Research Article

Rolling-Element Bearing Fault Data Automatic Clustering Based on Wavelet and Deep Neural Network

Yanli Yang ⓘ **and Peiying Fu** ⓘ

Tianjin Key Laboratory of Optoelectronic Detection Technology and Systems, Tianjin Polytechnic University, Tianjin 300387, China

Correspondence should be addressed to Yanli Yang; yyl070805@163.com

A method based on wavelet and deep neural network for rolling-element bearing fault data automatic clustering is proposed. The method can achieve intelligent signal classification without human knowledge. The time-domain vibration signals are decomposed by wavelet packet transform (WPT) to obtain eigenvectors that characterize fault types. By using the eigenvectors, a dataset in which samples are labeled randomly is configured. The dataset is roughly classified by the distance-based clustering method. A fine classification process based on deep neural network is followed to achieve accurate classification. The entire process is automatically completed, which can effectively overcome the shortcomings such as low work efficiency, high implementation cost, and large classification error caused by individual participation. The proposed method is tested with the bearing data provided by the Case Western Reserve University (CWRU) Bearing Data Center. The testing results show that the proposed method has good performance in automatic clustering of rolling-element bearings fault data.

1. Introduction

Rolling-element bearings are widely used in industry. Such bearings play a pivotal role in the rotating machine because they can reduce friction between moving parts and allow the machine to operate efficiently. However, rolling-element bearings are also the most vulnerable component of a machine. Damage to the bearing can cause faults in the machine, potentially leading to severe accidents [1]. According to statistics, rolling-element bearing failure is an important factor causing rotating machinery failures [2]. Even in some devices, bearing faults can constitute 44% of the total number of faults [3]. Thus, regular inspections and diagnosis are critical to ensure that rolling-element bearings work properly.

Vibration monitoring is one of the most useful techniques for condition monitoring because it is reliable and very sensitive to fault severity [4]. It was noted in [3] that vibration signal monitoring reveals early signs of abnormality, even several months prior to any permanent damage.

The literature [5] reviews typical bearing fault frequencies and their expressions, including the ball pass frequency of the outer race (BPFO), the ball pass frequency of the inner race (BPFI), the fundamental train frequency (FTF), and the ball spin frequency (BSF).

Consensus is that fault diagnosis relies on reliable signal acquisition and efficient signal processing. Wavelet is characterized by its multiresolution and the ability to characterize local features of signals in the time-frequency domain, so it has been widely used in the field of fault diagnosis. The wavelet transform (WT) can be used not only to extract fault features [6], but also to combine with other methods, such as the random forests classifier [7], the particle swarm optimization and the nearest neighbor classifier [8], and the support vector machine [9], to achieve fault diagnosis. Moreover, the WT can work with artificial neural network (ANN) to estimate the fault location [10]. The review about the application of WT in rotating machinery diagnostics and prognostics was presented in [11, 12]. Although fault diagnosis based on signal processing methods such as the WT

has achieved many important results in bearing fault diagnosis, it is still necessary to continue to explore better fault diagnosis methods in the era of big data.

Deep learning is a powerful tool for processing big data [13], and it has achieved much success in speech recognition, visual object recognition, object detection, and many other domains [14]. Deep learning has led to breakthroughs in the field of bearing fault diagnosis [15]. It was shown in [1] that the three deep neural network (DNN) models, deep Boltzmann machine (DBM), deep belief network (DBN), and stacked autoencoders (SAE) are highly reliable and applicable in fault diagnosis of rolling-element bearings. A recent review on the applications of deep learning in machine health monitoring was presented in [16], which holds that deep learning is able to act as a bridge connecting big machinery data and intelligent machine health monitoring. The combination of the WT and deep learning has begun to detect faults of machines. The wavelet packet transform (WPT) used to extract fault features was employed to help DBN [17] and CNN [18] to achieve intelligent fault diagnosis.

It is ongoing research for automatic fault detection of mechanical systems. A key technology for intelligent fault diagnosis is the automatic classification of fault data. There is a large amount of unlabeled data in practice, which makes it more difficult to realize intelligent fault diagnosis. Deep learning which can learn fault features autonomously without human intervention is a powerful tool for intelligent fault diagnosis. Thus, deep learning is used here to achieve automatic signal classification without the help of human expert knowledge.

The main contribution of this paper is to realize the automatic classification of fault data by combining the wavelet with DNN. In the proposed method, time-domain vibration signals are decomposed by the WPT to build a training dataset in which the samples are labeled randomly. The training dataset is roughly classified by clustering methods. The labels of samples are adjusted according to the clustering results. A fine classification process based on the DNN is followed to achieve accurate classification of the samples. The features are automatically extracted, and fault data are automatically classified. The testing results on the bearing data provided by Case Western Reserve University (CWRU) Bearing Data Center show that this method can realize automatic classification of fault signals. Detailed about the proposed algorithm is introduced with the following steps. In Section 2, features extraction through the WPT is analyzed. Section 3 gives a detailed description of the proposed fault data automatic clustering method based on a DNN. In Section 4, testing of the proposed method with the CWRU bearing data is presented. Finally, the conclusions are drawn in Section 5.

2. Features Extraction by the WPT

The WT is known to deal with signals with mutation or functions with isolated singularity. This approach is able to characterize the local features of signals in both time and frequency domains. However, the WT has the poor resolution in the high-frequency regions because it just redecomposes the low-frequency part of signals.

The WPT overcomes the poor resolution of the WT through further decomposing of the high-frequency part of signals and offers a more comprehensive signal analysis [19, 20]. This approach decomposes a signal into a set of wavelet packet (WP) nodes with the form of a full binary tree [6]. A three-level WPT tree structure is shown in Figure 1. Each level of the WPT provides a frequency range that is half as wide as the preceding level and twice as wide as the proceeding level [21].

