

## Research Article

# Parallel Cross-Sparse Filtering Networks and Its Application on Fault Diagnosis of Rotating Machinery

**Shan Wang** <sup>1</sup>, **Baokun Han** <sup>1</sup>, **Huaiqian Bao** <sup>1</sup>, **Jinrui Wang** <sup>1</sup>,  
and **Zongzhen Zhang** <sup>1,2</sup>

<sup>1</sup>College of Mechanical and Electronic Engineering, Shandong University of Science and Technology, Qingdao 266590, China

<sup>2</sup>College of Energy and Power Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Correspondence should be addressed to Zongzhen Zhang; zhzz18@126.com

Received 29 November 2021; Accepted 11 January 2022; Published 7 February 2022

Academic Editor: Haidong Shao

Copyright © 2022 Shan Wang 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.

Intelligent diagnosis method has become a new focus for researchers, which can get rid of the dependence of diagnostic experience and prior knowledge. However, in practical application, to deal with the new fault type of mechanical equipment, the number of fault labels of the diagnosis model needs to be increased. We must retrain the whole training model, which is a time-consuming process. To solve this problem, higher requirements are put forward for the generalization ability and universality of the algorithm. In view of the feature extraction advantages of cross-sparse filtering (Cr-SF), which can be regarded as an unsupervised minimum entropy learning method using the maximization of the proxy of sparsity, this paper proposed a parallel network based on Cr-SF. The feature extraction process of each sample is independent, and the feature extraction and classifier training process are separated. Therefore, the most prominent advantage of the proposed method is that when a new fault occurs, it only needs to extract the feature of the new fault separately and then input it to the classifier at the last layer for training. The experimental results show that the proposed method can obtain high accuracy and stability and can significantly improve the adaptability of intelligent fault diagnosis in practical application.

## 1. Introduction

As the key parts of mechanical transmission, bearings and gears are prone to failure during the runtime, which may reduce the working efficiency and even cause accidents and disasters [1]. Therefore, accurate early warning and corresponding maintenance measures when bearing and gear failure are of great significance to ensure the safe operation of mechanical equipment [2]. With the rapid development of machine learning theory, intelligent fault diagnosis method of rotating machinery has become an important topic in the area of health monitoring of mechanical equipment [3, 4].

Recently, deep learning-based intelligent rotating machinery fault diagnosis, which can automatically extract the features from original data, has achieved remarkable success [5–7]. These methods have great performance to overcome the inherent disadvantages of traditional machine learning methods, such as Support Vector Machine (SVM),

Artificial Neural Network (ANN), and Principal Component Analysis (PCA). Jia et al. [8] proposed a novel intelligent method based on DNNs to overcome the shortcomings of the traditional methods, which can adaptively learn the fault features and obtain superior diagnostic accuracy and robustness. Wang et al. [9] improve the computation efficiency of feature extraction using batch normalization based Stacked Autoencoders (SAEs). Li et al. [10] enhanced the feature learning ability using S-transform (ST) algorithm and connected convolutional neural networks (CNN). Shao et al. introduced Gaussian visible units to the electric locomotive bearing fault diagnosis based on Convolutional Deep Belief Network (CDBN) [11, 12]. The literature [13] presented a novel deep autoencoder loss function based on maximum corr-entropy to eliminate the effects of background noise. To estimate the irregularity of the collected time series, Li et al. [14] proposed a multiscale symbolic Lempel-Ziv- (MSLZ-) based intelligent fault diagnosis method and successfully

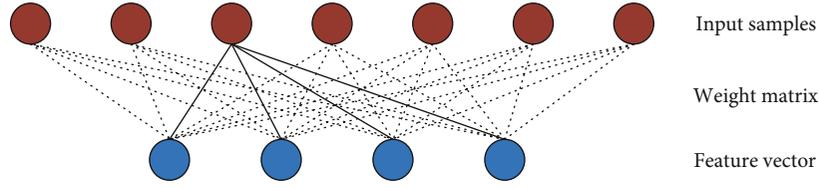


FIGURE 1: Architecture of sparse filtering.

applied it the multiple fault diagnosis of railway vehicle systems. In [15], a novel feature learning method named multi-scale symbolic dynamic entropy (MSDE) was firstly proposed and then combined with transfer learning to obtain the mapping matrix and achieve the novel cross-domain intelligent fault diagnosis. In [16], the MSDE-based fault diagnosis method was applied to the planetary gearboxes and shows superior advantages in terms of computation efficiency and robustness. Wang et al. [17] proposed subdomain adaptation transfer learning network to reduce adaptively marginal and conditional distribution biases. Jia et al. [18] proposed a partial transfer fault diagnosis model based on a weighted subdomain adaptation network (WSAN), which focuses on the distribution of local feature while aligning the global distribution.

However, deep learning requires training all types of fault samples simultaneously. Therefore, if new fault types appear in the application process, we need to retrain and reoptimize a lot of parameters simultaneously with a large number of training data. This is a time-consuming process and will affect the real-time application of the monitoring system. The main reason for this problem is that deep learning requires all label data to participate in feature extraction and classification training. Lei et al. [19] applied sparse filtering and softmax classifier to the intelligent diagnosis bearing faults, which is a two-stage model, and feature learning and classification training are independent of each other. Although the feature learning and classifier training of these methods are separated, the feature learning is still carried out simultaneously for all fault types. Parallel network structure is an effective way to solve this problem, in which each fault can be trained separately. In this way, the feature learning process is an independent process for each type of fault condition. In [20], a concurrent convolution neural network (C-CNN) composed of multiple branches was proposed for bearing fault diagnosis, in which the convolutional layer of different branches selects the kernels with different scales in same level. In view of gearbox structure and operating condition, Guo et al. [21] established reinforced input-based multitask parallel convolutional neural network for coupling fault diagnosis of gearbox that has parallel submultiple classifiers and convolutional neural networks. This method is to overcome the problem that all kinds of shared features of multiparts cannot be adequately extracted simultaneously. The independent training of single fault sample cannot be realized.

The independent training of samples puts forward higher requirements for the feature extraction algorithm. In an unsupervised learning method, sparse representation is the core principle [22]. Data sparsity corresponds to information entropy. The sparser the data is, the smaller the

entropy is. In [23], Zennaro and Ken's study a thorough theoretical analysis of SF and the corresponding performance and proved that the SF works by explicitly maximizing the entropy of the learned representations through the maximization of the proxy of sparsity. Zhang et al. introduced generalized normalization to the sparse filtering and discussed the lifetime and population sparsity [24]. In [25, 26], Intrinsic Component Filtering (ICF) and Cr-SF, the improved variants of standard SF, were proposed for the intelligent fault diagnosis, weak feature extraction, and compound separation. Cr-SF is a variant of SF. Therefore, Cr-SF can be regarded as an unsupervised minimum entropy feature learning method, in which the entropy of the extracted features is measured as cross-sparsity. Considering the advantages of Cr-SF in extracting features of small sample [25], this paper proposes parallel Cr-SF networks.

