IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 20, 2018, accepted November 3, 2018, date of publication November 16, 2018,
date of current version December 18, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2880990

Bearing Fault Automatic Classification Based on Deep Learning

YANLI YANG, PEIYING FU, AND YICHUAN HE
Tianjin Key Laboratory of Optoelectronic Detection Technology and Systems, Tianjin Polytechnic University, Tianjin 300387, China

Corresponding author: Yanli Yang (yyl070805@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61401305 and in part by the Natural Science Foundation of Tianjin, China, under Grant 15JCYBJC16500.

ABSTRACT An automatic classification method based on deep learning for bearing fault diagnosis is proposed. The method is designed with the ability of faulty signal automatic clustering without human knowledge. A dataset in which each sample is given a random label is configured after extracting the features of vibration signals from the frequency domain. The dataset is used to train a deep neural network (DNN) to obtain the initial classification. The classification results are assessed by testing the subsignals extracted from the raw data, and the sample labels are modified according to the evaluation result. The modified dataset is used to train the DNN a second time. Samples with characteristic faults are clustered in various classes after iterating the DNN training and testing. The proposed method is tested with the bearing data provided by the Case Western Reserve University (CWRU) Bearing Data Center, which is a standard reference to test fault detection algorithms. The 12k drive end, 48k drive end, and 12k fan end CWRU bearing data are classified into 7, 6, and 4 groups, respectively. The testing results show that the proposed method can achieve automatic clustering for vibration signals with a variety of faults.

INDEX TERMS Rotating machinery, fault diagnosis, deep learning, artificial neural network, bearing, fast Fourier transforms, feature extraction, classification algorithms.

I. INTRODUCTION

Rotating machinery is one of the most widely used types of machinery in industry [1], and its reliability, maintainability, and security have received substantial attention, with growing safety awareness. Rolling element bearings are the most prevalent components in rotating machinery [2], [3]. Such bearings are the primary components in many heavy-duty machines, and their integrity often has a significant influence on the performance and reliability of the machinery in which they are installed [4], [5]. Damage to the bearing can cause faults in the machine, potentially leading to severe accidents [6], [7]. The failure of rolling element bearings is one of the most frequent reasons for machine breakdown [3]. Thus, bearing fault diagnosis has become an important research direction in the field of equipment failure detection.

In bearing fault diagnosis, diagnostic results are typically based on a large amount of data measured in industrial production cases. With the rapid development of the industrial Internet of Things, fault diagnosis has entered the era of big data. Deep learning, introduced by Hinton and Salakhutdinov [8], attempts to model high-level representations behind data and to classify patterns via stacking multiple layers of information processing modules in hierarchical architectures. Deep learning can be used to extract useful knowledge, make appropriate decisions from big data, and help achieve intelligent machine health monitoring [9]. This approach has already had great success in speech recognition, visual object recognition, object detection and many other domains [10]. For example, AlphaGo, an artificially intelligence algorithm based on deep learning that was designed to play the game Go, has already defeated a world-class master player. Relative to traditional diagnosis algorithms, the main advantage of deep learning is that the fault features are learned via a general-purpose learning procedure instead of engineered manually or with prior knowledge of the signal processing techniques [4].

For fault diagnosis based on a deep neural network (DNN), deep Boltzmann machines (DBMs), deep belief networks (DBNs) and stacked autoencoders (SAEs) are the most popular models of deep learning. By combining two-layer sparse autoencoder neural networks with a DBN, a multisensor feature fusion method was proposed in [11] for

bearing fault diagnosis. By using a stack of autoencoders, a DNN-based intelligent method was proposed in [12] for diagnosing rolling element bearing faults from the frequency domain. This method can mine fault characteristics from the frequency spectra adaptively for various diagnosis issues and effectively classify the health conditions of machinery. The stacked denoising autoencoder (SDAE) was adopted in [4] to diagnosis bearing faults; this approach can adaptively mine salient fault characteristics and effectively identify the health states with high diagnosis accuracy and strong robustness. A deep autoencoder feature learning method with a new loss function was developed in [13] to diagnose rotating machinery faults. An integrated deep fault recognizer model based on SDAE was applied to denoise the raw signals and represent fault features during the fault pattern diagnosis of bearing rolling faults [14]. A fault diagnosis method based on the stacked SAE was proposed in [15], and a multimode method based on deep learning was proposed in [16] for fault classification. In [6], three DNN models, namely, the DBM, DBN and SAE, were found to enable a high-accuracy and reliable fault diagnosis of rolling bearings through a comparative analysis of their performance. In addition, convolution neural networks (CNNs), which have had great success in image recognition, have been used for fault classification [17], [18], and for bearing fault diagnosis and fault severity evaluation [2].

Although bearing fault diagnosis based on deep learning has had some success, certain issues remain to be resolved. The amount of time consumed is one problem. A DNN with too many input units and hidden layers can have higher computational costs (although such a network probably also offers better diagnosis performance). Furthermore, a DNN with too few input units may not have enough power to learn the global structure information [13]. Finding the optimal parameters of a DNN is a challenging task. Thus, feature extraction and dimensionality reduction of raw data are necessary steps for DNN-based fault diagnosis of machinery [6]. In fact, an integrated dimensionality reduction method combined with feature extraction techniques that yield a more sensitive and lower dimensional feature set is a way to reduce the computation burden for fault diagnosis and improve the separability of samples [19]. Note also that many methods do not perform well in practical applications. Some approaches work properly in the laboratory, but the performance may decrease dramatically when used outside laboratory settings [20].

The consensus is that the expert knowledge of humans is helpful for fault diagnosis, even for neural networks [4], [5], [12], [14], [16], [21]. For example, a two-stage detector [22] and a semi-supervised diagnostic framework [23] were proposed for fault diagnosis. Deep learning applied to fault diagnosis performs well for presorted sample cases [4], [11]–[14]. An unsupervised method based on the SDAE that can be trained in an unsupervised way was proposed in [21] to extract the fault characteristic; however it still requires labeled data for fine-tuning the fault diagnosis model. Furthermore, expert datasets, when available, are often expensive or unreliable [24]. Thus, the development of a fault diagnosis system with an autonomous learning function merits specific study.

Recently, differing from AlphaGo, which depends on human expert moves, AlphaGo Zero was developed to play Go without incorporating human knowledge [24]. Inspired by the success of AlphaGo Zero, we propose an automatic classification method based on deep learning for bearing fault diagnosis. Considering the computational constraints, we reduce the high data dimension to a small dimension before inputting the data to the DNN to train. After feature extraction from the frequency domain, instead of setting a specific label for each sample, we specify a random label to build a sample dataset. It is clear that the dataset is not based on an expert or previous knowledge. A training result assessment method is designed to evaluate the classification result and adjust the labels in the training dataset. The method is tested by using bearing data provided by the Case Western Reserve University (CWRU) Bearing Data Center [25], which has become a standard reference to test fault detection algorithms.

Deep learning is combined with different neural networks to create various network structures, such as DNNs, DBNs and recurrent neural networks (RNNs). Here, we use a DNN as the deep learning model for analysis of the bearing data by using the DeepLearn Toolbox [26]. The proposed algorithm is introduced with the following steps. In Section II, the characteristics of the CWRU bearing data are analyzed. Section III gives a detailed description of the proposed fault automatic classification method based on a DNN. In Section IV, testing of the proposed method with the CWRU bearing data is presented. Finally, the conclusions are drawn in Section V.

II. ANALYSIS OF CWRU BEARING DATA

The data provided by CWRU were collected from the bearing testing platform shown in Fig. 1. The platform comprises a 2 HP motor, a torque transducer and encoder, and a dynamometer. The test bearings, SKF bearing (SKF deep groove ball bearings: 6205-2RS JEM and 6203-2RS JEM) and NTN equivalent bearing support the motor shaft. By using electro-discharge machining, fault diameters of 0.007, 0.014, 0.021 inches on the SKF bearings and 0.028, 0.04 inches on the NTN bearings were seeded. The motor speed ranged from 1730 to 1797 rpm. The acceleration was measured in the vertical direction on the drive end bearing housing (DE), and in some tests, the acceleration was measured in the vertical direction on the fan end bearing housing (FE) and on the motor-supporting base plate (BA) [3]. The data were saved as different files in MATLAB format. Each file contains one or more of the recorded DE, FE, and BA. Detailed information is shown on the CWRU bearing data center website [25]. A benchmark to test the fault diagnosis method was provided in [3] through studying the CWRU dataset by using envelope analysis, cepstrum prewhitening and benchmark methods.