The WP coefficients of a signal $x(t)$ can be calculated by [20]

$$d_j^n(k) = \langle x, W_{j,k}^n \rangle = \int_{-\infty}^{\infty} x(t) W_{j,k}^n(t) dt, \tag{1}$$

where $\langle \cdot, \cdot \rangle$ stands for the inner product operator, $d_j^n(k)$ denotes the nth set of WP coefficients at the jth scale parameter, k is the translation parameter, and $W_{j,k}^n(t)$ is the WP function. The WP function is defined as [20]

$$W_{j,k}^n(t) = 2^{j/2} W^n\left(2^j t - k\right). \tag{2}$$

Equation (2) shows that WP function can be obtained by a recursive equation in which $W^0(t) = \phi(t)$ and $W^1(t) = \psi(t)$ are also called the scaling function and the mother wavelet, respectively. The WP function $W^n(t)$ can be calculated by [6]

$$\begin{aligned} W^{2n}(t) &= \sqrt{2} \sum_k h(k) W^n(2t - k), \\ W^{2n+1}(t) &= \sqrt{2} \sum_k g(k) W^n(2t - k), \end{aligned} \tag{3}$$

where $h(\cdot)$ and $g(\cdot)$ are the low-pass and high-pass filters associated with the mother wavelet, respectively. For a discrete signal, the decomposition coefficients of WP can be computed iteratively by [6]

$$\begin{aligned} d_{j+1}^{2n}(\tau) &= \sum_k h(k - 2\tau) d_j^n(k), \\ d_{j+1}^{2n+1}(\tau) &= \sum_k g(k - 2\tau) d_j^n(k). \end{aligned} \tag{4}$$

Fault features can be obtained from WPT energy [17, 19–22]. WPT energy contains the wealth of fault knowledge, and its fluctuation in a specific component corresponds to the occurrence of a specific fault [20]. Compared with the original signal, the length of $d_j^n(k)$ becomes shorter. As the number of decomposition level increases, the length of $d_j^n(k)$ becomes shorter and shorter. Too short of the length of $d_j^n(k)$ will cause WPT energy to distort. Instead of WPT coefficients, reconstructed signals using WPT coefficients are used to calculate WPT energy. After a signal is decomposed into several subsignals, the energy of each node can be calculated by

$$E_j^n = \sum_k \left[s_j^n(k)\right]^2, \tag{5}$$

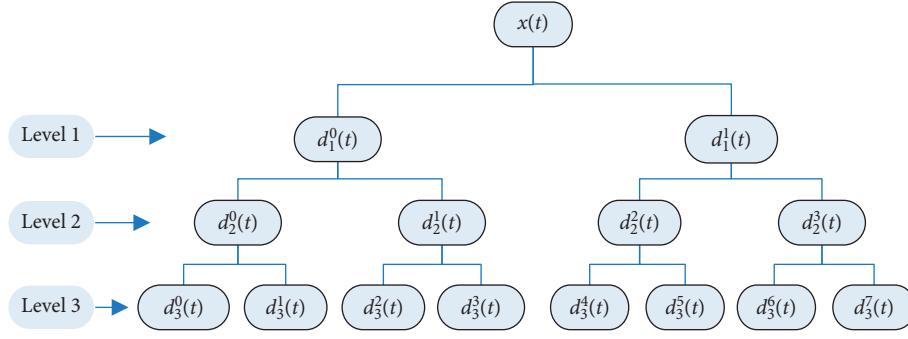where $s_j^n(k)$ denotes the reconstructed signal from the WPT coefficients $d_j^n(k)$.

FIGURE 1: Illustration of the wavelet packet transform process.

For the level j, all the energies of nodes are combined to form a vector:

$$E_j = \left[E_j^0, E_j^1, \cdots, E_j^{2^j-1} \right]. \tag{6}$$

To normalize the energy vector, (6) is rewritten as

$$\overline{E}_j = \left[\overline{E}_j^0, \overline{E}_j^1, \cdots, \overline{E}_j^{2^j-1} \right], \tag{7}$$

where

$$\overline{E}_j^n = \frac{E_j^n}{\sum_{i=0}^{2^{j-1}} E_j^i}. \tag{8}$$

Different levels of decomposition result in large differences in energy vector $\overline{E}_j$. For example, the WPT energy of a signal provided by the CWRU Bearing Data Center [23] is shown in Figure 2. The signal and its spectrum are shown in Figure 3. More layers of decomposition can reflect more details of the signal. Thus, we choose as many decomposition layers as possible.

Different wavelet bases have different WPT energy vectors. It was shown in [24] that the Daubechies 44 (db44) is the most similar mother wavelet function across the vibration signals. An example comparing the db3, db10, and db44 is shown in Figure 4. This figure shows that the WPT energy vectors can better reflect the real situation of the signal components by using db10 and db44. However, db44 takes more CPU time than the two others, and it is not the appropriate function for the fault identification algorithm [24]. Thus, db10 is selected to deal with the bearing signals.

3. Fault Data Automatic Clustering

The flowchart of fault data automatic clustering algorithm based on the WPT and a DNN is shown in Figure 5. Here, a time-domain signal is decomposed by using WPT to obtain some subsignals. After calculating the WPT energy, a training dataset which has the characteristic of giving the label randomly is configured. In fact, randomly given sample labels avoids preassigning specific category labels to samples, which in turn facilitates automatic classification of the samples. However, it is challenging to classify all samples successfully through a training process for the case of randomly assigned sample labels in the dataset.

The training dataset is roughly classified by clustering methods. The labels of samples are adjusted according to the clustering results. Next, a fine classification process based on a DNN is designed to achieve accurate classification of samples. A detailed description of the proposed algorithm is as follows.

3.1. Rough Classification. If the number of samples is not too large, the samples can be classified directly. However, for a dataset with a large number of samples, assigning labels to each sample will result in a large number of output-layer neurons, which will increase training time. Thus, a preliminary classification on the samples is required. Here, a rough classification process is designed by using clustering methods.

Clustering can group a set of samples into the same class with more similar features. There are already many clustering algorithms, such as distance-based clustering, density-based clustering, and fuzzy clustering. Here, the distance-based clustering is used to achieve rough classification. For two samples $x_1(n)$ and $x_2(n)$ with length N, the Euclidean distance is defined as

$$d_{12} = \sqrt{\sum_{k=1}^{N} \left[x_1(n) - x_2(n) \right]^2}. \tag{9}$$

If the distance between two samples is less than the user-specified tolerance, then we say that the two samples are similar. Similar samples can be classified into the same category and be assigned the same label. The rough classification ended after calculating all the distance of different two samples in a dataset and adjusting the label of samples. The result of clustering is to form a new dataset in which similar samples have the same label.