Firstly, each type of fault sample is trained through Cr-SF to learn the weights of each fault condition. The second step is feature selection and optimization, in which the most representative features are selected to arrange and combine the features of the whole fault conditions. The third step is to input the entire features and label data into the classifier for training. In this way, when new fault type data appears and the trained monitoring system needs to be upgraded, we only need to extract the features of the newly added data and add the features to the existing feature matrix to retrain the classifier. There is no need to retrain all training samples, which saves the time of system upgrade.

The rest of this study is organized as follows. Section 2 introduces Cr-SF and the proposed structure of parallel network. Section 3 verifies the effectiveness of the proposed method through rolling bearing and planet gear fault datasets. The visualization of weights and features is discussed in Section 4. Finally, the conclusions are given in Section 5.

## 2. Proposed Method

**2.1. Cross-Sparse filtering.** As shown in Figure 1, cross sparse filtering is the variant of SF, Which can be regarded as a two-layer neural network. The optimization process of Cr-SF is simultaneous for the rows and columns of feature matrix. Therefore, the objective function is composed of two terms:  $l_{1/2}$ -norms of rows and  $l_{1/2}$ -norms of columns of feature matrix. Suppose the input dimension and output dimension of Cr-SF are  $N_{in}$  and  $N_{out}$ . The collected original signal  $\mathbf{x} \in \mathcal{R}^N$  is randomly segmented into a matrix segment matrix  $\mathbf{S} \in \mathcal{R}^{N_{in} \times M}$ , where  $M = N_{s1} \times m$ ,  $m$  is training samples number, and  $N_{s1}$  is the segment number.

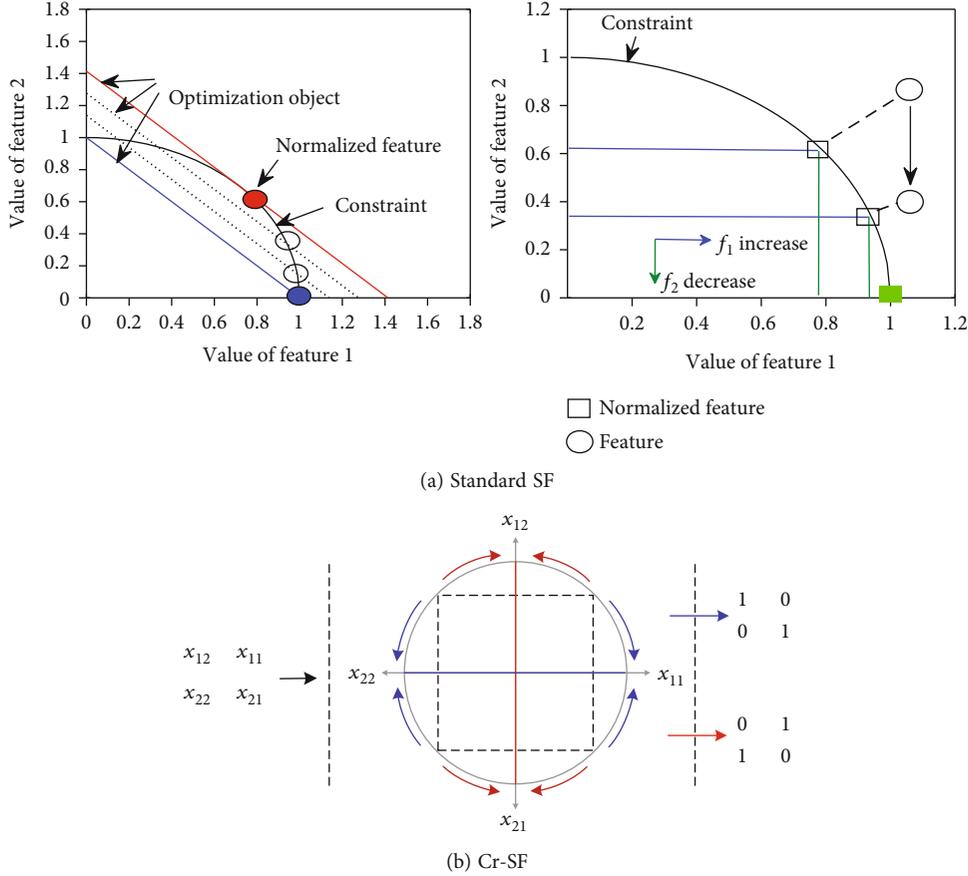


FIGURE 2: Schematic diagram of the optimization target and the constraints of various methods.

First, the local feature matrix  $\mathbf{F} \in \mathfrak{R}^{N_{out} \times M}$  is activated by the product of weight matrix  $\mathbf{W} \in \mathfrak{R}^{N_{out} \times N_{in}}$  and input matrix  $\mathbf{S} \in \mathfrak{R}^{N_{in} \times M}$ , that is,  $\mathbf{F} = \mathbf{W}\mathbf{S}$ .

Second, the objective function of Cr-SF can be constructed by the  $l_{1/2}$ -norms of rows and  $l_{1/2}$ -norms of columns of the feature matrix  $\mathbf{F}$ . The two terms can be written as

$$\mathcal{L}_1 = \left\| \left\| \frac{\mathbf{F}^i}{\|\mathbf{F}^i\|_2} \right\|_1 \right\|, \quad (1)$$

$$\mathcal{L}_2 = \left\| \left\| \frac{\mathbf{F}_j}{\|\mathbf{F}_j\|_2} \right\|_1 \right\|. \quad (2)$$

Then, weight vectors are constrained to its  $l_2$ -norm sphere to eliminate the influence of redundancy in an optimization process. Therefore, the final objective function of the Cr-SF can be expressed as

$$\mathcal{L} = \sum_i^M \mathcal{L}_1 + \lambda \sum_j^{N_{in}} \mathcal{L}_2 S \cdot T |W^i|_2 = 1, \quad (3)$$

where  $\lambda \geq 0$  is an adjustable parameter that adjusts the priority of two items in the sparse optimization process. According to the discussion of references [26], the value of  $\lambda$  is 1 in this

paper. Due to the fact that  $\mathcal{L}$  is nonconvex and nonsmooth,  $|\mathbf{F}|$  is replaced by a soft-absolute function  $\sqrt{\mathbf{F}^2 + \varepsilon}$ , where  $\varepsilon$  is a small number and equals  $1 \times 10^{-8}$ .

Third, the sparse optimization process uses the L-BFGS algorithm. The gradient function of the objective function can be given by

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{w}} = & \left[ \left( \frac{\mathbf{O}}{\left( \sum_j^{N_{in}} |\mathbf{F}^j|^2 \right)^{1/2}} - \frac{\mathbf{F} \cdot \sum_j^{N_{in}} (\mathbf{F}^j)}{\left( \sum_j^L |\mathbf{F}^j|^2 \right)^{3/2}} \right) \right. \\ & \left. + \left( \frac{\mathbf{O}}{\left( \sum_i^M |\mathbf{F}_i|^2 \right)^{1/2}} - \lambda \frac{\mathbf{F} \cdot \sum_i^M (\mathbf{F}_i)}{\left( \sum_i^M |\mathbf{F}_i|^2 \right)^{3/2}} \right) \right] \cdot \frac{\mathbf{f}}{|\mathbf{f}|} \cdot \mathbf{x}', \end{aligned} \quad (4)$$

where  $\mathbf{o} \in \mathfrak{R}^{N_{in} \times M}$  is a matrix of all ones.