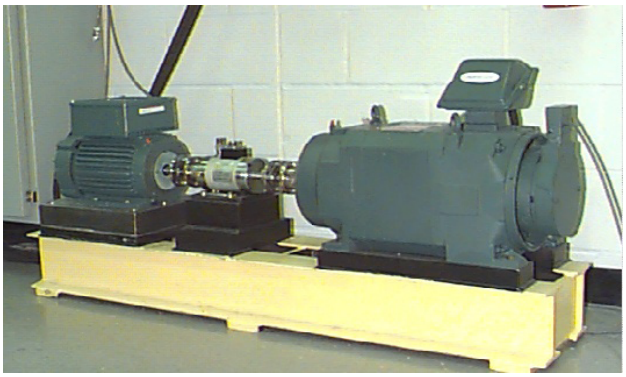

FIGURE 1. Bearing testing platform [25].

TABLE 1. Bearing defect frequency [3], [25].

Position	Model number	Defect frequencies (multiple of running speed in Hz)			
		BPFI	BPFO	FTF	BSF
Drive end	6205-2RS JEM	5.4152	3.5848	0.39828	2.3568
Fan end	6203-2RS JEM	4.9469	3.0530	0.3817	1.9937

The data were recorded at 12 kHz and 48 kHz for the drive end and 12 kHz for the fan end in bearing experiments and contain normal and faulty bearing measurements. The faulty data include three fault types, namely, the inner race (IR), outer race (OR), and rolling element (ball). The defect frequency of the IR, OR, cage train and rolling element are listed in Table. 1. The bearing fault frequency includes the ball pass frequency of the outer race (BPFO), ball pass frequency of the inner race (BPFI), fundamental train frequency (FTF), and ball spin frequency (BSF). The defect characteristic frequency in the sampled signals is the product of these frequencies and the main frequency, which is proportional to the running speed.

A characteristic of the CWRU bearing data is that the length is long and varied, as listed in Appendix Table 2. The data length is not an integer multiple of 2. The length changes from 120,801 to 122,917 for the 12k drive end bearing data. The maximum length of the 12k fan end bearing is 122,269, and the minimum length is 120,617, except for the datasets 277 and 278 with a length of 10,000. The maximum length of the 48k drive end bearing is 491,446 and the minimum length is 124,602, except for file 174 with a length of 63,788. The normal data length changes from 243,938 to 485,643. We use the largest integer data of 2 times the training data to calculate the fast Fourier transform (FFT) conveniently because the CWRU bearing data length is not an integer multiple of 2. The remainder is chosen as the test data. For example, a signal length of 65,536 from dataset 105 is used for the training data, and the remaining signal length of 55,729 is used as the test data.

Another characteristic is that many of the datasets are dominated by non-classical features of fault identification [3]. Not all the faults can be well recognized. Smith and Randall [3], for example, classified the data into three types. One type is clearly detected faults. The fault frequency is reflected clearly in certain data either in the time domain or the frequency domain, such as in file 105. The second type is probable or potential diagnosis of faults. The fault frequency is not dominant in the spectrum, such as in file 172, the spectrum of which is shown in Fig. 2. The third type is that the data cannot be detected with the specified bearing fault or with other problems or when severe noise exists, such as in file 118, the spectrum of which is shown in Fig. 3.

The third characteristic is that the spectra of measurements in the same file are different, and the difference can be large. Using files 105 to 108 as an example, the spectra are shown in Fig. 4. Here, the three columns correspond to BA, DE, and FE, and the four rows correspond to files 105, 106, 107, and 108. The spectra have a similar shape in the same column, but the shapes vary greatly across different columns. In other words, spectra have a similar shape for the same record but different shapes for different records. Thus, the features in the frequency domain belong to a variety mode although they are sorted into the same fault type. It was noted in [27] that the datasets with slight faults, medium faults, and severe faults in the IR, ball and OR have different spectra.

In addition, certain frequency components are intermittent or sometimes large or small. By using file 105 as an example, the short-time Fourier transform (STFT) is shown in Fig. 5. Here, spikes exist, and the amplitude of certain components is up and down. This phenomenon is more prominent when the data length is short.

The above analysis shows that the composition of the CWRU data is complex and varied. A well-designed classification method is necessary to classify these records into proper classes. In addition, the CWRU bearing data can be used to test the ability of the DNN in the field of fault diagnosis for real applications.

III. FAULT AUTOMATIC CLASSIFICATION BASED ON DNN

It was reported previously [4], [11]–[14], [28], [29] that a DNN can provide successful classification for datasets with given labels. However, labels based on expert knowledge are scarce. In real engineering applications, much unlabeled data exist. Therefore, it is necessary to develop a method to automatically classify fault samples.

To classify data with a variety of fault modes automatically, an algorithm based on a DNN is proposed that has four primary steps. The time-domain signal is transformed into the frequency domain, and features are extracted from the spectrum to reduce the data dimension. Then, a sample dataset is built after each sample is given a random label. By using this dataset, the constructed DNN is trained to achieve a rough classification, so called because some samples are not assigned the correct class labels. The DNN classification result is assessed, and the sample labels are adjusted to obtain a better classification result. A flow diagram of the proposed

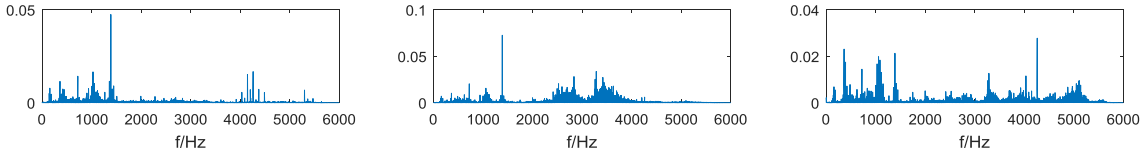

FIGURE 2. Spectra of file 172. The three columns are records of BA, DE, and FE.

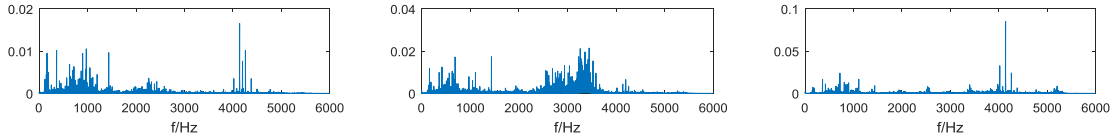

FIGURE 3. Spectra of file 118. The three columns are records of BA, DE, and FE.

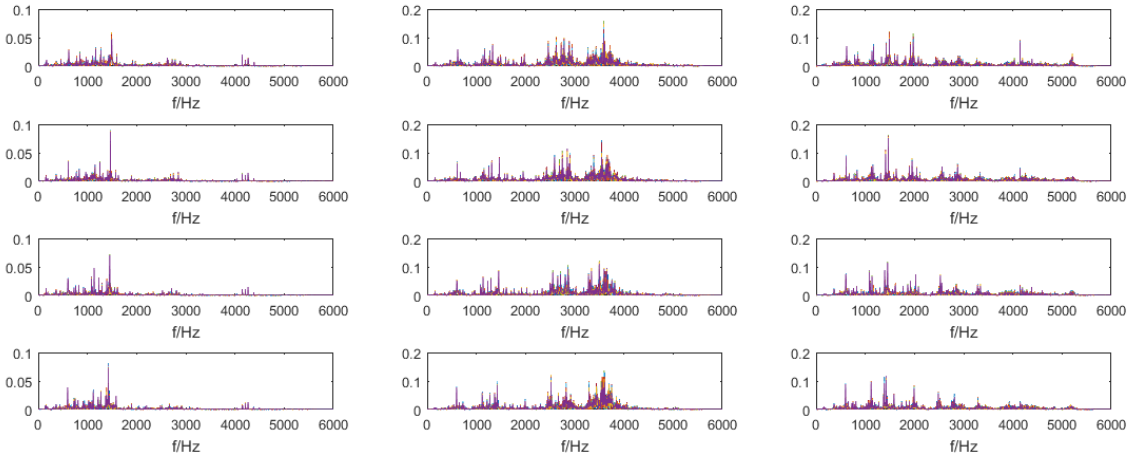

FIGURE 4. IR spectra for a fault diameter of 0.007 inch. The three columns are records of BA, DE, and FE, and the four rows are files 105, 106, 107, and 108.

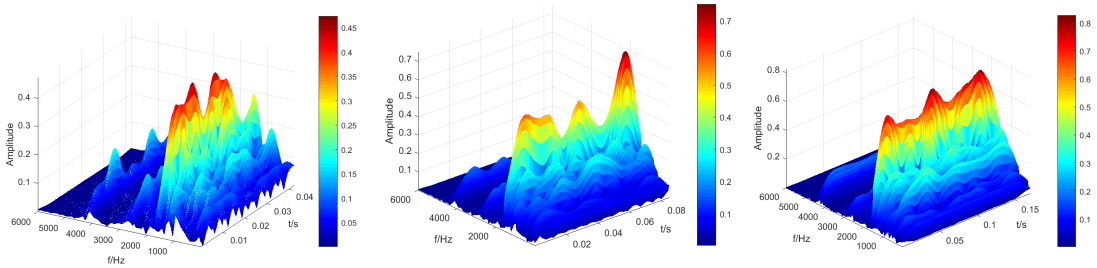

FIGURE 5. STFT of part of the data of file 105DE. The three subfigures from left to right show the STFT results of extracting lengths of 512, 1024 and 2048 points from file 105DE.

algorithm is shown in Fig. 6. A detailed description of the proposed algorithm is presented as follows.