However, the Euclidean distance is a very brittle distance measure [25, 26]. There is no guarantee that all similar samples will be assigned to the same class. This is the reason that it is called rough classification. Thus, a fine classification is needed to get an accurate classification.

3.2. Fine Classification. When the rough classification is completed, a fine classification process based on a DNN is followed to achieve accurate classification of the samples.
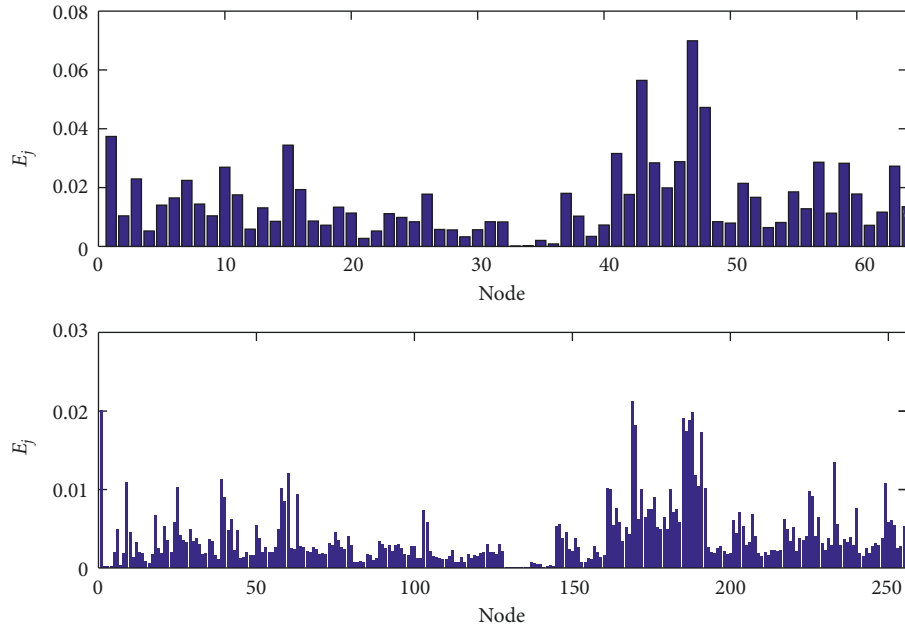
FIGURE 2: The WPT energy of the record 160FE provided by CWRU. The top and bottom figures show the WPT energy of levels 6 and 8, respectively.
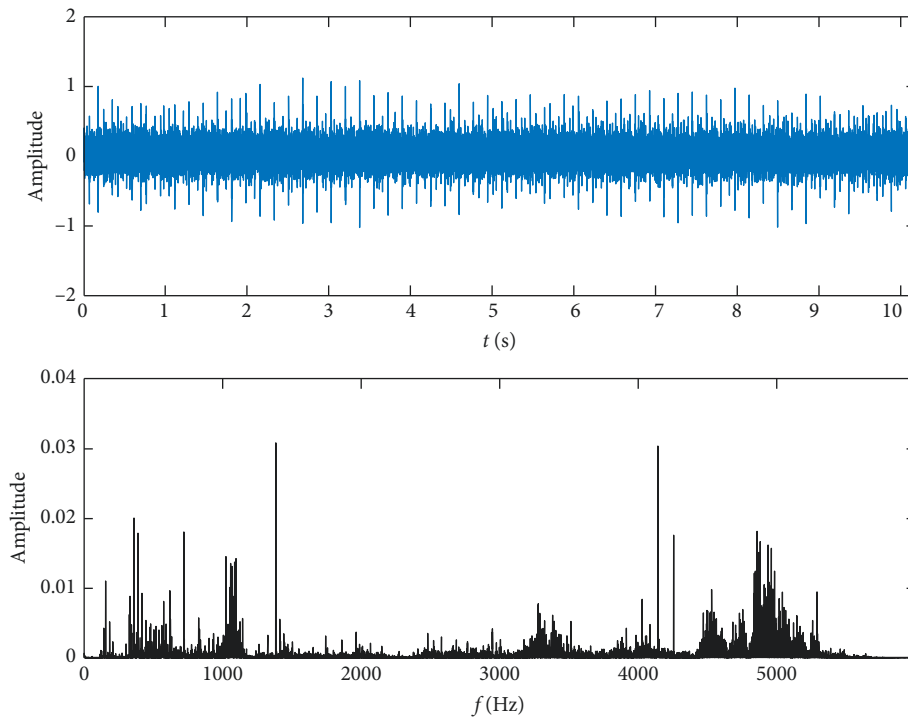

FIGURE 3: The signal of the record 160FE and its spectrum. The top and bottom figures show the raw data and their spectrum, respectively.

The flowchart of the fine classification process is shown in Figure 6. This figure shows that there are two main steps, training the DNN and assessing the training result to adjust the dataset.

First, a DNN is built, and it is trained with the training dataset. For a dataset with preset labels, the DNN can be trained with expected classification results if the training epoch is sufficiently large. Therefore, if the training is not successful, we increase the number of training epoch and train the DNN again. However, excessive training epoch consumes a lot of computation time. In this way, if the number of training epoch increases to a certain value and the training is still unsuccessful, we terminate the training process.