As discussed earlier, the optimization of Cr-SF is a cross-sparse optimization process. In order to show more clearly, we present the optimization diagram of SF and Cr-SF in Figure 2. It can be seen that the optimization of SF is a process in which the column features are gradually sparse under row competition constraints. As one feature increases, the other must decrease to ensure the direction of optimization. Cr-SF is a sparse optimization of column features and row

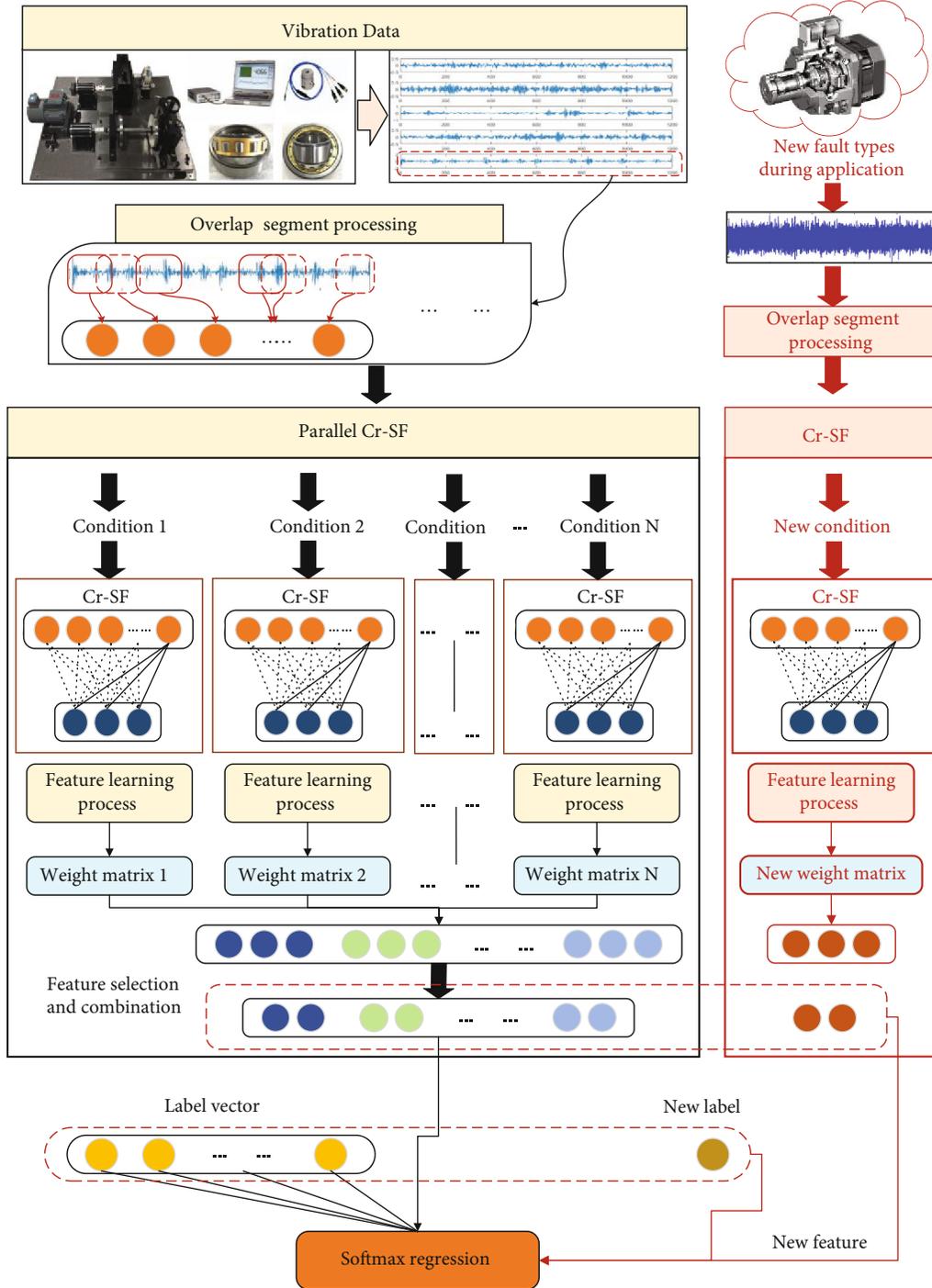


FIGURE 3: Parallel network structure of the proposed method.

features at the same time. The sparse process between column feature and row feature is separate. For a 2-to-2 matrix, the ideal result of Cr-SF will become a standard orthogonal matrix.

2.2. *Parallel Networks Using Cr-SF.* This section presents the parallel networks based on the Cr-SF. Figure 3 shows the flow chat of the parallel diagnosis model. The general procedures can be summarized as follows.

(Step 1) *Model training:* different from the traditional intelligent fault diagnosis methods, the proposed method trains different fault conditions separately. Specifically, suppose that the number of input fault types is  $L$ . The fault datasets with same fault type are input into Cr-SF, respectively. Each sample is randomly segmented into a matrix. The sample features are activated through formulas (1)–(3), and the

weights are optimized following formula (4). The trained weights  $\mathbf{W}_i$  are obtained, where  $i$  is the number of fault type.

- (Step 2) Feature selection and arrangement: local features  $\mathbf{F}_i \in \mathfrak{R}^{N_{\text{out}} \times M}$  can be calculated using the non-linear activation function and the trained weights,  $\mathbf{F}_i = \text{Log}(1 + \mathbf{W}_i \mathbf{S}_i)$ . The local features are averaged to obtain the features of each fault sample,  $\mathbf{f}_i = (\sum_{j=1}^{N_s} F_i^j) / N_s$ , where  $N_s$  is the number of segments. According to the order of eigenvalues from large to small,  $n$  features are selected as the final features of fault samples. Then, the features of all fault samples are arranged and combined into a feature matrix  $\mathbf{f} = \mathfrak{R}^{n \times m}$  as the input of softmax, where  $m$  is the number of training samples.
- (Step 3) Classifier training using the learned features  $\{\mathbf{f}^i\}_{i=1}^m$  and the corresponding label set  $\{y^i\}_{i=1}^m$ .  $\{\mathbf{f}^i\}_{i=1}^m$  and  $\{y^i\}_{i=1}^m$  are fed into the softmax classifier, where  $y^i \in \{1, 2, 3, \dots, k\}$  denotes the label vector of each fault type. The hypothesis  $h_\theta(\mathbf{f})$  can be written as

$$h_\theta(x^i) = \frac{1}{\sum_{j=1}^k \exp(\theta_j^T f^i)} \begin{bmatrix} \sum_{j=1}^k \exp(\theta_1^T f^i) \\ \sum_{j=1}^k \exp(\theta_2^T f^i) \\ \vdots \\ \sum_{j=1}^k \exp(\theta_k^T f^i) \end{bmatrix}, \quad (5)$$

where  $p(y^i = j | \mathbf{f})$  means the probability for each feature and  $\theta_1, \theta_2, \dots, \theta_k$  are the parameters of softmax model.

