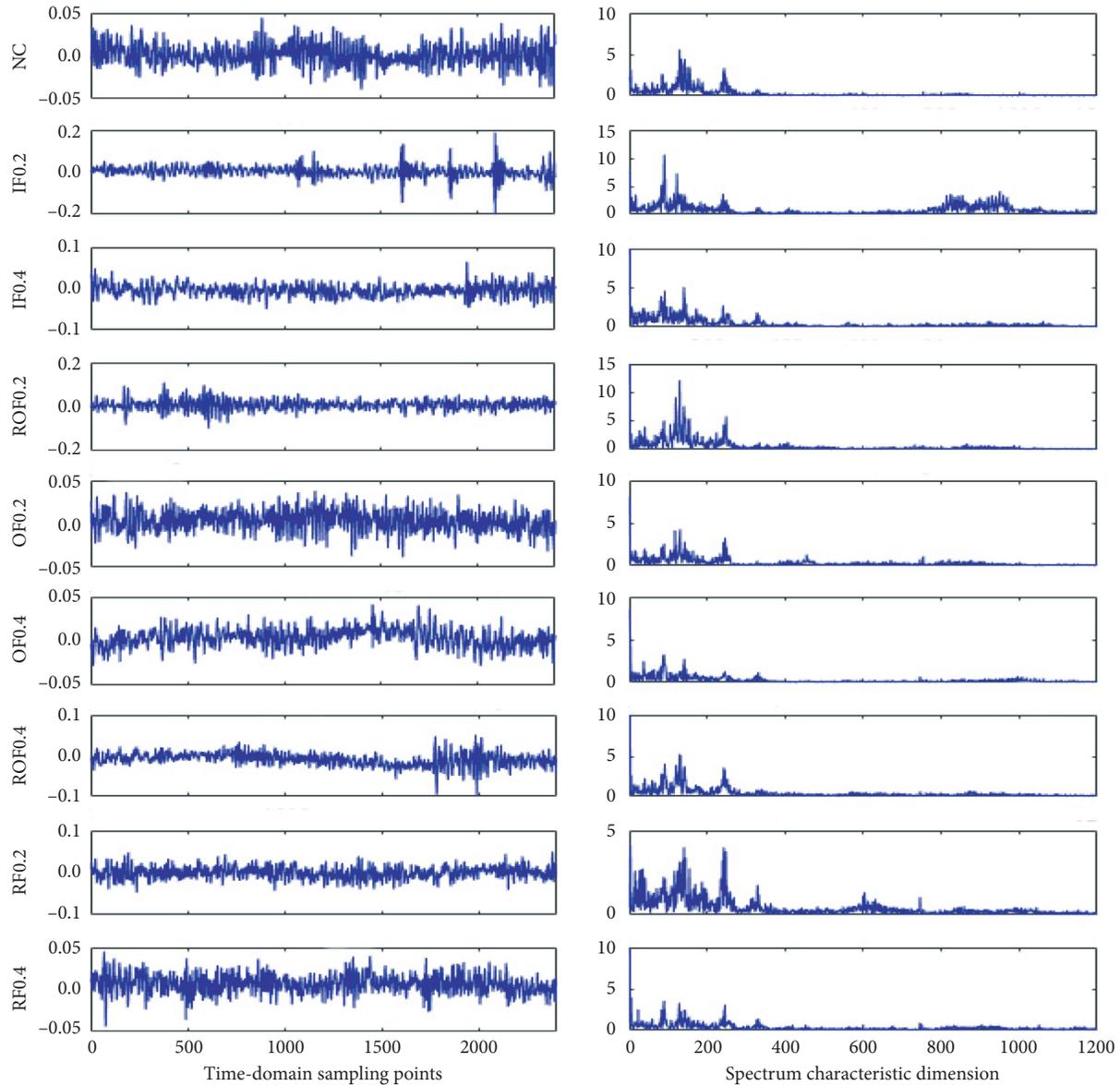


FIGURE 5: Time- and frequency-domain waveforms of bearing signals under the nine health conditions.