After the DNN is trained, the output of the DNN is as consistent as possible with the sample labels. This is
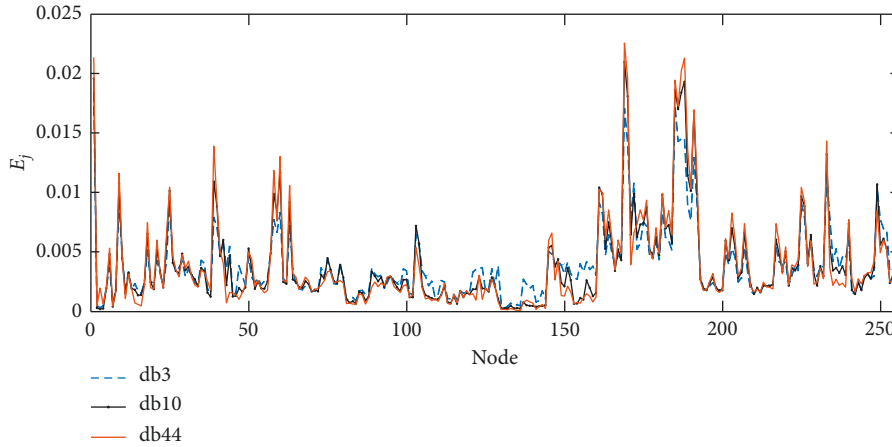
FIGURE 4: Illustration of the influence of mother wavelets.
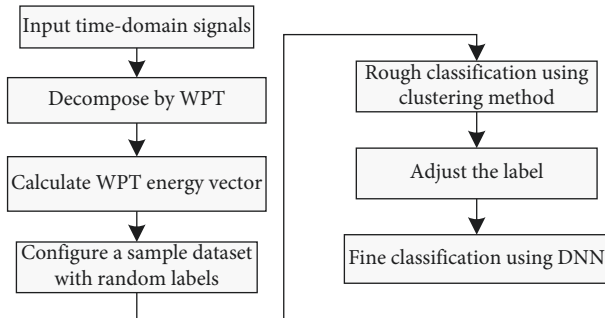

FIGURE 5: Flowchart of the proposed algorithm.

obviously not the result we want. In other words, we cannot achieve a truly effective classification of samples only through training the DNN. Thus, it is necessary to design a method or a strategy to evaluate the trained DNN.

A method is designed to assess the DNN classification result and to adjust the sample labels. We divide the raw data into smaller sections and calculate the WPT energy of each subsignal. Next, we build a testing dataset in which the sample labels are set according to the training results. As long as the length of subsignal is not too short, the WPT energy vector of the subsignal is similar to the raw data. An example is shown in Figure 7 where the distance function values between subsignal and raw data are calculated by using WPT energy. This figure shows that the fluctuations of the Euclidean distance values are small. At the same time, the Euclidean distance values are not zero, which indicates that the subsignal is different from the original signal. Thus, the subsignal can be used to assess the trained DNN and to test the generalization ability of the DNN.

If not all the testing results meet the requirement, we find out the misclassified samples. We change the sample labels and modify the sample dataset according to the test results. In other words, we adjust the training dataset by using the test result. The distance function shown in (9) is applied to prevent overadjustment of dataset. The distance function also prevents unrelated or less relevant samples from being classified as the same class.

If the dataset is changed, then the DNN is trained a second time. The fine process is ended after an iterative process of training and testing on the DNN. The result of the fine classification is that similar samples are grouped into the same class.

The above procedure shows that the entire procedure is unsupervised. The label of the sample is given randomly before starting the classification. Similar classes are continually merged in the rough classification and the fine classification processes. Therefore, the classification process is completely automatic without manual intervention. The effectiveness of the algorithm is tested by using the CWRU bearing data, which is provided in the next section.

4. Testing and Analysis

The proposed method is tested by using the CWRU bearing data with faults in 0.021 inch. These data include 12k drive-end, 48k drive-end, and 12k fan-end bearing data. The designed DNN has six layers. The number of decomposition layers of WPT is 8. Thus, the input layer of the DNN has 256 artificial neurons. The hidden layer of the DNN is set as 200-130-80-50. The number of output-layer neurons is determined by the types of sample labels, which can be adjusted dynamically. The weights of the DNN are initialized randomly. The learning rate and the batch size are set to 1. The activation function of neurons is the sigmoid function. The samples are selected randomly as the input to train the DNN.

In the rough classification process, all distances of different two samples in the dataset are calculated. The mean distance of all samples is calculated. If the distance between two samples is less than half of the mean distance, they are grouped together. In the fine classification process, the raw data are divided into small sections with lengths of 16,384, 8192, 4096, and 2048 to test the trained DNN.