The cost function of the softmax regression is a cross-entropy function and takes the form

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^k 1\{y^i = j\} \lg \frac{\exp(\theta_j^T f^i)}{\sum_{l=1}^k \exp(\theta_l^T f^i)} + \frac{\lambda}{2} \sum_{k=1}^K \sum_{l=1}^N \theta_{kl}^2, \quad (6)$$

where  $m$  is the training samples number and  $k$  is the category number and  $1\{\bullet\}$  is the indicator function.

After the model is trained, the features of the online monitoring data are extracted through the trained weights and then input into the softmax classifier to realize feature recognition.

As shown in Figure 3, when a new fault needs to be added, the features of the new fault sample data can be trained through Cr-SF, and then the trained weights and fea-

TABLE 1: Description of the ten bearing working conditions.

Fault condition	Fault diameter (mm)	Motor load (HP)	Sample number	Condition label
NC	—	0	400	1
OF	0.18	0	400	2
OF	0.36	0	400	3
OF	0.53	0	400	4
IF	0.18	0	400	5
IF	0.36	0	400	6
IF	0.53	0	400	7
BF	0.18	0	400	8
BF	0.36	0	400	9
BF	0.53	0	400	10

tures are selected and arranged according to the way in Step 2 to establish a new weight matrix, and then feature extraction and classifier training are carried out for all fault samples. In this way, the newly added fault data does not need to be trained at the same time with the existing fault data. It only needs to train the classifier after combining the trained features.

### 3. Experimental Validation

In this paper, the planetary gear and rolling bearing datasets are employed to demonstrate the proposed parallel diagnosis model. It should be noted that the proposed model is actually a process of adding new faults when training multiple fault types. Therefore, this study does not separately verify the accuracy of the experiment adding new faults. We only need to compare the accuracy of multifault diagnosis and the training time of adding a new fault.

*3.1. Rolling Bearing Data Verification.* In this experiment, the rolling bearings fault datasets, which provided from the Case Western Reserve University Lab, are used to demonstrate the diagnostic performance of the proposed method. The motor bearing test bench is mainly composed of an induction electrical motor, the testing bearings, and an acceleration sensor. Two vibration sensors are installed at the drive end and output end of the motor, respectively. This experiment consists of four different fault types: normal condition (NC), inner race fault condition (IF), outer race fault condition (OF), and roller fault condition (RF). For each fault type, three different severity levels (0.18, 0.36, and 0.53 mm) are designed. Therefore, this experiment totally contains ten kinds of fault conditions. During data acquisition, the sampling frequency is set to 12 kHz. Each sample contains 1200 data points. Each health condition includes 100 vibration samples under one load, and there are 1000 samples for this study, as shown in Table 1.

In the experimental verification, the output dimension  $N_{\text{out}}$  of the parallel neural unit is 20 and the input dimension  $N_{\text{in}}$  is 100. Figure 4 shows the comparison of accuracy, standard deviation, and computational time obtained by different methods with the change of percentage of training

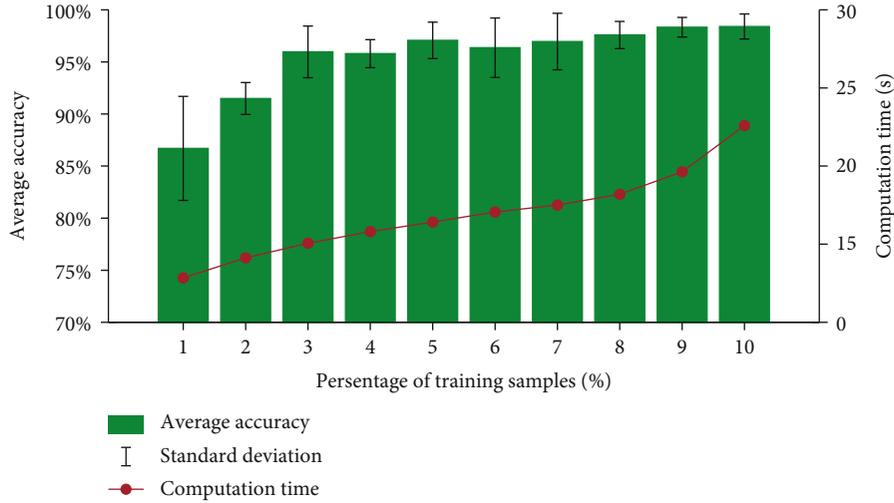


FIGURE 4: Diagnostic results of parallel SF with different percentage of training data.

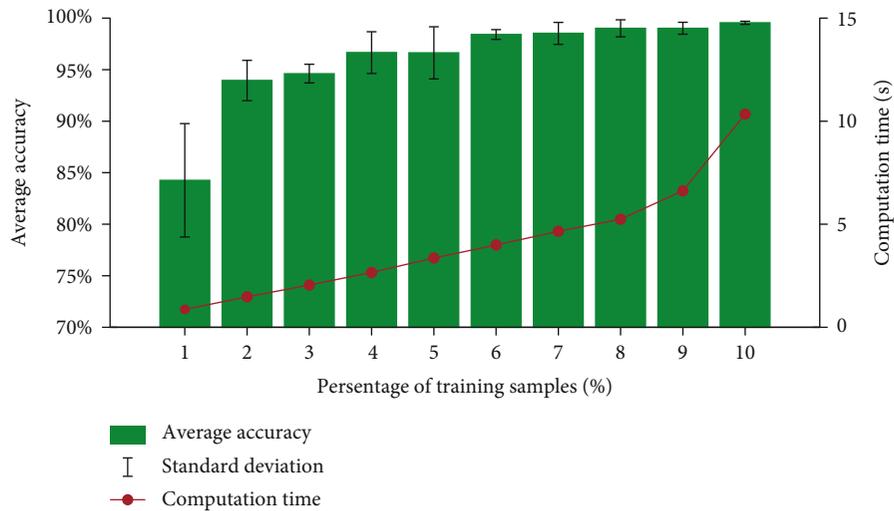


FIGURE 5: Diagnostic results of traditional structure with different percentage of training data.

data. Firstly, by comparing Figures 5 and 4, it can be seen that under the same output dimension, the parallel structure performs a negative impact on the accuracy and calculation time. When the standard SF algorithm is transformed to parallel network, the accuracy decreases and the calculation time is longer. The reason may be that the structure needs to calculate more times and the number of training samples of each parallel neural unit is much smaller than that of the traditional structure. Therefore, the experimental results also verify the problem mentioned in the preface: the parallel structure requires higher feature extraction ability of neural units in small samples.

As can be seen from Figure 6, compared with the other methods, a parallel Cr-SF network has significantly improved in the accuracy due to the stronger feature extraction ability of Cr-SF. For example, when the proposed method is trained with by only 2% data, the accuracy is above 99% and the standard deviation is only 0.81%. While when the parallel SF network is trained with 15% samples, the accuracy rate is only 98.4% and standard deviation is 1.2%. With the increasing of

the percentage of training data, the accuracy increases. When the percentage of training data is above 3%, the accuracy of the proposed method can reach a relatively stable state. The experimental results show that the proposed method has higher accuracy and better robustness in case of small samples. The accuracy and stability of feature extraction based on Cr-SF are the premise to ensure the use of parallel network structure.