(CNN) techniques and obtained 99.75% testing accuracy. In Method 5, the proposed method achieves the best testing accuracy of 99.93% when classifying ten different fault conditions, which outperforms all compared approaches.

To show the details of the diagnostic results of the four methods, the confusion matrixes on the bearing dataset are presented in Figure 9. It can be seen from Figures 9(a) and 9(b) that the classification results of SF and L1-SF are unsatisfactory. The concurrent faults such as ROF0.2 and ROF0.4 are not well distinguished, and the single faults such as RF0.4 and OF0.4 are not perfectly distinguished. The fault classification performance of L2-SF is slightly better than that of SF and L1-SF as shown in Figure 9(c), but it cannot distinguish different health conditions with high accuracy, which means that concurrent faults increase the difficulty of fault classification. As shown in Figure 9(d), the proposed

DSF model can distinguish not only single faults but also concurrent faults perfectly, which shows that the proposed method can better extract the deep features of acoustic signal.

4.2. Case Study 2: Planetary Gear Fault Diagnosis

4.2.1. Data Description. The gear fault signals are measured from the gearbox of the test bench as shown in Figure 4(a). The collected dataset contains one normal condition (NC) and four kinds of mechanical faults including sun wheel crack (WC), sun wheel pit (WP), pinion crack (PC), and pinion pit (PP), as shown in Figure 10. The gear speed is 2600 r/min and the sampling frequency is 10.24 kHz. 300 samples were collected for each health condition and each

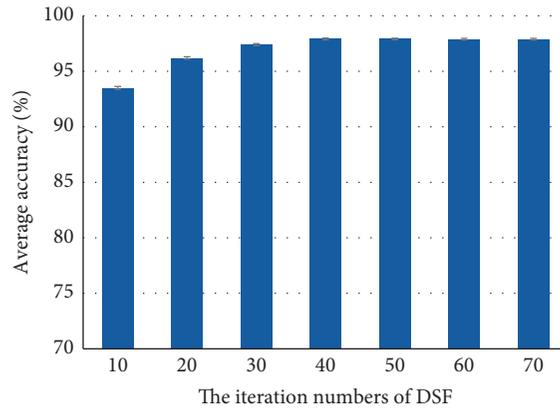


FIGURE 6: Average accuracy of different iteration number of DSF model.

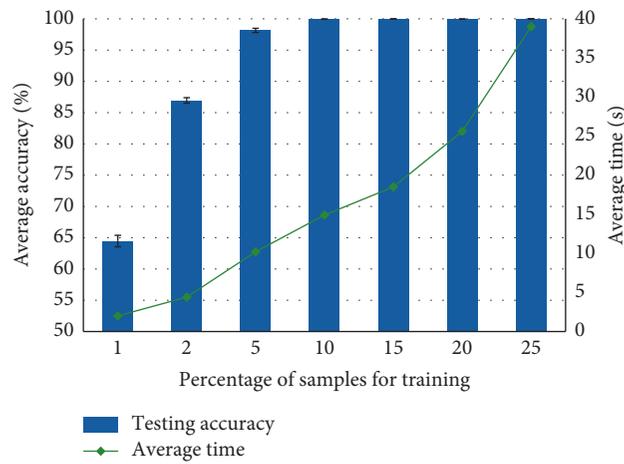


FIGURE 7: The diagnosis results of the DSF method using different percentage of training samples.