A. SAMPLE DATASET

The dataset is known to be important for neural networks. Inspired by previous work [12] in which frequency domain data were used as the network input to detect faults in the rolling element bearings and planetary gearboxes, the spectrum data instead of the time-domain data are selected to configure a dataset. For a long measurement time record, many data points are obtained after the Fourier transform. If all the spectrum data are used as the input for the DNN, then training the DNN is relatively time-consuming. Take the labeled 105 data as an example: the data length is 65,536, and the length of the spectrum is 32,768. The input layer is 32,768 if the spectrum is directly used as an input. If the DNN is set as 32768-10000-3000-1000-300-100, then the time for each epoch is approximately 790 s. Thus, the data

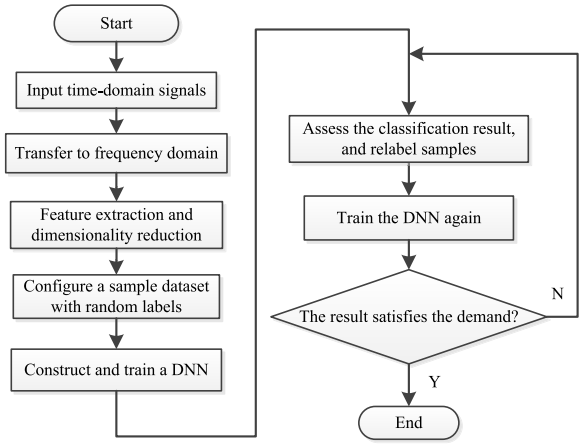

FIGURE 6. Flowchart of the proposed algorithm.

of dimensionality must be reduced to reduce the computing time.

Many spectral lines appear on the spectrum, as shown in Fig. 4. Although the spectrum is a useful tool for fault diagnosis, certain characteristic frequencies, not all of them spectral lines, contribute to diagnosing faults in real applications. Thus, we divide the spectrum into several sections and determine the characteristic value of each section. In other words, we add windows on the spectrum and determine the characteristic value in each window. Too few windows will cause feature loss, but too many will result in increased computing time. Determining the appropriate number of windows is thus challenging. It is clear that the main frequency and its harmonics are helpful for identifying faults. For bearings, BPFO, BPFI, FTF, and BSF are the main fault frequencies [3]. To prevent the characteristic frequency of the bearing and its multiple frequency from falling within the same window, we design a method to determine the window length

$$l_w < \min\left(\frac{f_m}{\Delta f}, \frac{f_{IR}}{\Delta f}\right), \tag{1}$$

where f_m denotes the main frequency, f_{IR} represents the characteristic frequency shown in Table 1, and Δf is the frequency resolution. In (1), $\Delta f = f_s/n$ where f_s denotes the sample frequency. The main frequency is calculated by using the rotating speed n. Then, the number of windows is

$$n_w > \left[\frac{n/2}{l_w}\right]. \tag{2}$$

where $[\cdot]$ denotes the rounding operation. Introducing (1) into (2), we obtain

$$n_w > \left[\frac{f_s/2}{\min(f_m, f_{IR})}\right]. \tag{3}$$

The characteristic value of each window can be the maximum point or the energy value.

After feature extraction from the spectrum, a dataset is configured. To achieve automatic classification of the samples, a dynamic dataset is necessary, meaning that the sample labels

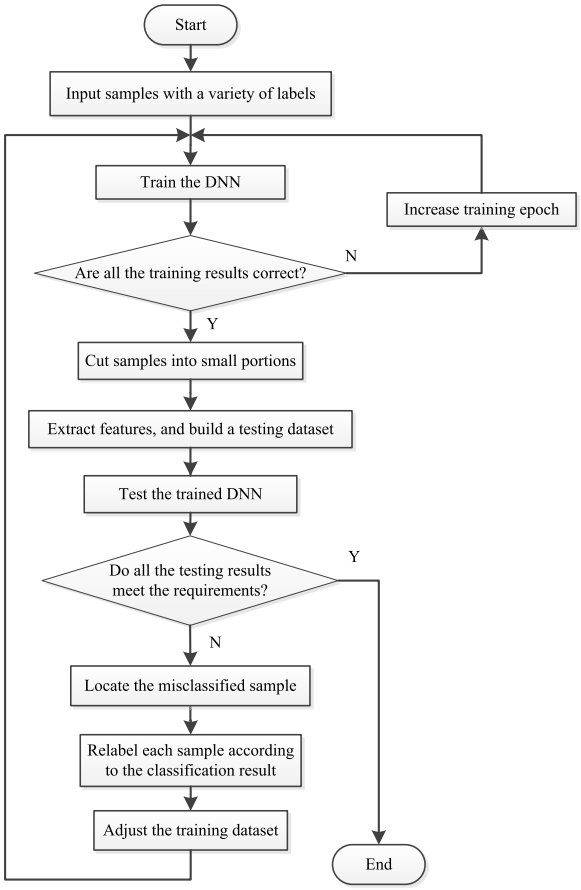

FIGURE 7. Flowchart of the DNN training strategy.

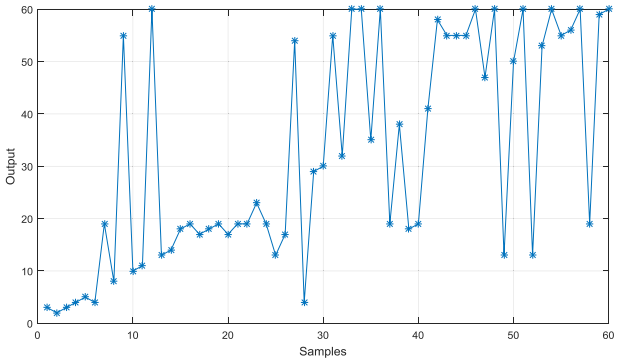

FIGURE 8. DNN output example when the training epoch is insufficiently large.

in the dataset must be modified in the DNN training process. For the initial sample dataset, each sample is given a random label. In fact, the use of random sample labels avoids prefixing a specific category label to a sample.

B. SAMPLE AUTOMATIC CLASSIFICATION

For a dataset with pre-set labels, the DNN can be trained with expected classification results. However, it is challenging to classify all the samples successfully through a training process for the case of randomly assigned sample labels in

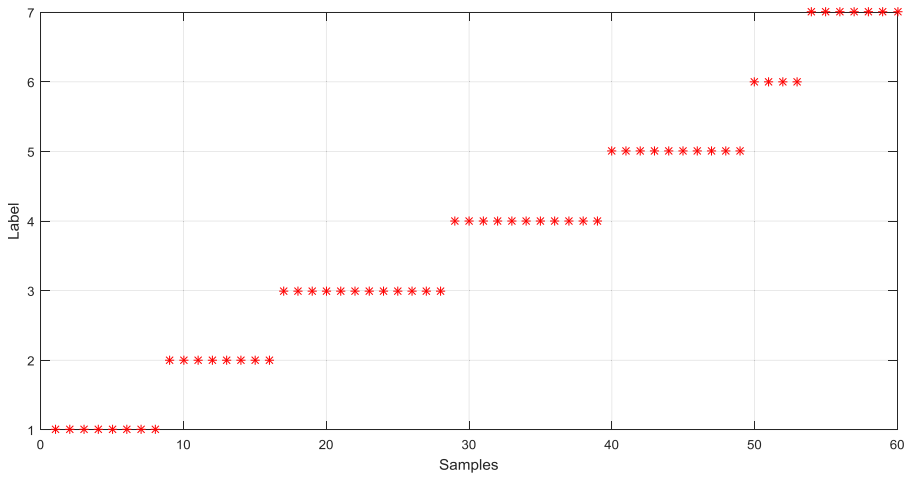

FIGURE 9. Fine classification results for the 12k drive end bearing data.