Figures 7 and 8 show the impact of the different feature dimension of each condition on the diagnosis performance. The output dimension  $N_{out}$  of the parallel neural unit is 20, the input dimension  $N_{in}$  is 100, and the percentage of training data is 10%. It can be observed that the test accuracy of parallel SF gradually increases, when the dimension is 20, which means the input dimension of softmax is 200, the accuracy is 98.4%, and the standard deviation is 1.2%. However, when the dimension is greater than 8, the accuracy of parallel Cr-SF is relatively stable, and the accuracy can achieve above 99.9%. In this case, the input dimension of softmax is only 80.

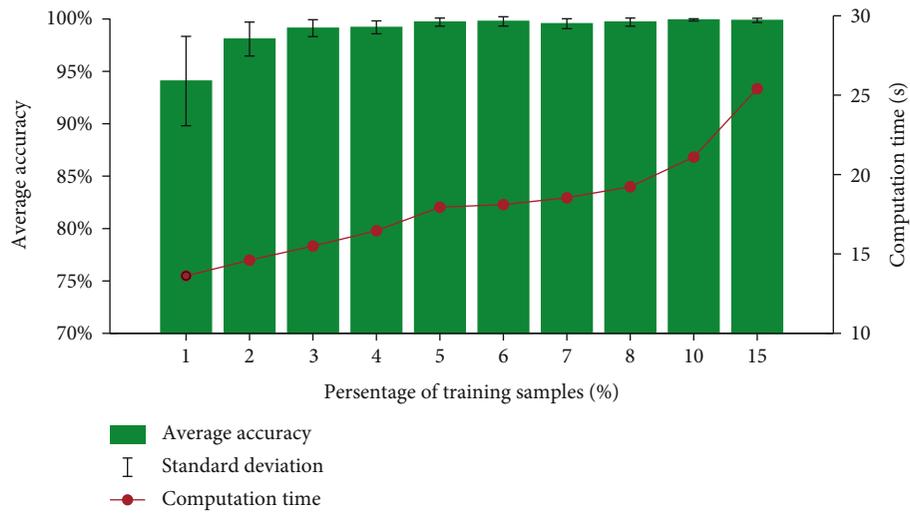


FIGURE 6: Diagnostic results of parallel Cr-SF with different percentage of training data.

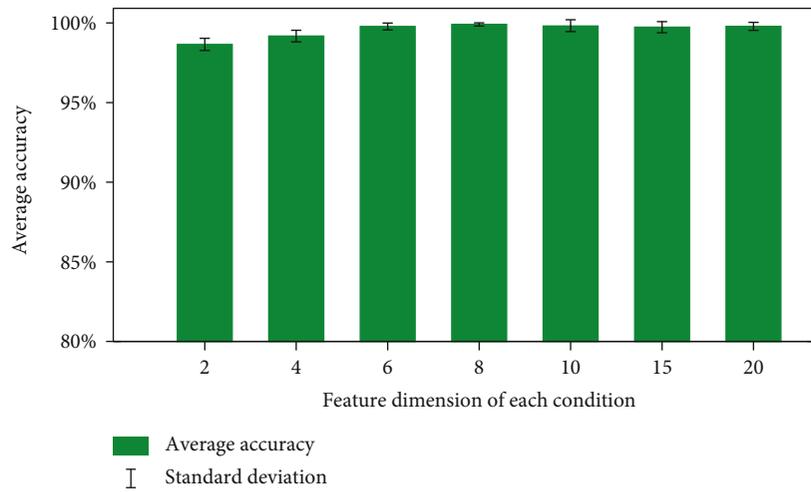


FIGURE 7: Diagnostic results of parallel Cr-SF with different feature dimension of each condition.

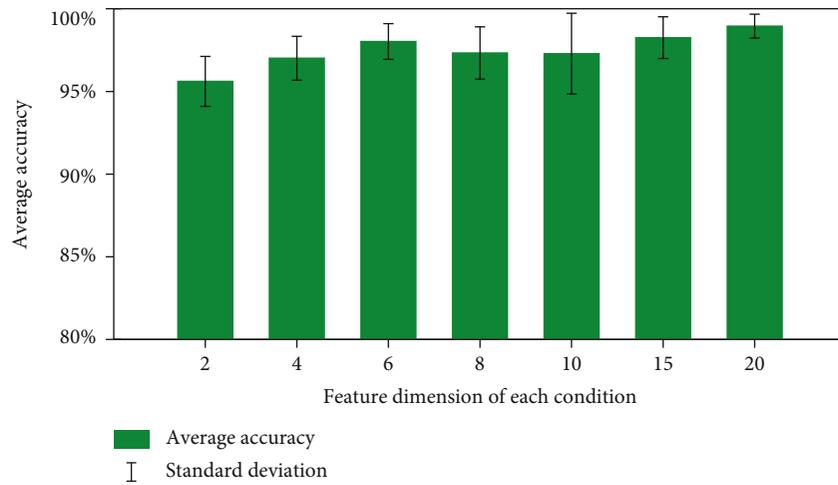


FIGURE 8: Diagnostic results of parallel SF with different feature dimension of each condition.

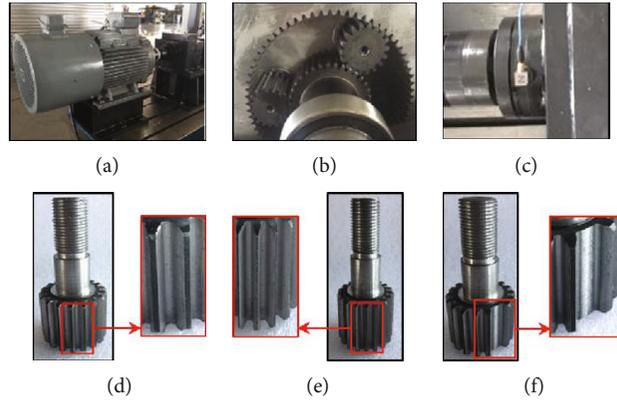


FIGURE 9: Gear fault feature and test bench diagrams. (a) Test rig, (b) schematic of the gearbox, (c) triaxial accelerometer, (d) normal gear, (e) worn teeth, and (f) broken teeth.

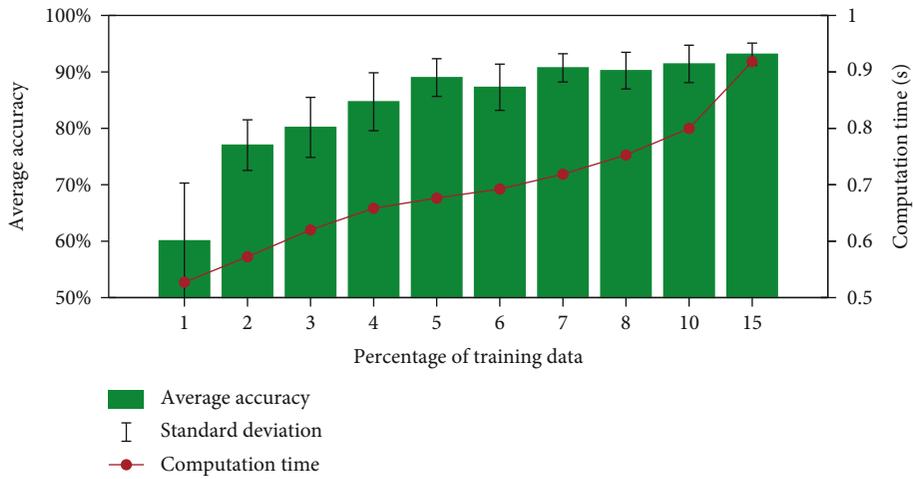


FIGURE 10: Diagnostic results of parallel SF with different percentage of training data.