For the 12k drive-end 0.021-inch bearing data, the 60 samples are randomly signed a label from 1 to 60. These samples are grouped into 48 categories after the rough classification. The rough classification results are further classified using the DNN, and the results are shown in
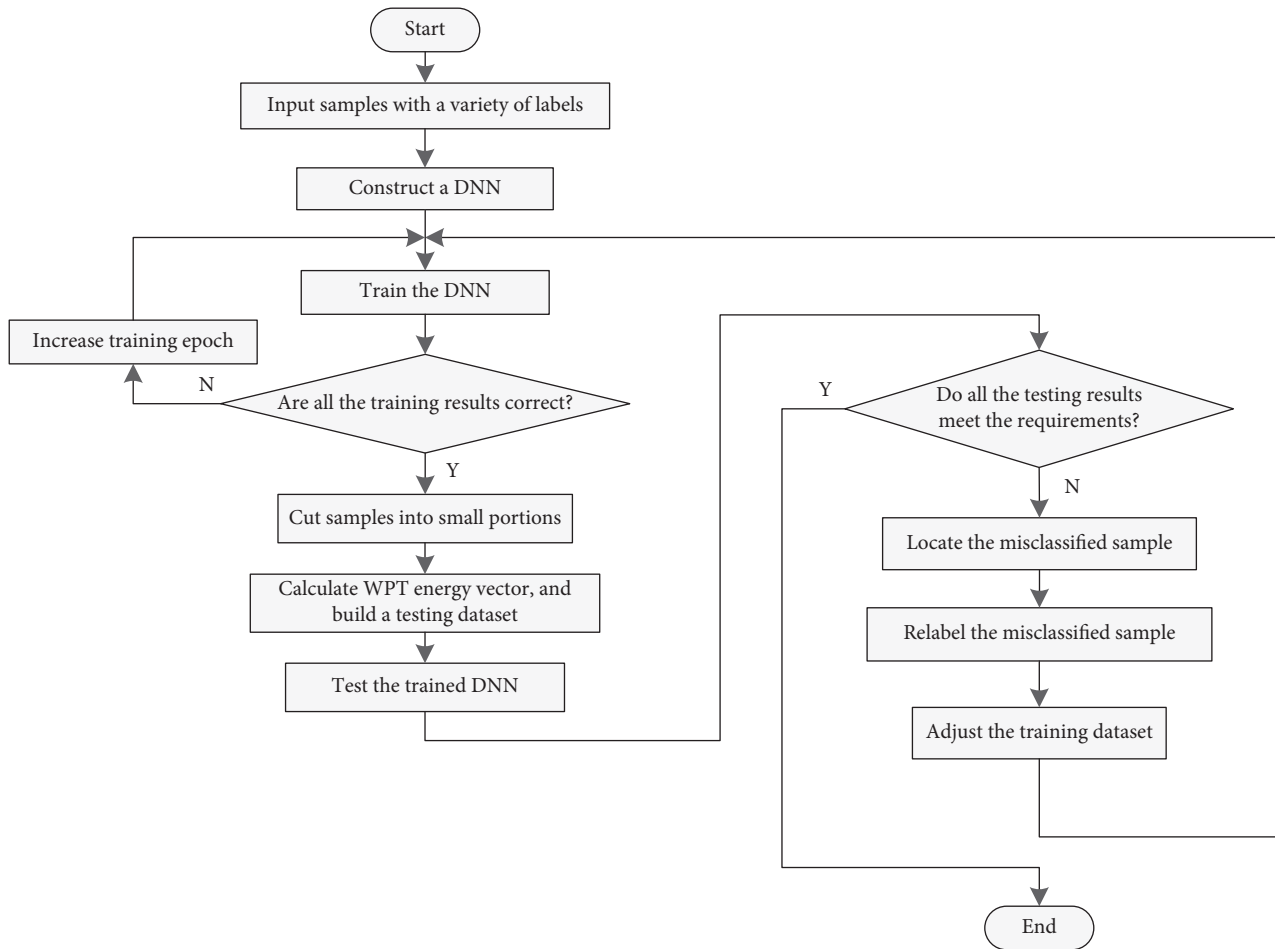
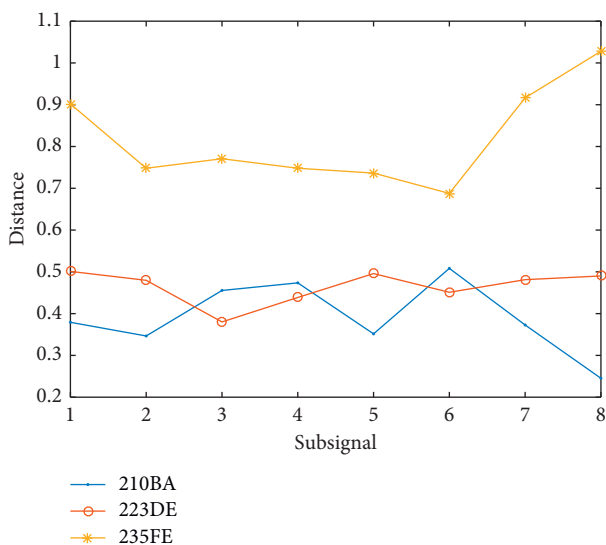Figure 6: Flowchart of the DNN training strategy.

Figure 7: The Euclidean distance values between subsignal and raw data by using WPT energy.

Figure 8. The raw data are divided into six classes. The WPT energy vectors of samples show significant differences in various categories, as shown in Figure 9.

The testing results on the trained DNN are shown in Table 1. We learn from this table that all the testing results are accurate except for the subsignals of length 2048. There are only 13 samples whose testing result is not consistent with the label of the subsignals of length 2048. The testing accuracy rates are 100%, 100%, 100%, and 99.63% for the samples with lengths 16,384, 8192, 4096, and 2048, respectively.

To further demonstrate the rationality of the classification, principal component analysis (PCA) [27, 28] was used to visualize the features extracted by the DNN. The first three principal components (PCs) of the features are shown in Figure 10. This figure shows that the features of the raw data are clearly gathered into six groups. However, as the sample length becomes shorter, the boundaries between certain gathering points begin to blur. Thus, we learn that the length of the testing samples cannot be too short. Combined with Figures 8–10 and Table 1, we can say that the overall classification results of the algorithm are satisfactory.

For the 48k drive-end 0.021-inch bearing data, the 40 records are given a unique label that is a random number between 1 and 40. These records are divided into 28 categories through the rough classification, and they are divided into 3 groups after the fine classification. The classification results are shown in Figure 11, and the WPT energy vectors
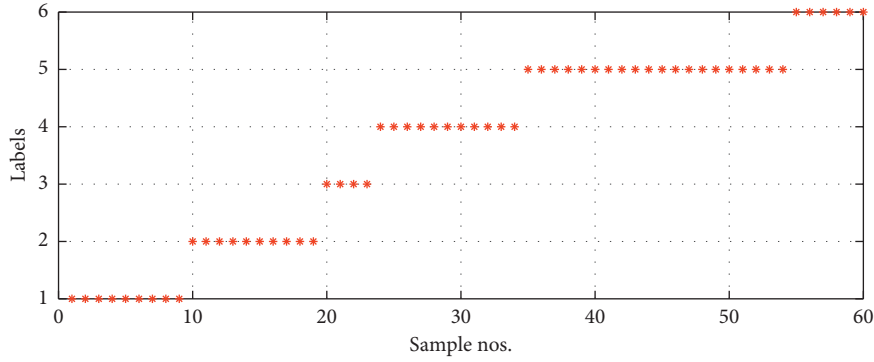
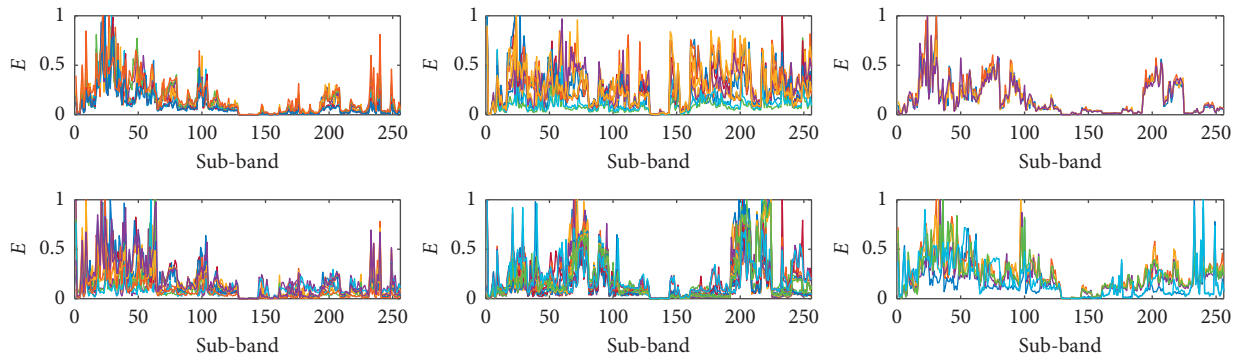FIGURE 8: Classification results for the 12k drive-end bearing data.