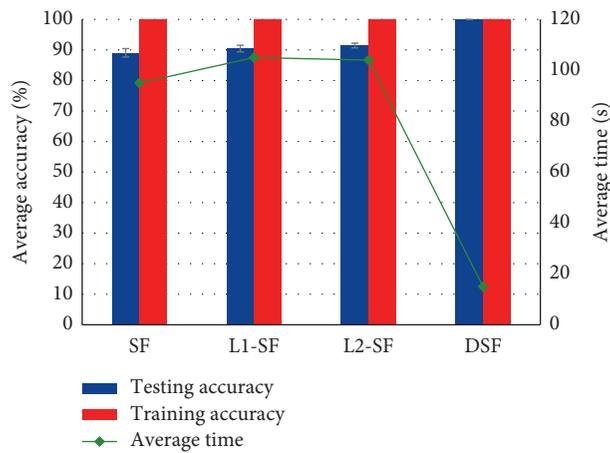


FIGURE 8: Comparison of the average diagnostic accuracies of four SF models.

sample contains 1600 data points. Each sample gets 800 frequency-domain points through FFT as the input of the model.

4.2.2. Results and Analysis. 10% of gear samples were randomly selected to train DSF model. After testing, we set 600 and 100 as the output dimension of the two SF layers.

TABLE 1: Performance comparison between the proposed DSF method and some classical method.

Method	Description	Training samples (%)	Testing accuracy (%)
1	EMMD features + SVM	25	97.04
2	SAE + DNN	50	99.68
3	BPNN + frequency	50	73.74
4	EMD features + CNN	75	99.75
5	DSF	10	99.93

True labels	NC	IF0.2	IF0.4	ROF0.2	OF0.2	OF0.4	ROF0.4	RF0.2	RF0.4
	159	0	0	0	0	0	0	0	21
	0	166	2	0	0	0	7	0	5
	0	0	173	0	0	4	1	0	2
	4	0	1	130	9	6	5	7	18
	0	0	0	0	174	0	1	0	5
	1	0	0	0	0	160	3	0	16
	0	0	0	0	1	2	156	1	20
	0	0	0	0	1	1	15	156	7
	0	0	0	0	0	0	1	0	179

(a)

True labels	NC	IF0.2	IF0.4	ROF0.2	OF0.2	OF0.4	ROF0.4	RF0.2	RF0.4
	168	0	0	0	1	10	1	0	0
	0	169	1	0	5	0	5	0	0
	0	0	150	0	0	23	6	0	1
	1	0	0	131	36	7	1	0	4
	0	0	0	0	180	0	0	0	0
	0	0	1	0	0	179	0	0	0
	0	0	0	0	10	9	160	0	1
	0	0	0	0	7	11	7	149	6
	0	0	0	0	0	10	1	0	169

(b)

True labels	NC	IF0.2	IF0.4	ROF0.2	OF0.2	OF0.4	ROF0.4	RF0.2	RF0.4
	174	0	0	0	6	0	0	0	0
	0	174	3	0	1	0	2	0	0
	0	0	172	0	0	7	1	0	0
	2	0	0	143	31	0	2	0	2
	0	0	0	0	180	0	0	0	0
	0	0	0	0	0	179	1	0	0
	0	0	0	0	22	7	148	0	3
	0	0	0	0	6	2	18	152	2
	0	0	0	0	5	1	3	0	171

(c)

True labels	NC	IF0.2	IF0.4	ROF0.2	OF0.2	OF0.4	ROF0.4	RF0.2	RF0.4
	180	0	0	0	0	0	0	0	0
	0	180	0	0	0	0	0	0	0
	0	0	180	0	0	0	0	0	0
	0	0	0	180	0	0	0	0	0
	0	0	0	0	180	0	0	0	0
	0	0	0	0	0	180	0	0	0
	0	0	0	0	0	1	179	0	0
	0	0	0	0	0	0	0	180	0
	0	0	0	0	0	0	0	0	180

(d)

FIGURE 9: Confusion matrixes for bearing dataset: (a) SF, (b) L1-SF, (c) L2-SF, and (d) DSF.