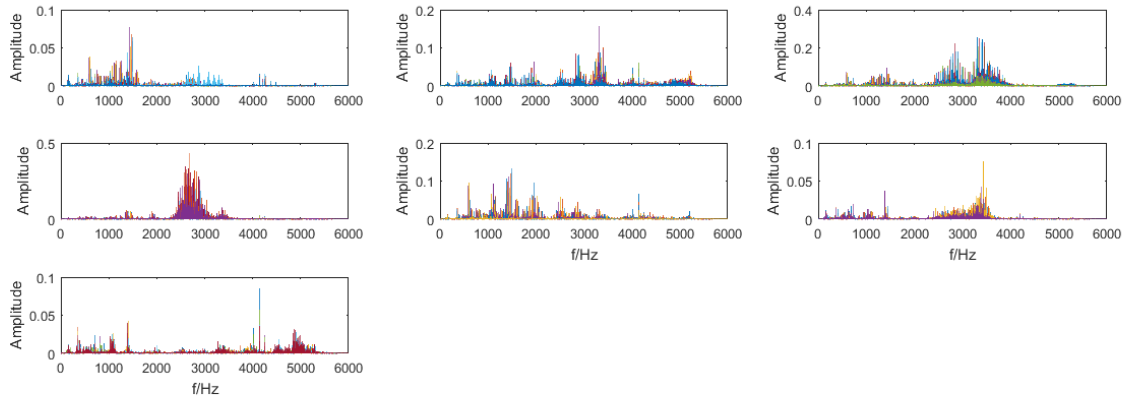

FIGURE 10. Spectra of samples in various classes.

the dataset. Thus, certain special methods are necessary for automatic fault classification. We design an iterative procedure to achieve automatic sample classification, as shown in Fig. 7. The iterative procedure consists of two parts, training the DNN on the dataset and assessing the training result to adjust the training dataset. As shown in Fig. 7, the DNN is trained to learn the dataset knowledge, and a testing process is used to assess the training result and adjust the dataset. In fact, the process involves gradually gathering similar samples into the same class. The process is also a class reduction process.

First, a constructed DNN is trained with the dataset. Note that the neural network can be overfitted if the epoch is too large and that an epoch that is too small cannot train the DNN completely. In other words, the samples are classified completely for epochs that are too large, and some samples with few similarities are considered the same class for epochs that are too small. For a dataset with random signed labels, each sample can be trained as a class if the training epoch is sufficiently large. Otherwise, some samples with similarities are considered the same class. If not all the output of the DNN is consistent with the sample labels, we increase the number

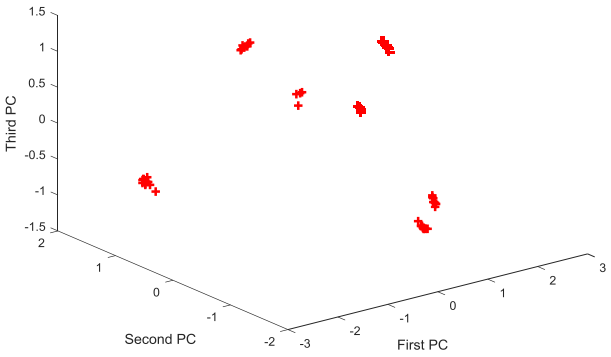

FIGURE 11. First three PCs of the features.

of training epochs unless the number is too large. If the training time is increased to a large number, then the DNN is still not trained successfully. Then, we end the training process and adjust the dataset according to the training results.

For example, consider the 12k drive end bearing data: there are 60 records for the faults associated with the 0.007-inch bearing. We assign each record a label

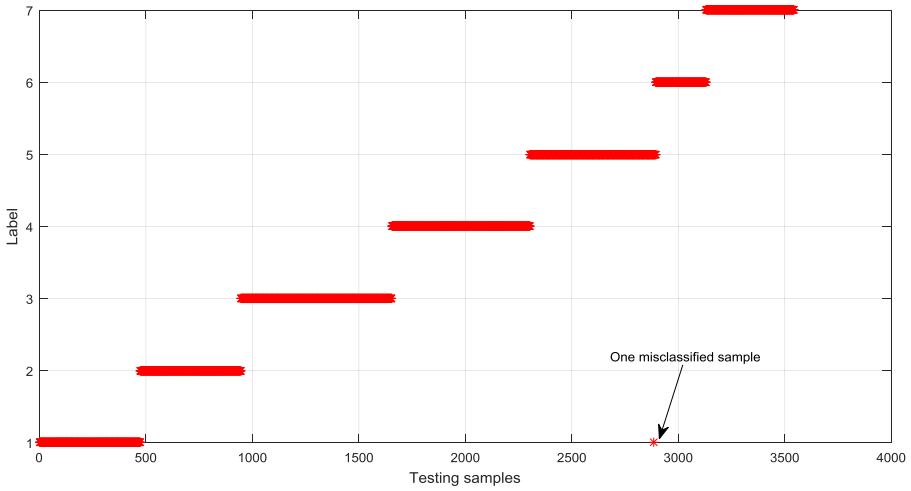

FIGURE 12. Testing results for subsignals of length 2048.

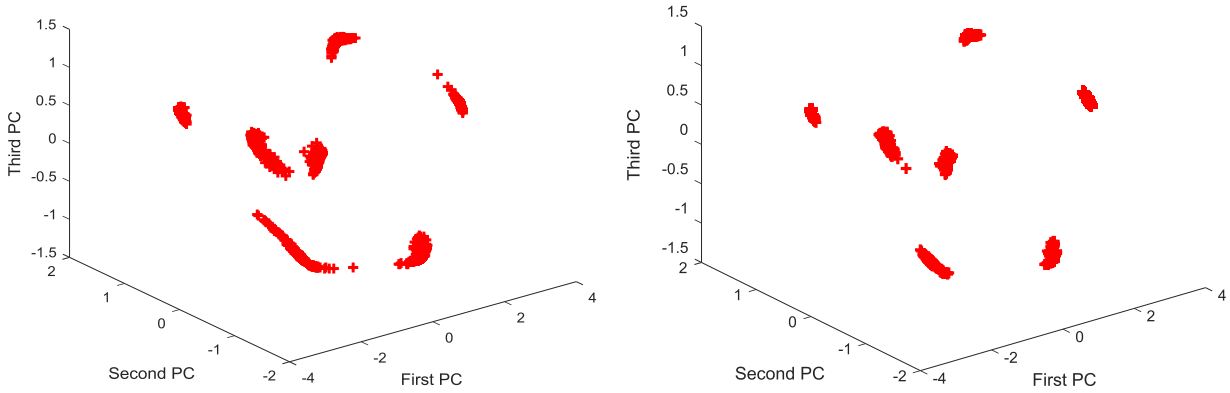

FIGURE 13. First three PCs of the features for subsignals of length 2048. The left and right subfigures show the results of subsignals of length 2048 and 4096, respectively.

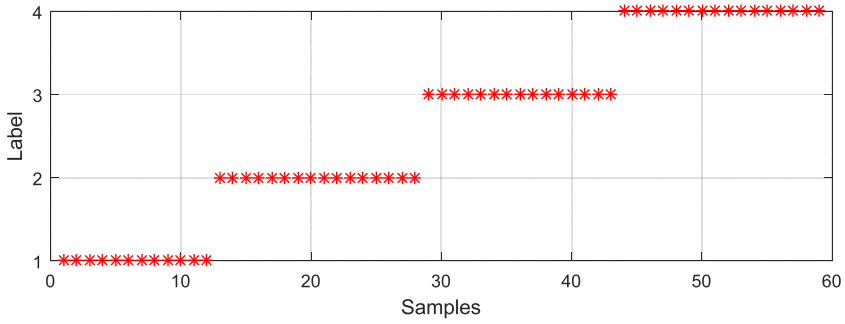

FIGURE 14. Classification results for the 12k fan end bearing data.

from 1 to 60. The DNN structure is shown in the next section. If the training epoch is set as 300, then the DNN output is as shown in Fig. 8. Here, the 60 records are divided into 28 categories, and not all DNN outputs are consistent with the labels. Actually, it is an advantage of the DNN that the input data can be classified independent of the label. The main deficiency is that the training epoch is too small. Thus, we increase the training epoch to 500, and the 60 records are divided into 60 categories.

Although the DNN training is successful, we anticipate that the results may not be as expected because some samples should be classified into the same class. An evaluation of the classification result is necessary, and the sample dataset should be adjusted. Thus, the assessment method and the

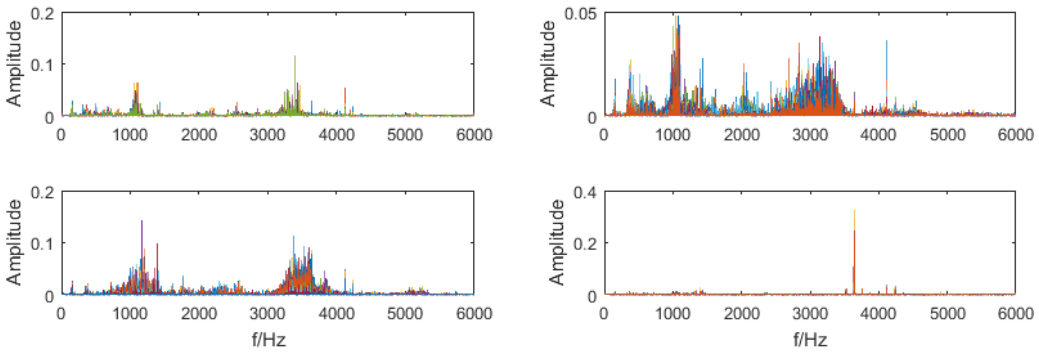

FIGURE 15. Spectra of samples for each class.