FIGURE 9: WPT energy vectors of the samples in various classes.

TABLE 1: The testing results on the 12k drive-end bearing data.

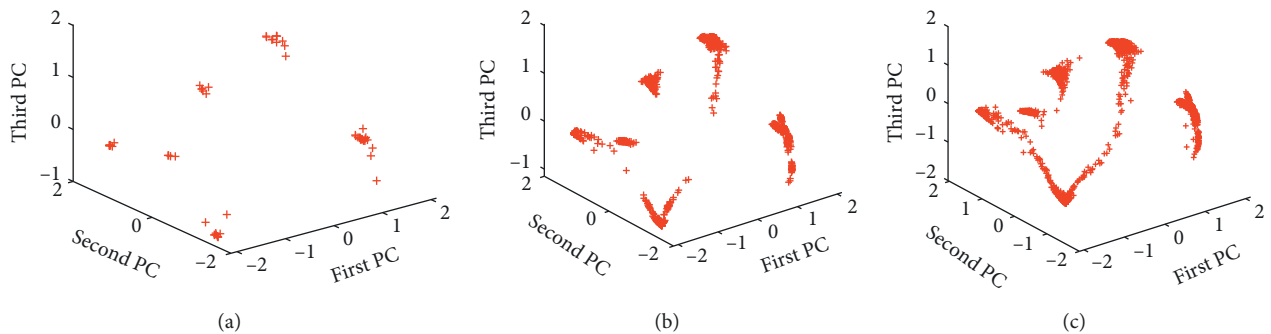| Length of subsignal	The number of samples	Misclassified samples	Accuracy rate (%)
16,384	420	0	100
8192	840	0	100
4096	1740	0	100
2048	3540	13	99.63

(a) (b) (c)

FIGURE 10: First three PCs of the features for the 12k drive-end bearing data. The results of (a) the raw data, (b) the subsignals of length 4096, and (c) the subsignals of length 2048, respectively.

of each class are shown in Figure 12. The WPT energy vectors of samples in different classes display distinct differences though there are somewhat similar for the first and third categories.

The testing results with the 48k drive-end bearing data are shown in Table 2. Similar to the testing results with the

12k drive-end bearing data, all the testing results are accurate except for the subsignals of length 2048 with 19 misclassified samples. The testing accuracy rates are 100%, 100%, 100%, and 99.77% for these four types of the subsamples.

To further prove the validity of the classification, the first three PCs of the features extracted the DNN are
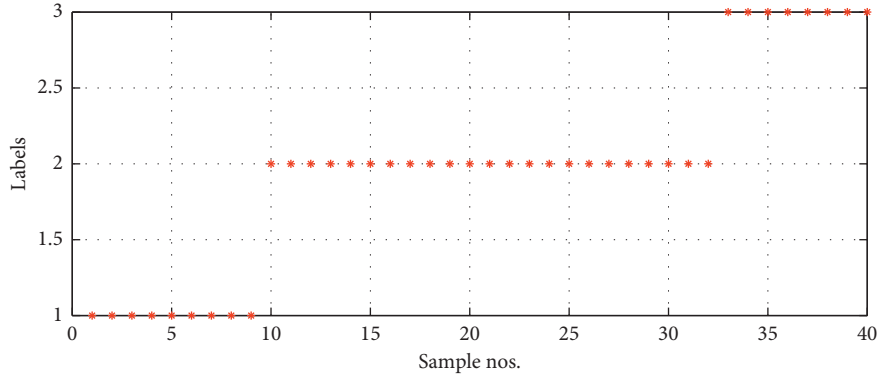
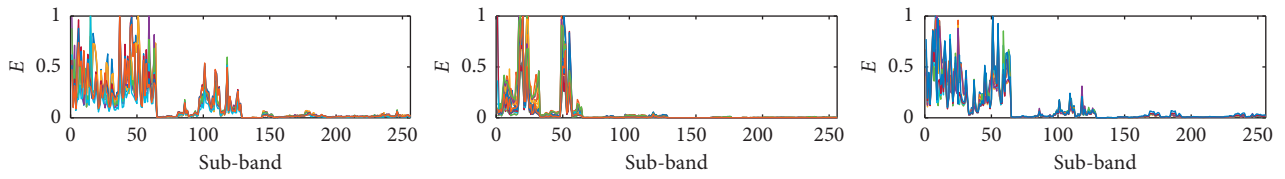FIGURE 11: Classification results for the 48k drive-end bearing data.

FIGURE 12: WPT energy vectors of the samples in various classes for the 48k drive-end bearing data.

TABLE 2: The testing results on the 48k drive-end bearing data.

Length of subsignal	The number of samples	Misclassified samples	Accuracy rate (%)
16,384	984	0	100
8192	2006	0	100
4096	4032	0	100
2048	8090	19	99.77

(a)   (b)   (c)
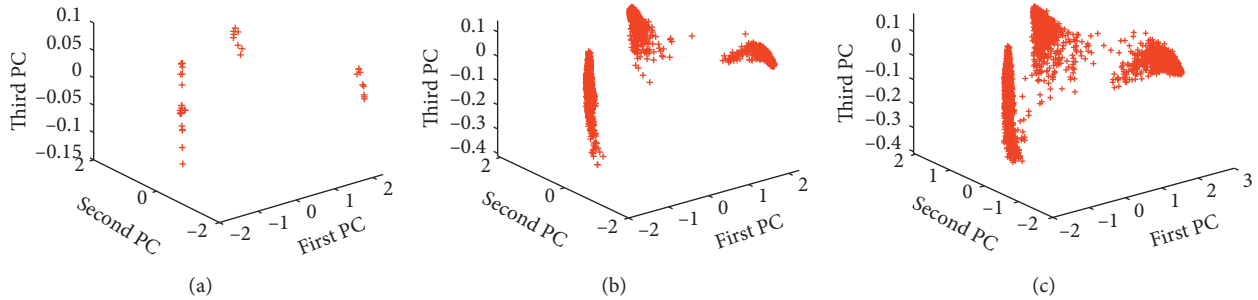
FIGURE 13: First three PCs of the features for the 48k drive-end bearing data. The results of (a) the raw data, (b) the subsignals of length 4096, and (c) the subsignals of length 2048, respectively.

presented in Figure 13. There is no overlap for the raw data features in different groups, but some discrete feature points appear between two classes. Overall, the boundaries between the three gathering points are still very clear. Thus, the proposed method offers satisfactory performance.