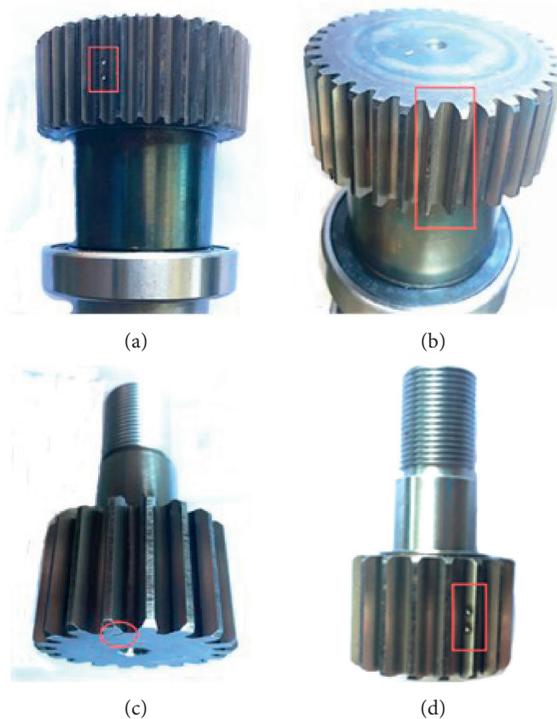


FIGURE 10: Four kinds of gear fault feature diagrams.

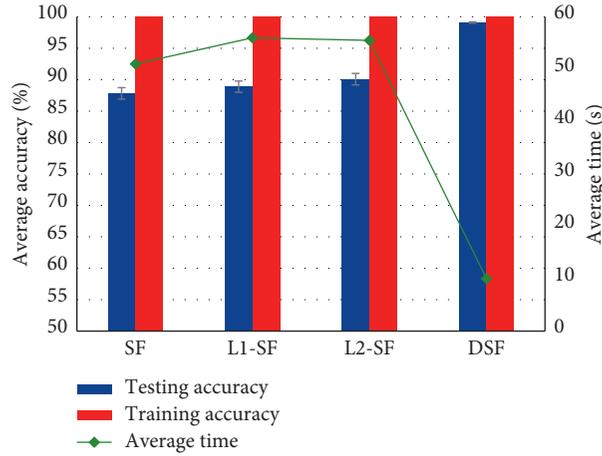


FIGURE 11: Comparison of the average diagnostic accuracies of four SF models.

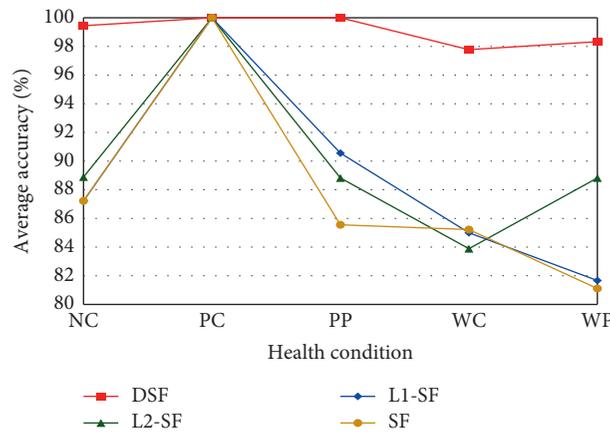


FIGURE 12: Comparison of average diagnosis accuracies of each health condition.

The numbers of iterations of the two SF layers are all 20, and the output dimension of softmax is 5. The iteration number of the BP algorithm is 50, and the batch size is 20. The output dimension of the three comparison methods is set as 800, and other parameters set are the same as Case 1.

Each experiment is repeated 20 times to reduce the effects of randomness. The gear fault diagnosis results of the four methods are shown in Figure 11. Specifically, the training accuracy of the four methods is 100%. The performance of SF model is the most unsatisfactory, and the testing accuracy is $87.82\% \pm 0.91\%$. The testing accuracies of L1-SF and L2-SF are 88% and 90%, respectively. By contrast, the proposed DSF model achieved the highest testing accuracy of $99.11\% \pm 0.11\%$. Meanwhile, the average computing time of DSF model is 9.58 s, and the computing time of three comparison methods is about 5 times that of DSF model. In conclusion, the proposed DSF method can achieve the highest accuracy and robustness in acoustic signal fault diagnosis.