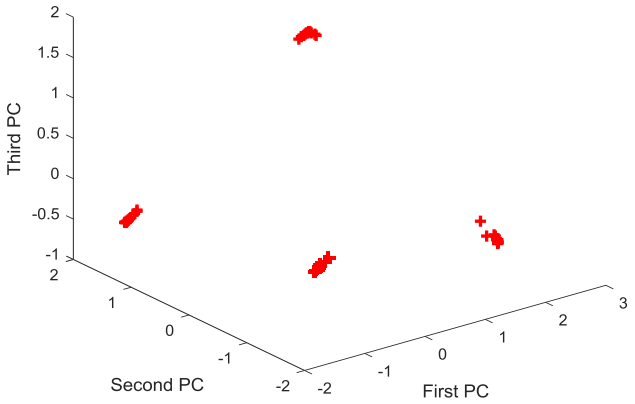

FIGURE 16. First three PCs of the features for the 12k fan end bearing data.

dataset adjustment strategy are the focus of the following work.

We divide the raw data into smaller sections and build a testing dataset. The test sample labels are set according to the training results. For each sub-signal, we calculate the FFT and extract the features. The process of setting a testing dataset is the same as building a training dataset. For a stationary bearing signal, the changes in the signal over time are not particularly intense. Although the FFT of subsignals is similar to that of the raw data, the testing dataset and training dataset are different. Thus, the generalization ability of the DNN is tested with the testing dataset.

Then, we test the trained DNN. The testing output with the sample labels are compared, and misclassified samples are determined if not all testing results meet the requirement. The raw data may have many misclassified samples in the testing result. The misclassified samples S_m in the testing result are counted. Then, a threshold is calculated as follows:

$$V_T = \frac{S_m}{S_T} \tag{4}$$

where S_T is the number of subsamples for the raw data. For raw data, if the misclassified rate is greater than a certain threshold V_T, such as 0.5, the corresponding sample is re-labeled. In other words, we adjust the training dataset according to the testing result.

For the adjusted dataset, the DNN is trained again by using the supervised training method. The above steps are repeated until all the testing results meet the requirements. After an iterative process of training and testing on the DNN, samples with similarities can be classified into the same class. Note that the use of smaller lengths of subsignals is not always better. As shown in Fig. 5, some components are not well reflected if the data length is too small.

It is clear from the above procedure that the dataset is not based on human knowledge but is based only on the data itself. Thus, the entire procedure is an unsupervised learning process. We test the procedure by using the CWRU bearing data in the next section to determine if these data achieve automatic data classification.

IV. TESTING AND ANALYSIS

We test the proposed method by using the 12k drive end, 48k drive end, and 12k fan end bearing data. In each group, we consider only faults in the 0.007-inch bearings. The shaft speed ranges between 1730 and 1797 rpm. The minimum main frequency f_m is 28.833Hz. From Table 1, we can see that the minimum characteristic frequency f_{IR} is 11.484Hz. By using (3), we attain a spectrum that is divided into 522 windows for the 12 kHz sampling and 2090 windows for the 48 kHz sampling. Thus, input layers of 1024 and 4096 are selected for the 12 kHz and 48 kHz sampling data, respectively.

The hidden layer of the DNN is set as 260-130-80-50. The number of output layers is determined by the types of sample labels. Thus, the designed DNN has six layers. The weights of the DNN are initialized randomly. The learning rate is 1, and the batchsize is 1. The encoder activation function is a sigmoid function. The threshold V_T is not always a constant. For each testing dataset, the threshold V_T is first set as 0.5. Then, V_T is changed to 0.3 or 0.1 to evaluate whether the classification results are accurate. The samples are selected randomly as the input to train the DNN.

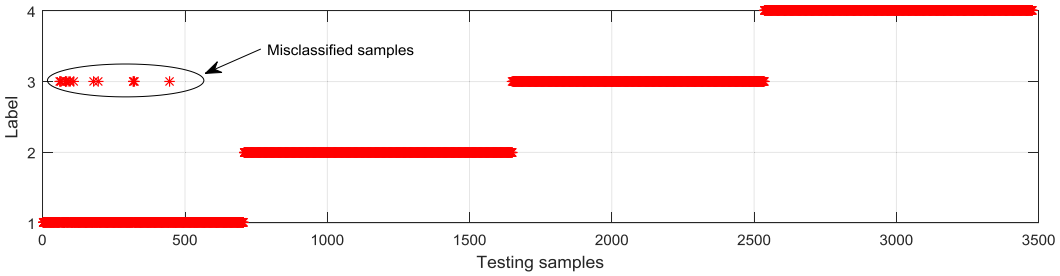

FIGURE 17. Testing results for subsignals of length 2048 extracted from the 12k fan end bearing data.

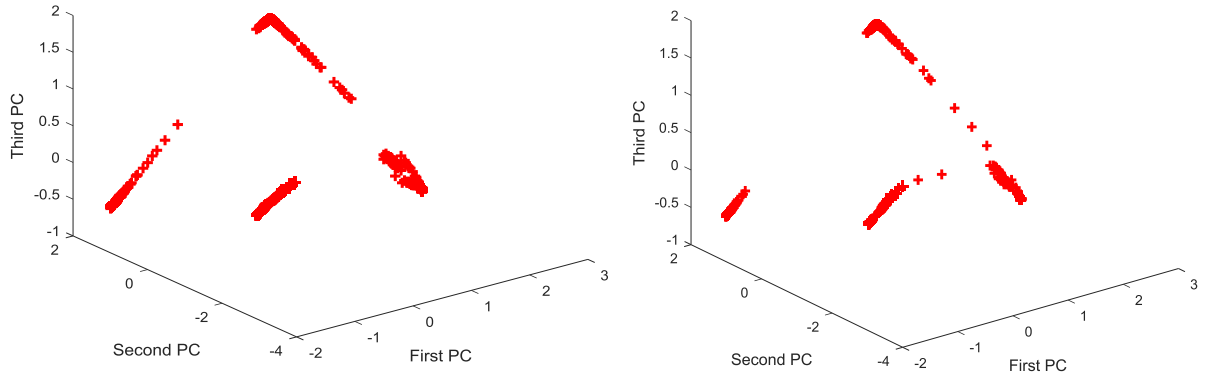

FIGURE 18. First three PCs of the features for subsignals extracted from the 12k fan end bearing data. The left and right subfigures show the results for subsignals of length 2048 and 4096, respectively.

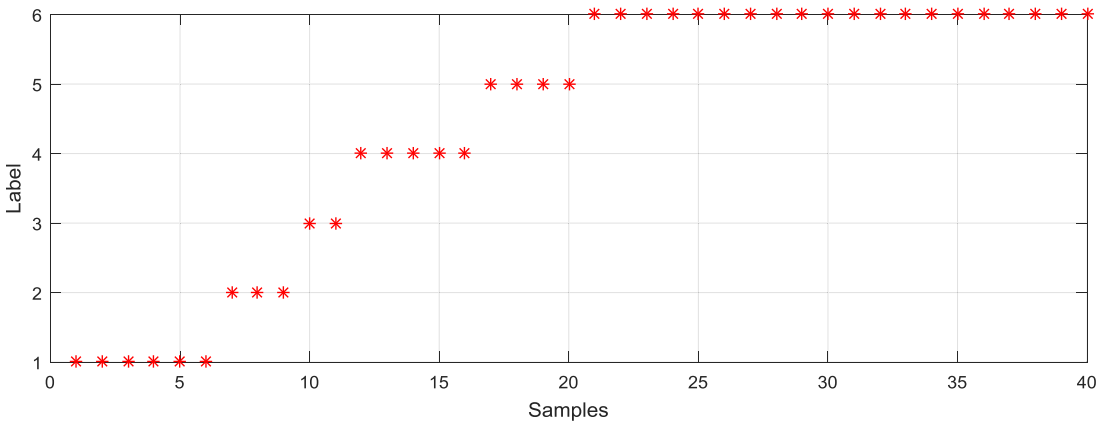

FIGURE 19. Fine classification results for the 48k drive end bearing data.

The training time for each epoch is approximately 0.32 s. The original number of epochs is set to 500. The maximum number of epochs is 20 times the initial number of epochs.