For the 12k fan-end 0.021-inch bearing data, the 36 records are given a unique label that is a random number between 1 and 36. These records are divided into 29 and 4 categories through the rough and the fine classifications, respectively. The classification results are shown in Figure 14, and the WPT energy vectors of each class are shown in

Figure 15 which displays a clear difference between the WPT energy vectors of samples for different classes.

The testing results are shown in Table 3. The results show that there are 2 and 13 misclassified samples for the subsignals of lengths 4096 and 2048, respectively. The testing accuracy rates are 100%, 100%, 99.81%, and 99.38% for these four types of the subsamples.

The first three PCs of the features extracted the DNN are presented in Figure 16. There are 4 gathering points displayed on the feature map of the raw data. Nevertheless, the boundaries between two gathering points on the feature map for the subsignal of length 4096 are blurred. Also, 3 gathering
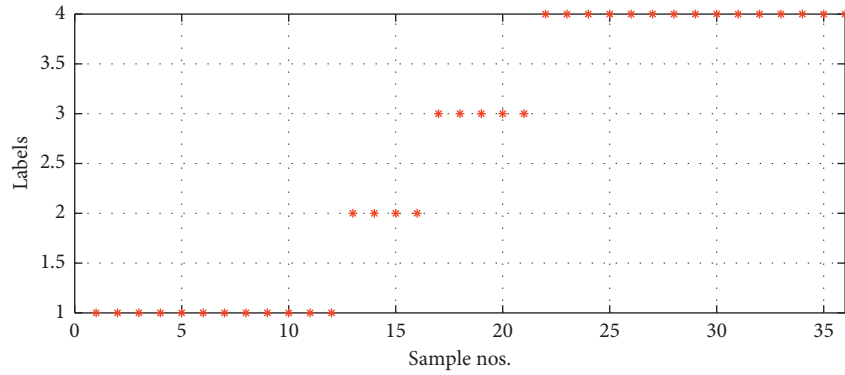
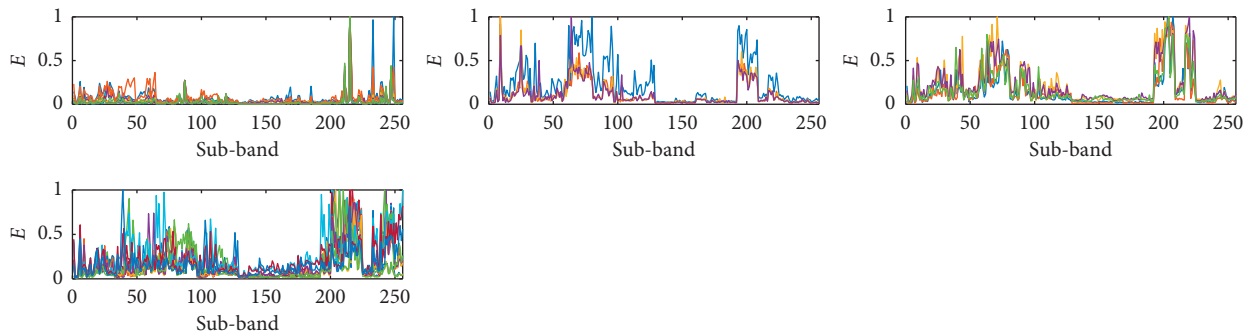FIGURE 14: Classification results for the 12k fan-end bearing data.

FIGURE 15: WPT energy vectors of samples in various classes for the 12k fan-end bearing data.

TABLE 3: The testing results on the 12k fan-end bearing data.

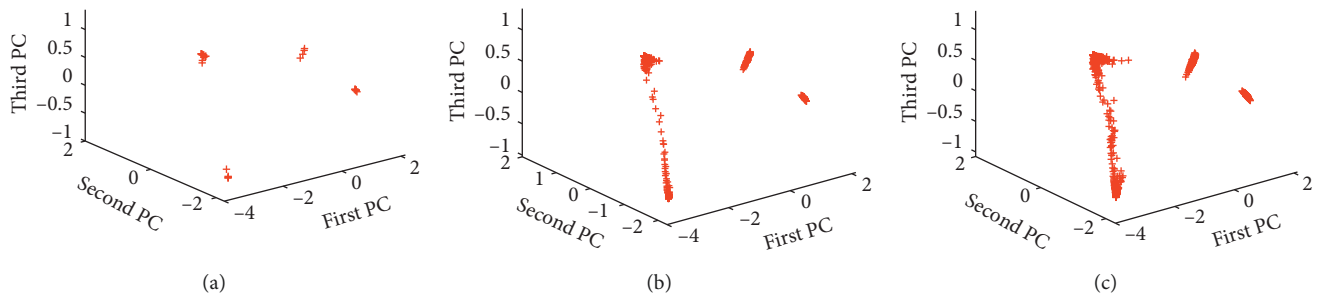| Length of subsignal	The number of samples	Misclassified samples	Accuracy rate (%)
16,384	252	0	100
8192	504	0	100
4096	1044	2	99.81
2048	2112	13	99.38

(a)

(b)

(c)

FIGURE 16: First three PCs of the features for the 12k fan-end bearing data. The results of (a) the raw data, (b) the subsignals of length 4096 and (c) the subsignals of length 2048, respectively.

points occur on the feature map for the subsignal of length 2048. There are two class features connected into one piece. This also indicates that the length of the subsignals cannot be too short.

As mentioned earlier, the testing dataset is different from the training dataset. That is, the samples used for testing differ from the training samples. Thus, we obtain that the proposed method has strong generalization ability.