The average accuracies of the five health conditions are shown in Figure 12. It can be determined that the four

methods can precisely diagnose the health condition of PC. However, the three comparison methods have lower testing accuracies for health conditions NC, PP, WC, and WP. In contrast, DSF model can overcome this shortcoming and accurately diagnose the five health conditions.

To confirm the performance of the proposed DSF model, the t-distributed stochastic neighbor embedding (t-SNE) is applied to obtain the first two dimensions of learned features and the results are shown in Figure 13. It can be seen from Figure 13(a) that the features of the same health condition are well clustered. However, ten points of WC were misclassified to PP, and five points of NC were misclassified to WC, which explains the phenomenon that this method obtains low diagnosis accuracy for the health conditions of PP and WC. In Figures 13(b) and 13(c), the interval between NC, WC, and NC is not obvious; there are also more points that are misclassified. It is worth noting that each feature cluster of the DSF method is separated, and only four points of WC were mistakenly assigned to NC, and three points of WP were mistakenly assigned to PC, which means that

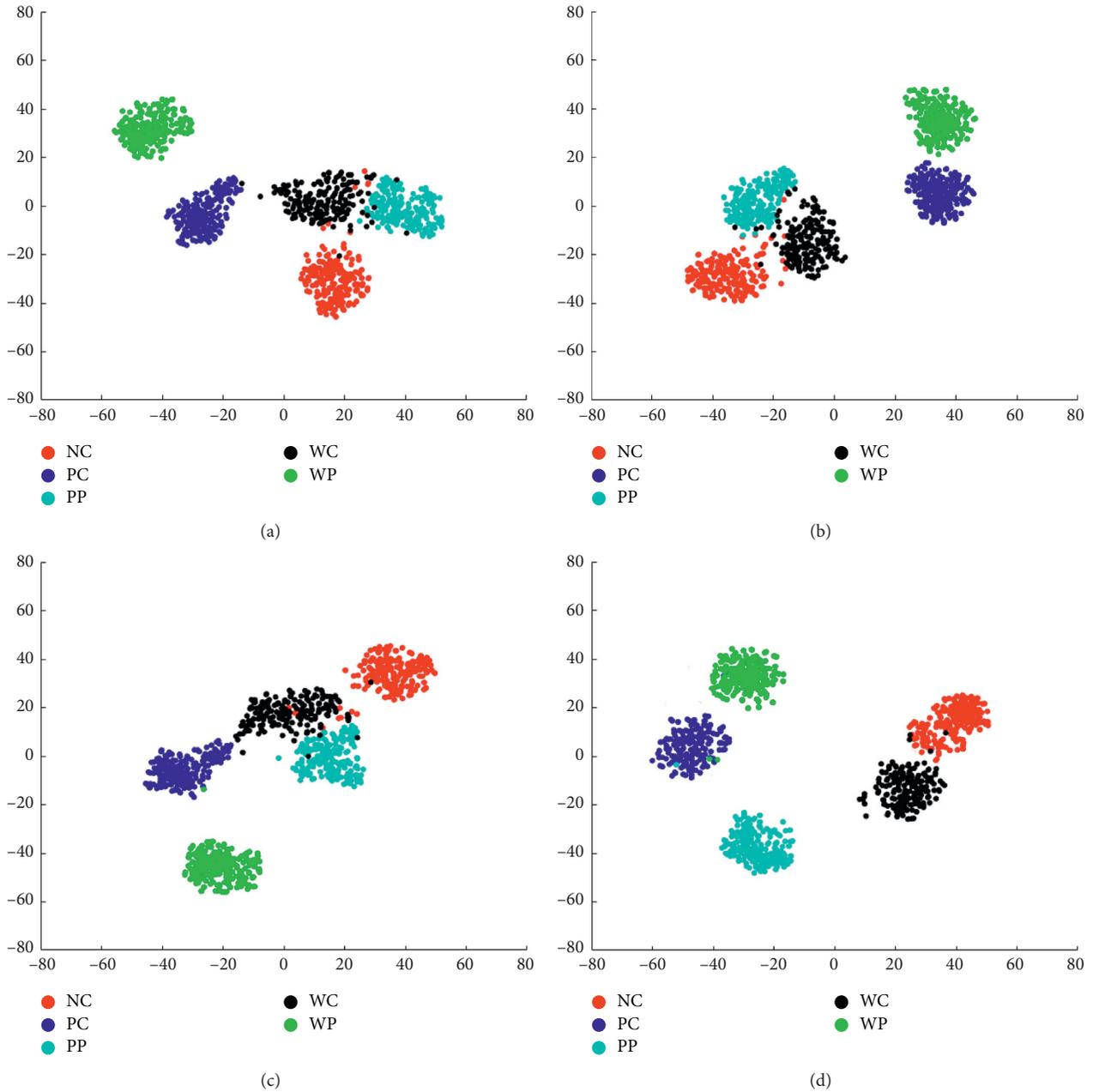


FIGURE 13: First two dimensions of learned features of the four methods: (a) SF, (b) L1-SF, (c) L2-SF, and (d) DSF.

the extracted features of DSF model are more recognizable.

5. Conclusion

In this paper, a batch-normalized DSF model is proposed to process acoustic signals for fault diagnosis. Two SF models are stacked to extract the deep features of acoustic signals, and the optimal weight is obtained by fine-tuning process of BP algorithm. The experiment results of bearing and gearbox dataset show that the DSF model can also achieve high test accuracy in the case of insufficient training samples. Meanwhile, compared with other SF models, DSF can get the

highest accuracy and the least computing time, which shows that the proposed method is more efficient in feature extraction.

Data Availability

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

Conflicts of Interest

The authors declare no conflicts of interest regarding this publication.

Acknowledgments

This work was supported by China Postdoctoral Science Foundation (2019M662399) and the Project of Shandong Province Higher Educational Young Innovative Talent Introduction and Cultivation Team (Performance Enhancement of Deep Coal Mining Equipment).