A. ANALYSIS OF THE 12k DRIVE END BEARING DATA

After the initial classification, the 60 records of fault data for the 12k 0.007-inch drive end bearings are divided into 60 categories. Each category has only one sample. Thus, the classification results must be refined. The raw data are divided into small sections with lengths of 32,768, 16,384, 8192, 4096 and 2048 to test the trained DNN. For the subsignal of length 2048, the first 1024 coefficients are used to test the trained DNN because of the symmetry of the FFT result, which means that every spectrum line is used to test the trained DNN. After the fine classification, the results are shown in Fig. 9. The raw data are divided into seven classes. The records in each group are listed in Appendix Table 3. The classification results are not consistent with the given fault types presented previously by [25]. This result is not unexpected. Smith et al. noted previously [3] that some of the datasets have severe noise and cannot be detected in terms of the specified bearing fault by using only the spectrum.

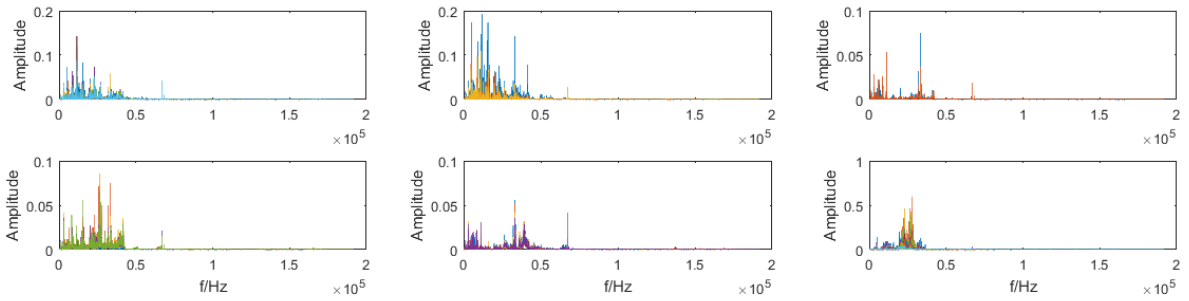

FIGURE 20. Spectra of samples in various classes for the 48k drive end bearing data.

To testify the validity of the classification, the spectra of samples for each class are shown in Fig. 10. The spectra of the records in each class are shown separately in Appendix Figs. 24 to 30. The spectra of samples exhibit prominent differences for various classes.

To further test whether the classification is reasonable, principal component analysis (PCA) is used to visualize the features extracted by the DNN. As in [12] and [30], the output of the last hidden layer is analyzed by PCA because the last layer is actually a classifier. The first three principal components (PCs) of the features are shown in Fig. 11. The features of the raw data are gathered together in seven groups.

In addition, we test the classification result by using subsignals of lengths 32,768, 16,384, 8192, 4096 and 2048. There are 180, 420, 843, 1743, and 3543 subsignals for these five length samples. All the testing results are accurate except for the subsignals of length 2048. There is only 1 sample whose testing result is not consistent with the label for the subsignals of length 2048, as shown in Fig. 12. The first three PCs of the features for the subsignals of length 2048 and 4096 are shown in Fig. 13. We find that the features of these data are clearly gathered into seven groups. Thus, the overall classification result is satisfactory. In fact, these subsignals are not in the training dataset; thus, the proposed method is shown to have strong generalization ability.

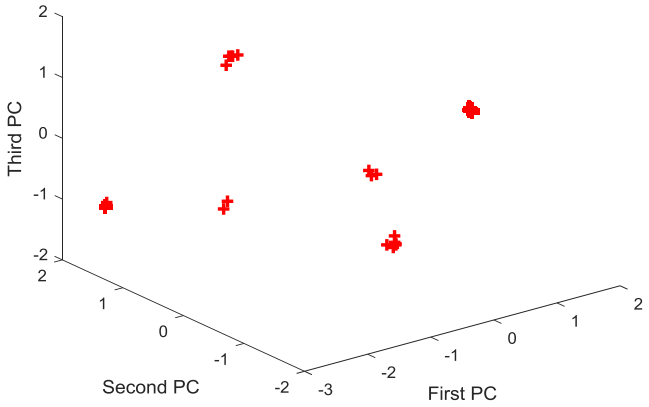

FIGURE 21. First three PCs of the features from the 48k drive end bearing data.

B. ANALYSIS OF THE 12k FAN END BEARING DATA

There are 59 records in the fault data of the 12k fan end 0.007-inch bearings. These records are classified into 4 groups by using the proposed method, as shown in Fig. 14. The spectra of samples in each class are shown in Fig. 15, and the first three PCs of the features are shown in Fig. 16. Fig. 15 displays a clear difference between the spectra of samples in different classes. Fig. 16 shows that the raw data features in different groups are distributed in four corners of a rectangle. These features are clearly separated with a certain distance.

We test the classification result by using subsignals of length 32,768, 16,384, 8192, 4096 and 2048 extracted from the 12k fan end bearing data. There are 177, 413, 826, 1711, and 3478 subsignals for these five types of samples. There are only 2, 4, and 15 misclassified samples for the subsignals of length 8192, 4096, and 2048, respectively. The testing result is shown in Fig. 17 for subsignals of length 2048. The first three PCs of the features are shown in Fig. 18. It is clear that the features are grouped into four classes. Although some classes are misclassified, the testing accuracy rate is 100%, 100%, 99.76%, 99.77%, and 99.57% for these five types of samples. Thus, the proposed method has a strong ability of automatic classification.

C. ANALYSIS OF THE 48k DRIVE END BEARING DATA

There are 40 records for the fault data of the 48k drive end 0.007-inch bearings. Each record is given a unique label that is a random number between 1 and 40. After approximate classification, these records are divided into 40 categories. The raw data are split into small pieces with lengths of 32,768, 16,384 and 8192 to test the trained DNN. After an iteration of training and testing the DNN, the samples are divided into six classes, as shown in Fig. 19. The spectrum of each class is shown in Fig. 20. The first three PCs of the features are shown in Fig. 21. Fig. 20 shows that the spectra of samples in different classes display distinct differences. There is no overlap for the raw data features in different groups.

We test the classification result by using subsignals of lengths 32,768, 16,384 and 8192, which were extracted from

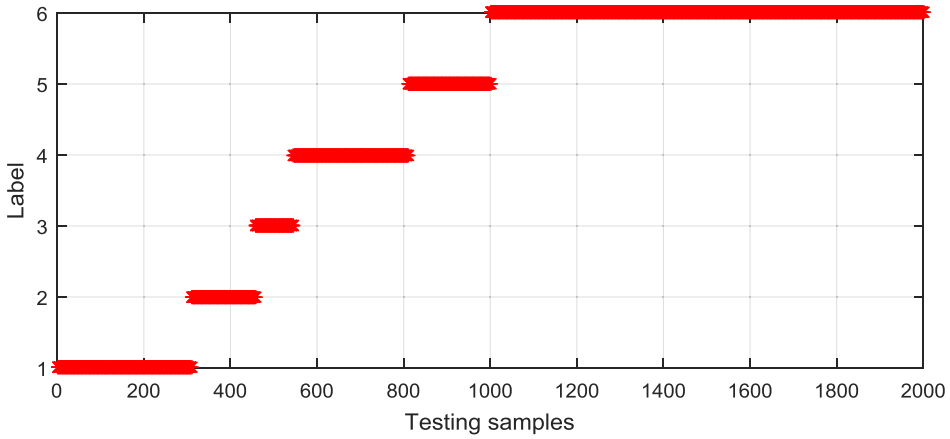

FIGURE 22. Testing result of subsignals of length 8192 extracted from the 48k drive end bearing data.

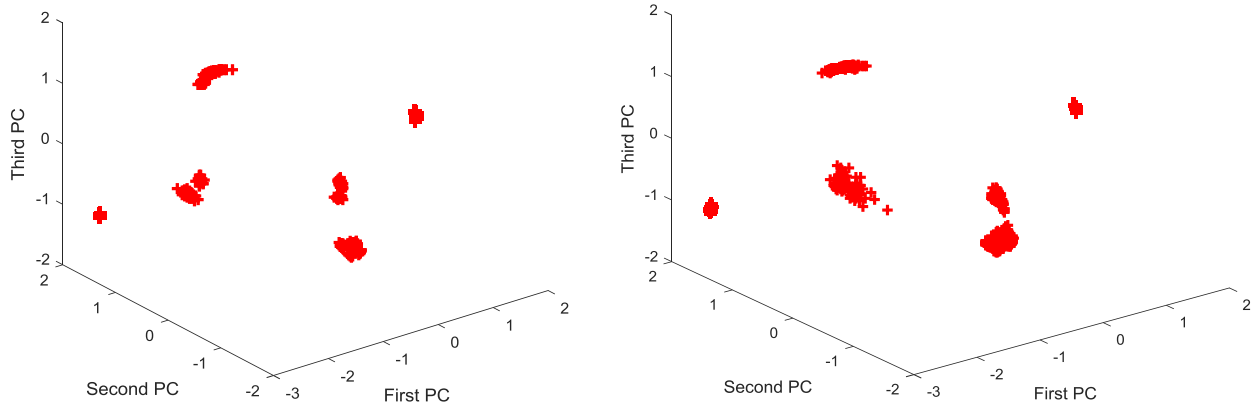

FIGURE 23. First three PCs of the features for subsignals extracted from the 48k drive end bearing data. The left and right subfigures show the results for subsignals of length 16,384 and 8192, respectively.