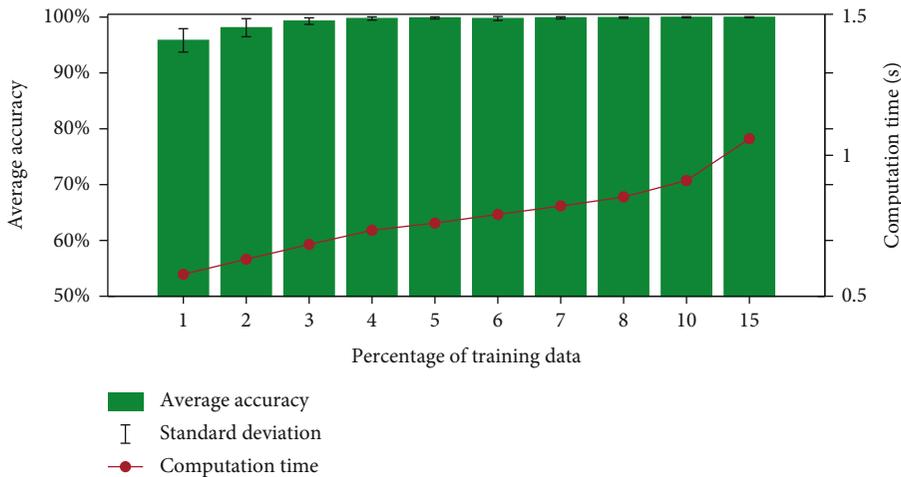


FIGURE 11: Diagnostic results of parallel Cr-SF with different percentages of training data.

3.2. Gear Data Verification and Analysis. In this experiment, the vibration data is collected from the driving end of the planetary gearbox on the test bench, as shown in Figure 9. The test bench is composed of a motor with a rated speed of 1500 rpm, a planetary gearbox with three planetary gears,

a multichannel data acquisition system, a three-way acceleration sensor, and a tachometer. In order to increase the interference of the environment, the test-bed is placed in an assembly workshop, and the base of the test-bed has no shock absorber. This test simulates four health conditions

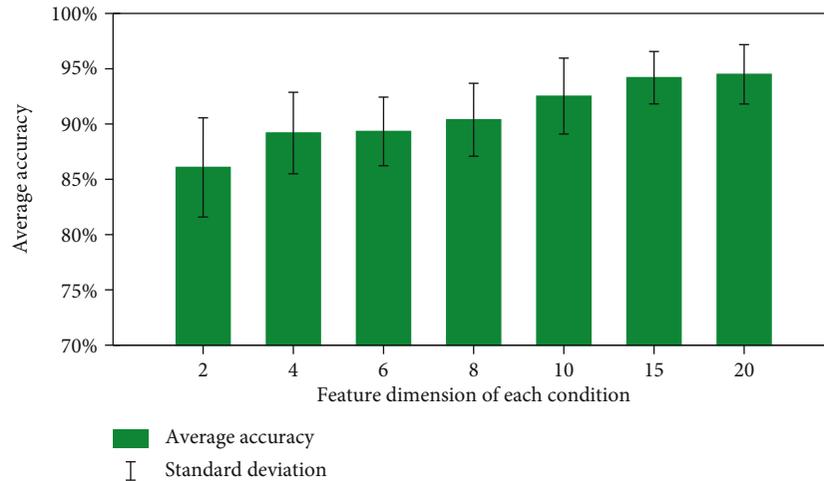


FIGURE 12: Diagnostic results of parallel SF with different feature dimensions of each condition.

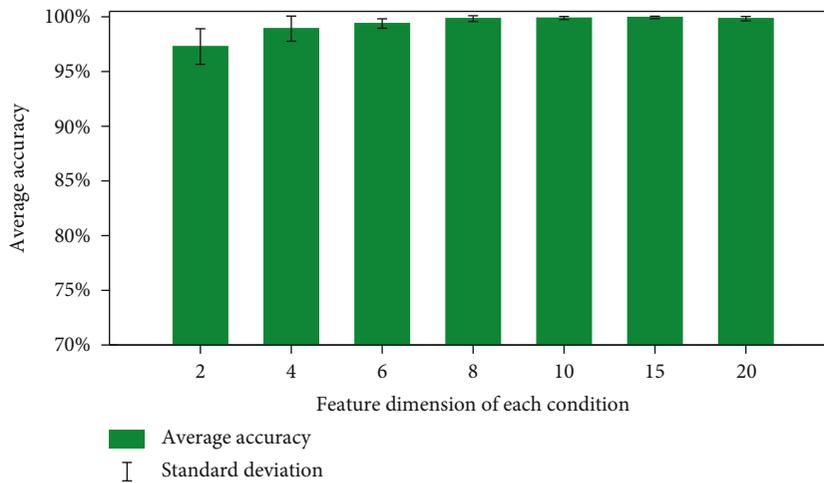


FIGURE 13: Diagnostic results of parallel Cr-SF with different feature dimensions of each condition.

of planetary gears, namely, normal working condition (NC), planetary gear wear fault (WF), planetary gear broken tooth (BT), and compound fault of broken tooth and wear (CF). The sampling frequency is set as 12.8 kHz. There are 600 samples under each health condition. Each test data includes 2400 sample points.

The diagnostic results using parallel SF and Cr-SF with various percentage of training data are shown in Figures 10 and 11. In this experiment, the parameter settings are consistent with the previous verification. The diagnostic results show a pattern similar to that in the previous section. With the increase of the percentage of training data, the accuracy increases gradually. The diagnostic results of parallel SF were significantly worse than those of parallel Cr-SF. Taking the training percentage of 3% as an example, the accuracy of parallel Cr-SF is  $99.28\% \pm 0.59\%$ , but the accuracy of parallel SF is only  $80.17\% \pm 5.3\%$ . When the training percentage is greater than 3%, the accuracy is above 99%. The diagnostic results show that the proposed method can ensure higher diagnosis accuracy and robustness with fewer samples.

Figures 12 and 13 show the changes of accuracy and standard deviation of the two methods with the increase of feature dimension of each condition. The percentage of training samples was 10%. When only two features are used to express the fault condition, the accuracy of the proposed method can reach 97.28%, and the standard disassembly is only 1.62%, which is much better than parallel SF. When the feature dimension is equal to 8, the accuracy can reach 99.83%.

#### 4. Discussion on Feature and Weight Visualization

In order to further intuitively explain the performance of the proposed method and its feature extraction process, the visualization of weights and features are studied in case of 10% of training samples.