References

- [1] C. K. Tan, P. Irving, and D. Mba, "A comparative experimental study on the diagnostic and prognostic capabilities of acoustics emission, vibration and spectrometric oil analysis for spur gears," *Mechanical Systems and Signal Processing*, vol. 21, no. 1, pp. 208–233, 2007.
- [2] C. Shen, Y. Qi, and J. Wang, "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.
- [3] H. Liu, L. Li, and J. Ma, "Rolling bearing fault diagnosis based on STFT-deep learning and sound signals," *Shock and Vibration*, vol. 2016, Article ID 6127479, 12 pages, 2016.
- [4] J. Wang, S. Li, B. Han et al., "Construction of a batch-normalized autoencoder network and its application in mechanical intelligent fault diagnosis," *Measurement Science and Technology*, vol. 30, no. 1, p. 15106, 2019.
- [5] 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.
- [6] S. A. Khan and J. M. Kim, "Automated bearing fault diagnosis using 2D analysis of vibration acceleration signals under variable speed conditions," *Shock and Vibration*, vol. 2016, Article ID 8729572, 11 pages, 2016.
- [7] X. Jiang and J. Shi, "Non-dominated Solution Set Based on Time-Frequency Infograms for Local Damage Detection of Rotating Machines," *Isa Transactions*, New York, NY, USA, 2019.
- [8] K. Shibata, A. Takahashi, and T. Shirai, "Fault diagnosis of rotating machinery through visualisation of sound signals," *Mechanical Systems and Signal Processing*, vol. 14, no. 2, pp. 229–241, 2000.
- [9] F. Elasha, M. Greaves, D. Mba, and D. Fang, "A comparative study of the effectiveness of vibration and acoustic emission in diagnosing a defective bearing in a planetary gearbox," *Applied Acoustics*, vol. 115, no. 2, pp. 181–195, 2017.
- [10] H. Zhang, S. Lu, Q. He, and F. Kong, "Multi-bearing defect detection with trackside acoustic signal based on a pseudo time-frequency analysis and Dopplerlet filter," *Mechanical Systems and Signal Processing*, vol. 71, pp. 176–200, 2016.
- [11] Z. Gao, J. Lin, X. Wang, and X. Xu, "Bearing fault detection based on empirical wavelet transform and correlated kurtosis by acoustic emission," *Materials*, vol. 10, no. 6, p. 571, 2017.
- [12] X. Zhang, Y. Wang, and K. Wang, "Rail crack detection based on the adaptive noise cancellation method of emd at high speed," in *Proceedings of 2017 IEEE International Instrumentation and Measurement Technology Conference*, IEEE, Turin, Italy, May 2017.
- [13] A. Coates, A. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 215–223, Sicily, Italy, June 2011.
- [14] K. Noda, Y. Yamaguchi, K. Nakadai, H. G. Okuno, and T. Ogata, "Audio-visual speech recognition using deep learning," *Applied Intelligence*, vol. 42, no. 4, pp. 722–737, 2015.
- [15] X. Zhu, Z. H. Liu, F. H. Yang, Shi, G. Qi, and S. Z. Li, "Large-scale bisample learning on ID versus spot face recognition," *International Journal of Computer Vision*, vol. 127, no. 6–7, pp. 684–700, 2019.
- [16] L. Yi, B. J. Liu, K. M. Yang, Sun, and S. Abbas, "A deep multi-modal CNN for multi-instance multi-label image classification," *IEEE Transactions on Image Processing*, vol. 27, no. 12, pp. 6025–6038, 2018.
- [17] J. Sun, Z. Chen, S. A. Bhaskar, P. W. Koh, and A. Y. Ng, "Sparse filtering," in *Advances in Neural Information Processing Systems*, pp. 1125–1133, 2011.
- [18] F. M. Zennaro and K. Chen, "Towards understanding sparse filtering: a theoretical perspective," *Neural Networks*, vol. 98, pp. 154–177, 2018.
- [19] 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.
- [20] Z. Yang, L. Jin, D. Tao, S. Zhang, and X. Zhang, "Single-layer unsupervised feature learning with l2 regularized sparse filtering," in *Proceedings of IEEE China Summit & International Conference on Signal and Information Processing*, pp. 475–479, Xi'an, China, July 2014.
- [21] W. Qian, S. Li, J. Wang, Z. An, and X. Jiang, "An intelligent fault diagnosis method of rotating machinery using L1-regularized sparse filtering," *Journal of Vibroengineering*, vol. 20, no. 8, pp. 2839–2854, 2018.
- [22] J. Wang, S. Li, Y. Xin, and Z. An, "Gear fault intelligent diagnosis based on frequency-domain feature extraction," *Journal of Vibration Engineering & Technologies*, vol. 7, no. 2, pp. 159–166, 2019.
- [23] C. Hou, F. Nie, and X. Li, "Joint embedding learning and sparse regression: a framework for unsupervised feature selection," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 44, no. 6, pp. 793–804, 2014.
- [24] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines vinod nair," in *Proceedings of the 27th International Conference on Machine Learning*, pp. 807–814, Haifa, Israel, June 2010.
- [25] X. Zhang and J. Zhou, "Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines," *Mechanical Systems and Signal Processing*, vol. 41, no. 1–2, pp. 127–140, 2013.
- [26] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mechanical Systems and Signal Processing*, vol. 73, pp. 303–315, 2016.
- [27] Y. Xie and T. Zhang, "Fault diagnosis for rotating machinery based on convolutional neural network and empirical mode decomposition," *Shock and Vibration*, vol. 2017, Article ID 3084197, 12 pages, 2017.