TABLE 2. Length of the 12k drive end bearing data.

Fault diameter	IR	Length	Ball	Length	OR (centered)	Length	OR (orthogonal)	Length	OR (opposite)	Length
0.007 in.	105	121,265	118	122,571	130	121,991	144	122,281	156	122,281
	106	121,991	119	121,410	131	122,426	145	121,846	158	121,991
	107	122,136	120	121,556	132	121,410	146	121,556	159	122,281
	108	122,917	121	121,556	133	122,571	147	122,281	160	122,136
0.014 in.	169	121,846	185	121,846	197	121,846				
	170	121,846	186	122,136	198	122,136				
	171	121,846	187	121,991	199	121,846				
	172	121,701	188	122,136	200	121,991				
0.021 in.	209	122,136	222	121,991	234	122,426	246	121,701	258	121,846
	210	121,556	223	121,701	235	121,991	247	121,991	259	122,426
	211	121,846	224	122,136	236	122,281	248	122,281	260	122,716
	212	121,991	225	122,136	237	121,991	249	122,136	261	121,701
0.028 in.	3001	120,801	3005	120,801						
	3002	121,351	3006	121,351						
	3003	121,351	3007	120,984						
	3004	121,535	3008	120,984						

the 48k drive end bearing data. There are 474, 982 and 2000 subsignals for these three types of samples. All testing output is consistent with the associated label. The accuracy rate is 100%. The testing result is shown in Fig. 22, and the first three PCs of the features are shown in Fig. 23 for subsignals of length 8192. For comparison,

TABLE 3. Classification results for the 12k drive end bearing data.

Category	IR	Ball	OR	
1	105BA	121BA	145BA	159BA
	106BA			160BA
	107BA			
	108BA			
2			130FE	144FE
			131FE	145FE
			132FE	146FE
			133FE	147FE
3	105DE		130DE	156DE
	106DE		131DE	158DE
	107DE		132DE	159DE
	108DE		133DE	160DE
4			130BA	144BA
			131BA	144DE
			132BA	145DE
			133BA	146BA
				146DE
				147BA
				147DE
5	105FE	118BA		156BA
	106FE	119BA		158BA
	107FE	120BA		
	108FE	121FE		
6		118DE		
		119DE		
		120DE		
		121DE		
7		118FE		156FE
		119FE		158FE
		120FE		159FE
				160FE

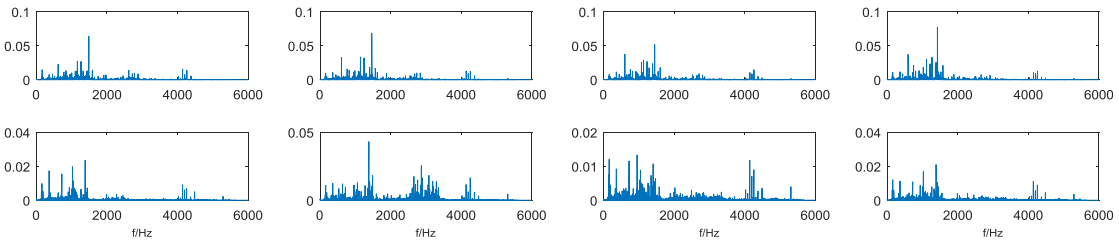

FIGURE 24. Spectra of samples in class 1 for the 12k drive end bearing data.

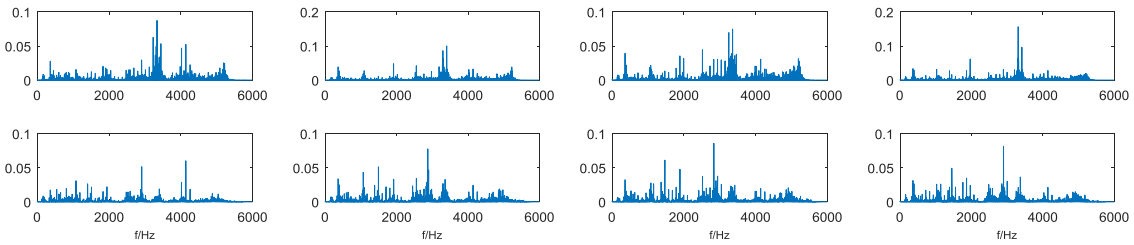

FIGURE 25. Spectra of samples in class 2 for the 12k drive end bearing data.

the first three PCs of the features for subsignals of length 16,384 are also presented in Fig. 23. Comparing the two subfigures, we observe that there are obvious boundaries between classes. Thus, the proposed method offers satisfactory performance and strong generalization ability.

V. CONCLUSIONS

An automatic classification method based on deep learning for bearing fault diagnosis is proposed. Two measures are used to ensure that the proposed method can classify samples automatically. One measure is that the sample label is signed randomly and modified in the following course. The other

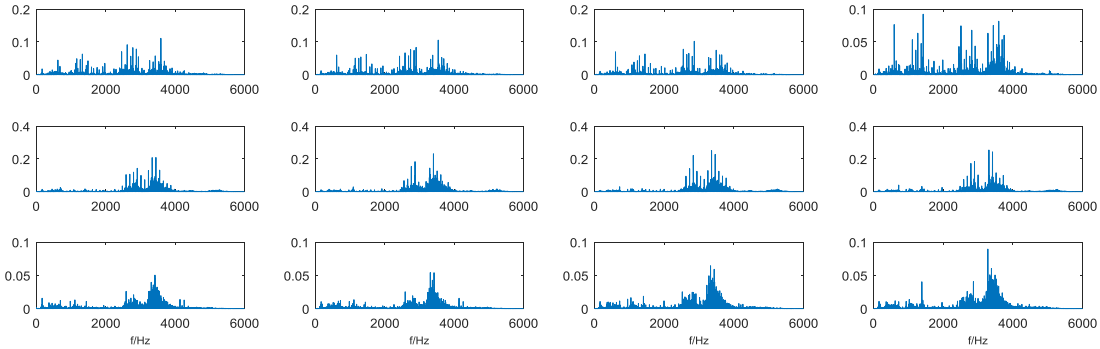

FIGURE 26. Spectra of samples in class 3 for the 12k drive end bearing data.

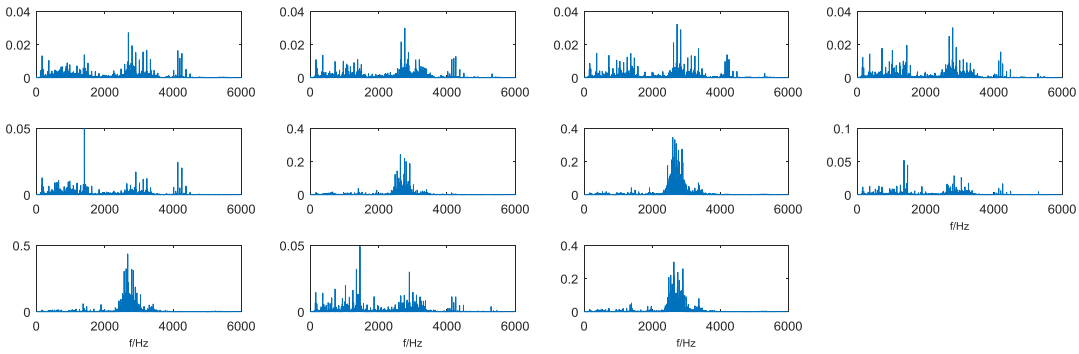

FIGURE 27. Spectra of samples in class 4 for the 12k drive end bearing data.

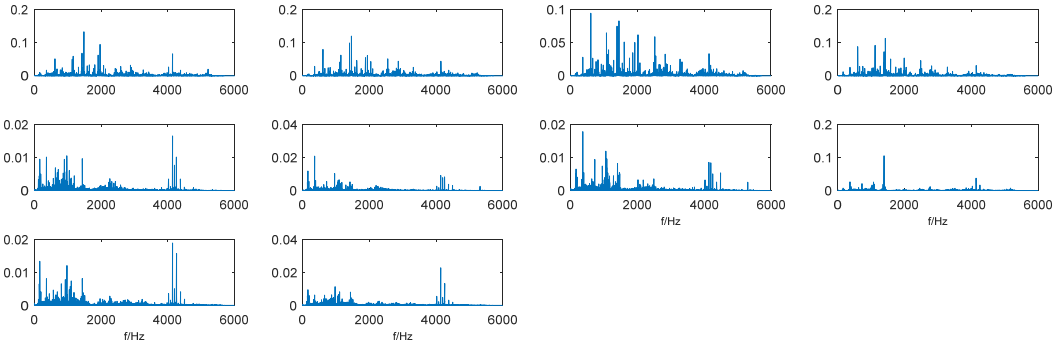

FIGURE 28. Spectra of samples in class 5 for the 12k drive end bearing data.