As shown in Figure 14, it is obvious that the learning features of Cr-SF are sparser, and there are only a few principal components in the features, which can explain why the

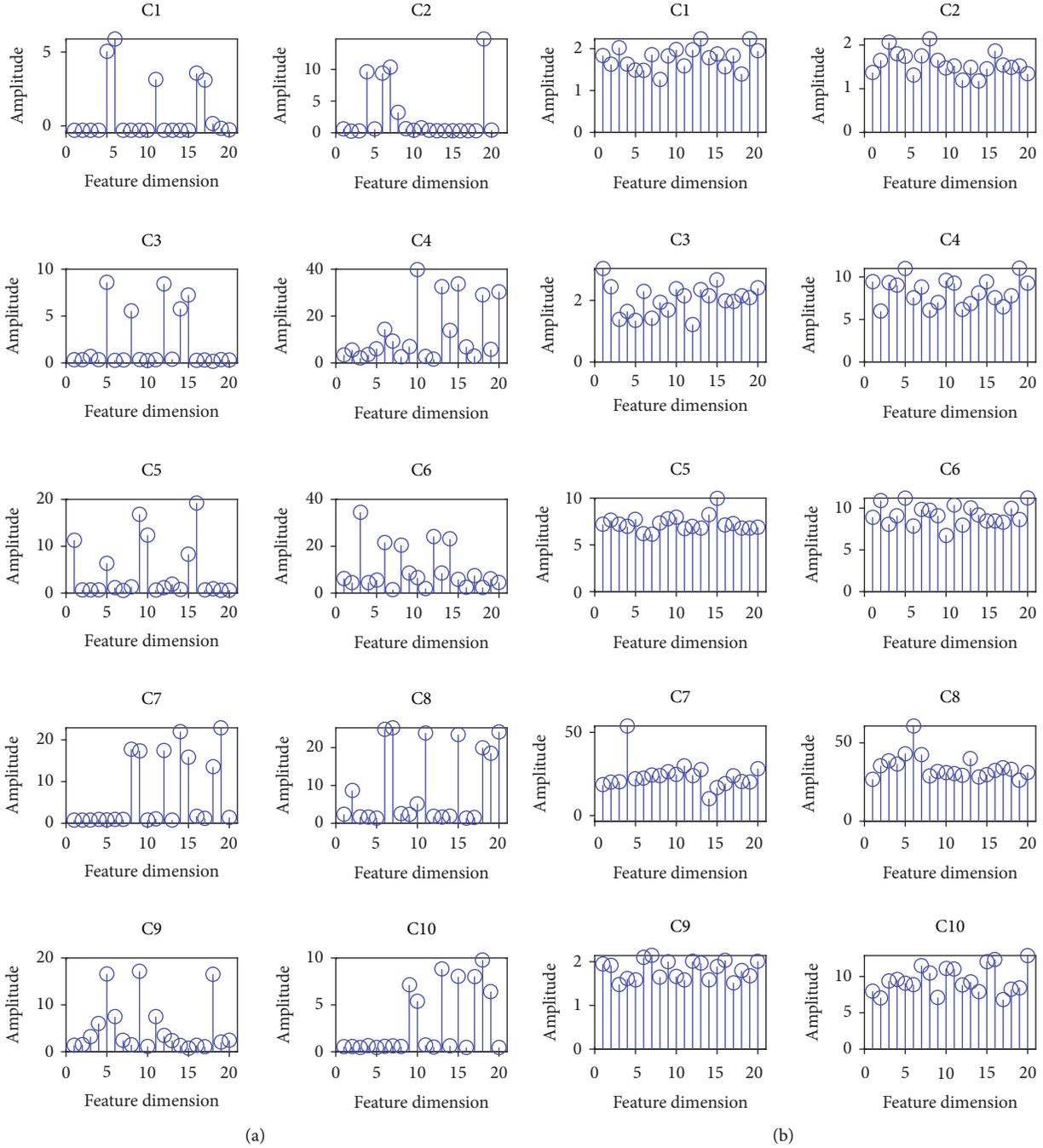


FIGURE 14: Learned feature vectors of each condition: (a) parallel Cr-SF and (b) parallel SF.

accuracy of the proposed method does not change significantly after the feature number is 8; all the main features have been selected in this case. The value of features of SF has little change and performs poor sparsity. At this case, more features need to be selected to express the main fault information to achieve high diagnostic accuracy. Therefore, the feature learning process of Cr-SF performs stronger robustness and better sparsity, which is the key to ensure the accuracy and robustness of the proposed algorithm in the diagnosis process.

In order to further explain the sparsity of features, Fourier transform is carried out on the learned filters to obtain the corresponding spectrum, as shown in Figure 15. It can be seen that the frequency components of the filters learned by the proposed method are clearer, and the energy of the filters is concentrated in a specific frequency range. However, the noise interference of the filters learned by SF is obvious. Compared with Cr-SF, there are many noise disturbances, especially in the low frequency band. It should be noted that some filters learned by Cr-SF contains high-

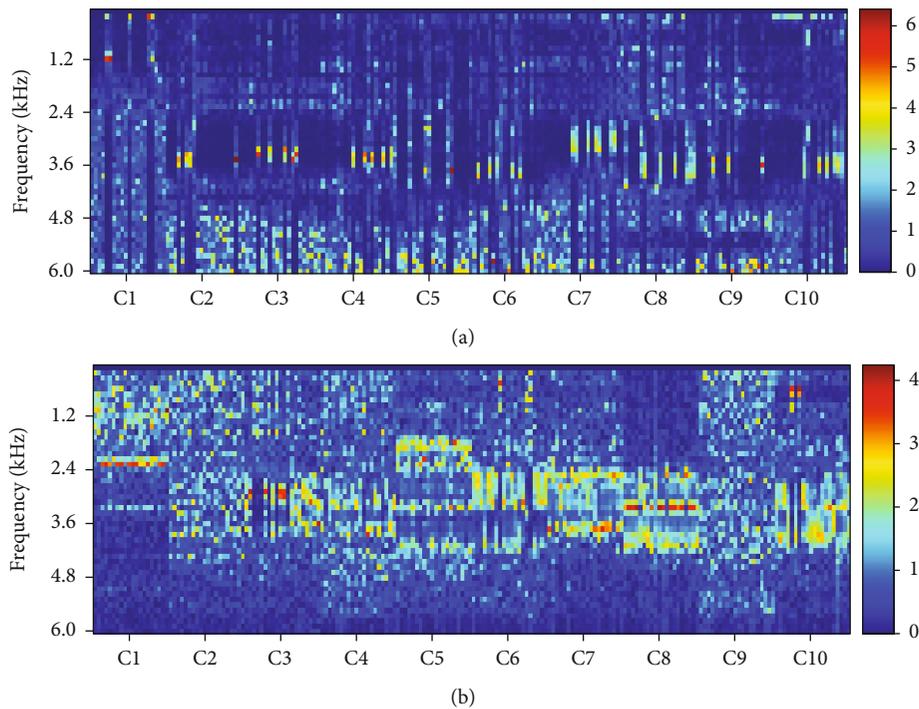


FIGURE 15: Learned weights of each condition: (a) parallel Cr-SF and (b) parallel SF.

frequency noise. However, we find that the peak value of characteristic frequency in these filters is small. The corresponding feature value is very small. The characteristic amplitudes corresponding to these filters are close to 0. Therefore, these features can be discarded in the process of feature selection.