5. Conclusion

An automatic clustering method of bearing fault data based on wavelet and deep neural network is proposed. The fault characteristics of signals are extracted by the WPT. The clustering process consists of a rough classification and a fine classification. The rough classification is designed to achieve preliminary clustering of samples. A fine classification based

on DNN is further to classify the samples accurately. The original label of the training sample is randomly given, and the labels of samples update automatically in the subsequent process without manual intervention. The entire clustering process is completely automated and does not rely on human expertise, which helps us to achieve intelligent diagnosis of faults.

The proposed method is tested by using the CWRU bearing data with faults in 0.021 inch. The testing results show that the proposed method has good performance and can be used for automatic classification of bearing fault data. However, in practical applications, the operating state of the device is very complicated. It is undeniable that some samples may not be classified into the expected category. The proposed method still requires a lot of data for further testing.

In engineering practice, different purposes lead to different expectations for classification results. In real application, manually adjusting the clustering results can be made to achieve accurate classification according to a certain purpose. Applying the proposed method to engineering practice and continuously improving and perfecting it, as well as exploring better methods, are the work needed to be done in the future.

Data Availability

The data used to support the findings of this study are available at [23].

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Acknowledgments

References

[1] Z. Q. Chen, S. C. Deng, X. D. Chen, C. Li, R. V. Sanchez, and H. F. Qin, "Deep neural networks-based rolling bearing fault diagnosis," *Microelectronics Reliability*, vol. 75, pp. 327–333, 2017.

[2] X. X. Qi, J. W. Ji, and X. W. Han, "Fault diagnosis methods-of rolling bearing: a general review," *Key Engineering Materials*, vol. 480-481, pp. 986–992, 2011.

[3] M. Cerrada, R. V. Sánchez, C. Li et al., "A review on data-driven fault severity assessment in rolling bearings," *Mechanical Systems and Signal Processing*, vol. 99, pp. 169–196, 2018.

[4] S. Kokil and S. Y. Gajjal, "Review on vibration based condition monitoring of rolling element bearing by numerical method and its validation by experimental approach," *International Journal of Scientific and Engineering Research*, vol. 5, no. 4, pp. 521–524, 2014.

[5] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—a tutorial," *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 485–520, 2011.

[6] Q. B. He, "Vibration signal classification by wavelet packet energy flow manifold learning," *Journal of Sound and Vibration*, vol. 332, no. 7, pp. 1881–1894, 2013.

[7] Z. W. Wang, Q. H. Zhang, J. B. Xiong, M. Xiao, G. X. Sun, and J. He, "Fault diagnosis of a rolling bearing using wavelet packet denoising and random forests," *IEEE Sensors Journal*, vol. 17, no. 17, pp. 5581–5588, 2017.

[8] Z. Q. Huo, Y. Zhang, P. Francq, L. Shu, and J. F. Huang, "Incipient fault diagnosis of roller bearing using optimized wavelet transform based multi-speed vibration signatures," *IEEE Access*, vol. 5, pp. 19442–19456, 2017.

[9] P. K. Kankar, S. C. Sharma, and S. P. Harsha, "Rolling element bearing fault diagnosis using wavelet transform," *Neurocomputing*, vol. 74, no. 10, pp. 1638–1645, 2011.

[10] S. Ekici, S. Yildirim, and M. Poyraz, "Energy and entropy-based feature extraction for locating fault on transmission lines by using neural network and wavelet packet decomposition," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2937–2944, 2008.

[11] R. Q. Yan, R. X. Gao, and X. F. Chen, "Wavelets for fault diagnosis of rotary machines: a review with applications," *Signal Processing*, vol. 96, pp. 1–15, 2014.

[12] J. L. Chen, Z. P. Li, J. Pan et al., "Wavelet transform based on inner product in fault diagnosis of rotating machinery: a review," *Mechanical Systems and Signal Processing*, vol. 70-71, pp. 1–35, 2016.

[13] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, pp. 504–507, 2006.

[14] Y. LeCun, Y. Bengio, and G. E. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[15] F. T. Wang, B. S. Dun, X. F. Liu, Y. H. Xue, H. K. Li, and Q. K. Han, "An enhancement deep feature extraction method for bearing fault diagnosis based on kernel function and autoencoder," *Shock and Vibration*, vol. 2018, Article ID 6024874, 12 pages, 2018.

[16] R. Zhao, R. Q. Yan, Z. H. Chen, K. Z. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 115, pp. 213–237, 2019.

[17] H. D. Shao, H. K. Jiang, F. Wang, and Y. N. Wang, "Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet," *ISA Transaction*, vol. 69, pp. 187–201, 2017.

[18] Y. Han, B. P. Tang, and L. Deng, "Multi-level wavelet packet fusion in dynamic ensemble convolutional neural network for fault diagnosis," *Measurement*, vol. 127, pp. 246–255, 2018.

[19] X. X. Ding and Q. B. He, "Energy-fluctuated multiscale feature learning with deep ConvNet for intelligent spindle bearing fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 8, pp. 1926–1935, 2017.

[20] M. Gan, C. Wang, and C. A. Zhu, "Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 72-73, pp. 92–104, 2016.

[21] M. Sohaib, C. H. Kim, and J. M. Kim, "A hybrid feature model and deep-learning-based bearing fault diagnosis," *Sensors*, vol. 17, no. 12, p. 2876, 2017.

[22] T. W. Rauber, F. de A. Boldt, and F. M. Varejão, "Heterogeneous feature models and feature selection applied to bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 637–646, 2015.

[23] CWRU, Case Western Reserve University Bearing Data Center Website, Cleveland, OH, USA, http://csegroups.case.edu/bearingdatacenter/home.

[24] J. Rafiee, M. A. Rafiee, and P. W. Tse, "Application of mother wavelet functions for automatic gear and bearing fault diagnosis," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4568–4579, 2010.

[25] F. K. Chan, A. W. Fu, and C. Yu, "Haar wavelets for efficient similarity search of time-series: with and without time warping," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 3, pp. 686–705, 2003.

[26] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowledge and Information Systems*, vol. 7, no. 3, pp. 358–386, 2005.

[27] 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. 72-73, pp. 303–315, 2016.

[28] H. D. Shao, H. K. Jiang, F. Wang, and H. W. Zhao, "An enhancement deep feature fusion method for rotating machinery fault diagnosis," *Knowledge-Based Systems*, vol. 119, pp. 200–220, 2017.