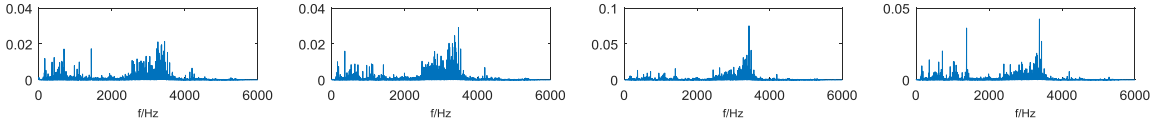

FIGURE 29. Spectra of samples in class 6 for the 12k drive end bearing data.

measure is to adjust the sample labels through testing on the trained DNN by using subsignals. In other words, the core of the proposed algorithm is that the DNN is trained on small-dimension data and that the sample labels are adjusted according to the testing result on the trained DNN. By using small-dimension data, only a small structure of the neural network is required, which can reduce the computation time. The proposed method is particularly suitable for fault signal classification processing of large data. The entire process has no human intervention; thus, the process does not depend on prior expert knowledge.

The proposed method is tested with CWRU bearing data, which feature a long data length, the existence of data with non-classical features of fault identification, and some components exhibiting changing amplitude. The results show that the 12k drive end bearing data, 48k drive end bearing

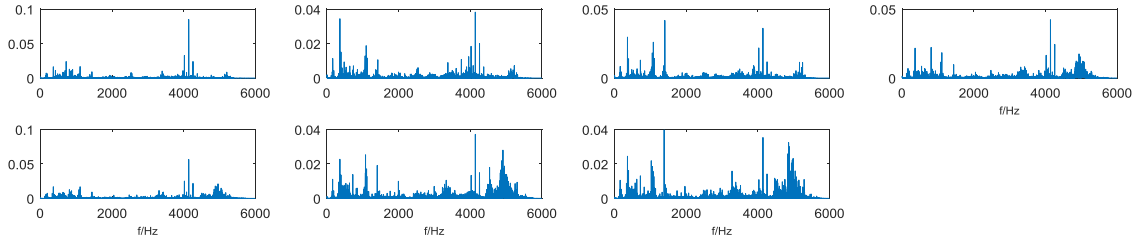

FIGURE 30. Spectra of samples in class 7 for the 12k drive end bearing data.

data, and 12k fan end bearing can be classified into 7, 6 and 4 categories, respectively. We show that the proposed method has satisfactory performance and strong generalization ability and can be used to classify the running state of the bearing. In fact, the proposed method is suitable not only for bearing fault diagnosis but also for the fault classification of rotating machinery.

In engineering practice, equipment is usually operated under complicated and variable conditions, and the measured signals are non-stationary. The proposed method also requires a large amount of data, especially real-world signals, for further testing. There is no doubt that this method needs to be improved for practical applications. Applying the proposed method in engineering practice and improving it to achieve intelligent fault diagnosis are the direction of our future research.

APPENDIX

See Tables 2 and 3 and Figs. 24–30.

REFERENCES

[1] A. Heng, S. Zhang, A. C. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," *Mech. Syst. Signal Process.*, vol. 23, no. 3, pp. 724–739, 2009.

[2] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, Nov. 2016.

[3] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mech. Syst. Signal Process.*, vols. 64–65, pp. 100–131, Dec. 2015.

[4] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Process.*, vol. 130, pp. 377–388, Jan. 2017.

[5] M. He and D. He, "Deep learning based approach for bearing fault diagnosis," *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 3057–3065, May/Jun. 2017.

[6] Z. Chen, S. Deng, X. Chen, C. Li, R.-V. Sanchez, and H. Qin, "Deep neural networks-based rolling bearing fault diagnosis," *Microelectron. Rel.*, vol. 75, pp. 327–333, Aug. 2017.

[7] M. Farajzadeh-Zanjani, R. Razavi-Far, M. Saif, J. Zarei, and V. Palade, "Diagnosis of bearing defects in induction motors by fuzzy-neighborhood density-based clustering," in *Proc. 14th IEEE Int. Conf. Mach. Learn. Appl.*, Miami, FL, USA, Dec. 2015, pp. 935–940.

[8] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[9] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring: A survey," *J. Latex Class Files*, vol. 14, no. 8, pp. 1–14, 2015.

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

[11] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1693–1702, Jul. 2017.

[12] 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," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 303–315, May 2016.

[13] H. Shao, H. Jiang, H. Zhao, and F. Wang, "A novel deep autoencoder feature learning method for rotating machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 95, pp. 187–204, Oct. 2017.

[14] X. Guo, C. Shen, and L. Chen, "Deep fault recognizer: An integrated model to denoise and extract features for fault diagnosis in rotating machinery," *Appl. Sci.*, vol. 7, no. 1, p. 41, 2017.

[15] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu, "Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery," *IEEE Access*, vol. 5, pp. 15066–15079, Jul. 2017.

[16] F. Zhou, Y. Gao, and C. Wen, "A novel multimode fault classification method based on deep learning," *J. Control Sci. Eng.*, vol. 2017, Mar. 2017, Art. no. 3583610.

[17] C. Lu, Z. Wang, and B. Zhou, "Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification," *Adv. Eng. Inform.*, vol. 32, pp. 139–151, Apr. 2017.

[18] J. Pan, Y. Zi, J. Chen, Z. Zhou, and B. Wang, "LiftingNet: A novel deep learning network with layerwise feature learning from noisy mechanical data for fault classification," *IEEE Trans. Ind. Electron.*, vol. 65, no. 6, pp. 4973–4982, Jun. 2018.

[19] H. Chen, J. Wang, B. Tang, K. Xiao, and J. Li, "An integrated approach to planetary gearbox fault diagnosis using deep belief networks," *Meas. Sci. Technol.*, vol. 28, no. 2, pp. 1–17, 2017.

[20] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, "Deep model based domain adaptation for fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 64, no. 3, pp. 2296–2305, Mar. 2017.

[21] M. Xia, T. Li, L. Liu, L. Xu, and C. W. de Silva, "Intelligent fault diagnosis approach with unsupervised feature learning by stacked denoising autoencoder," *IET Sci., Meas. Technol.*, vol. 11, no. 6, pp. 687–695, Sep. 2017.

[22] R. Razavi-Far, E. Hallaji, M. Saif, and G. Ditzler, "A novelty detector and extreme verification latency model for nonstationary environments," *IEEE Trans. Ind. Electron.*, vol. 66, no. 1, pp. 561–570, Jan. 2019.

[23] R. Razavi-Far, E. Hallaji, M. Farajzadeh-Zanjani, and M. Saif, "A semi-supervised diagnostic framework based on the surface estimation of faulty distributions," *IEEE Trans. Ind. Informat.*, to be published.

[24] D. Silver *et al.*, "Mastering the game of go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[25] *Case Western Reserve University Bearing Data Center Website*. Accessed: Oct. 2016. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/home

[26] *DeepLearnToolbox Website*. Accessed: Mar. 2016. [Online]. Available: https://github.com/rasmusbergpalm/DeepLearnToolbox

[27] W. Du, J. Tao, Y. Li, and C. Liu, "Wavelet leaders multifractal features based fault diagnosis of rotating mechanism," *Mech. Syst. Signal Process.*, vol. 43, pp. 57–75, Feb. 2014.

[28] X.-H. He, D. Wang, Y.-F. Li, and C.-H. Zhou, "A novel bearing fault diagnosis method based on Gaussian restricted Boltzmann machine," *Math. Problems Eng.*, vol. 2016, Dec. 2016, Art. no. 2957083.

[29] L. Guo, N. P. Li, F. Jia, Y. G. Lei, and J. Lin, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, pp. 98–109, May 2017.

[30] H. Shao, H. Jiang, F. Wang, and H. Zhao, "An enhancement deep feature fusion method for rotating machinery fault diagnosis," *Knowl.-Based Syst.*, vol. 119, pp. 200–220, Mar. 2017.

YANLI YANG received the Ph.D. degree from the Beijing Institute of Technology in 2010. He is currently an Associate Professor with the School of Electronics and Information Engineering, Tianjin Polytechnic University. His research interests include signal processing and artificial neural network.

PEIYING FU is currently pursuing the master's degree with the School of Electronics and Information Engineering, Tianjin Polytechnic University. Her research interest includes deep neural networks.

YICHUAN HE is currently pursuing the master's degree with the School of Electronics and Information Engineering, Tianjin Polytechnic University. His research interest includes equipment condition monitoring.