## 5. Conclusions

This paper presents a parallel network structure based on the Cr-SF, a minimum entropy unsupervised learning method. The proposed method solves the problem that the detection system is required to retrain the whole model when new faults are added in real industrial application. The experimental verification is carried out using the fault data of bearing and planetary gearbox. The results confirm that the proposed method can ensure the sparsity and training robustness of the feature extraction process and performs superior diagnostic accuracy, stability and computational efficiency comparing with the traditional methods. In the proposed method, fault feature training and classifier training are separated, and the training of each fault type is implemented independently. When a new fault is added, it only needs to extract the features of the new fault and then input them into the classification for retraining.

The limitation of the proposed method is that there is no clear judgment principle in the feature selection process. In some cases, the magnitude of the amplitude does not indicate the importance of the feature, especially in the strong interference environment. Therefore, the future works of authors will focus on more advanced feature selection methods to further improve the performance of the proposed algorithm.

## 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 that they have no conflicts of interest.

## Acknowledgments

The research was supported by the National Natural Science Foundation of China (52005303 and 52105110) and the Natural Science Foundation of Shandong Province (ZR2020QE157).

## References

- [1] X. Li, J. Cheng, H. Shao, K. Liu, and B. A. Cai, "A fusion CWSMM-based framework for rotating machinery fault diagnosis under strong interference and imbalanced Case," *IEEE Transactions on Industrial Informatics*, p. 1, 2021.
- [2] Z. He, H. D. Shao, Z. Y. Ding, H. K. Jiang, and J. S. Cheng, "Modified deep auto-encoder driven by multi-source parameters for fault transfer prognosis of aero-engine," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 1, pp. 845–855, 2022.
- [3] H. Chen, B. Jiang, and S. Ding, "Data-driven fault diagnosis for traction systems in high-speed trains: a survey, challenges, and perspectives," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–17, 2020.
- [4] Z. Z. Zhang, S. M. Li, J. T. Lu, J. R. Wang, and X. X. Jiang, "A novel intelligent fault diagnosis method based on fast intrinsic

- component filtering and pseudo-normalization,” *Mechanical Systems and Signal Processing*, vol. 145, article 106923, 2020.
- [5] S. S. Ji, B. K. Han, Z. Z. Zhang, and J. R. Wang, “Parallel sparse filtering for intelligent fault diagnosis using acoustic signal processing,” *Neurocomputing*, vol. 462, pp. 466–477, 2021.
  - [6] C. Li, S. H. Zhang, Y. Qin, and E. Estupinan, “A systematic review of deep transfer learning for machinery fault diagnosis,” *Neurocomputing*, vol. 407, pp. 121–135, 2020.
  - [7] R. Q. Yan, X. F. Chen, P. Wang, and M. Darian, “Deep learning for fault diagnosis and prognosis in manufacturing systems,” *Computers in Industry*, vol. 110, pp. 1–2, 2019.
  - [8] F. Jia, Y. G. 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.
  - [9] J. R. Wang, S. M. Li, Z. H. 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.
  - [10] G. Li, C. Deng, and J. Wu, “Sensor data-driven bearing fault diagnosis based on deep convolutional neural networks and S-transform,” *Sensors*, vol. 19, no. 12, p. 2750, 2019.
  - [11] H. Shao, H. Jiang, H. Zhang, W. Duan, T. C. H. Liang, and S. H. P. Wu, “Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing,” *Mechanical Systems and Signal Processing*, vol. 100, pp. 743–765, 2018.
  - [12] H. Shao, H. Jiang, H. Zhang, T. Liang, and S. H. P. Wu, “Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network,” *IEEE Transactions on Industrial Electronics*, vol. 65, no. 3, pp. 2727–2736, 2018.
  - [13] H. Shao, H. Jiang, H. Zhao, and F. Wang, “A novel deep auto-encoder feature learning method for rotating machinery fault diagnosis,” *Mechanical Systems and Signal Processing*, vol. 95, pp. 187–204, 2017.
  - [14] Y. B. Li, F. Liu, S. Wang, and J. Yin, “Multiscale symbolic Lempel-Ziv: an effective feature extraction approach for fault diagnosis of railway vehicle systems,” *IEEE Transactions on Industrial Electronics*, vol. 17, no. 1, pp. 199–208, 2021.
  - [15] Y. B. Li, Y. Ren, H. Zheng, Z. Deng, and S. Wang, “A novel cross-domain intelligent fault diagnosis method based on entropy features and transfer learning,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–14, 2021.
  - [16] Y. B. Li, S. Wang, N. Li, and Z. Deng, “Multiscale symbolic diversity entropy: a novel measurement approach for time-series analysis and its application in fault diagnosis of planetary gearboxes,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 2, pp. 1121–1131, 2022.
  - [17] Z. Wang, X. He, B. Yang, and N. Li, “Subdomain adaptation transfer learning network for fault diagnosis of roller bearings,” *IEEE Transactions on Industrial Electronics*, p. 1, 2021.
  - [18] S. X. Jia, J. R. Wang, X. Zhang, and B. K. Han, “A weighted subdomain adaptation network for partial transfer fault diagnosis of rotating machinery,” *Entropy*, vol. 23, no. 4, p. 424, 2021.
  - [19] Y. Lei, F. Jia, J. Lin, S. Xing, and S. 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] Y. H. Chang, J. L. Chen, C. Qu, and T. Y. 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.
  - [21] S. Guo, T. Yang, H. C. Hua, and J. W. Cao, “Coupling fault diagnosis of wind turbine gearbox based on multitask parallel convolutional neural networks with overall information,” *Renewable Energy*, vol. 178, pp. 639–650, 2021.
  - [22] J. Ngiam, Z. Chen, S. A. Bhaskar, P. W. Koh, and A. H. Ng, “Sparse filtering,” *Advances in Neural Information Processing Systems*, pp. 1125–1133, 2011.
  - [23] F. M. Zennaro and K. Chen, “Towards understanding sparse filtering: a theoretical perspective,” *Neural Networks*, vol. 98, pp. 154–177, 2018.
  - [24] Z. Z. Zhang, S. M. Li, J. R. Wang, Y. Xin, and Z. H. An, “General normalized sparse filtering: a novel unsupervised learning method for rotating machinery fault diagnosis,” *Mechanical Systems and Signal Processing*, vol. 124, pp. 596–612, 2019.
  - [25] Z. Z. Zhang, S. M. Li, J. T. Lu, Y. Xin, and H. J. Ma, “Intrinsic component filtering for fault diagnosis of rotating machinery,” *Chinese Journal of Aeronautics*, vol. 34, no. 1, pp. 397–409, 2021.
  - [26] Z. Z. Zhang, S. M. Li, J. T. Lu, J. R. Wang, and X. X. Jiang, “A novel unsupervised learning method based on cross-normalization for machinery fault diagnosis,” *IEEE Access*, vol. 8, pp. 92407–92417, 2020.