Energy-Fluctuated Multiscale Feature Learning With Deep ConvNet for Intelligent Spindle Bearing Fault Diagnosis

Xiaoxi Ding and Qingbo He, *Member, IEEE*

***Abstract*— Considering various health conditions under varying operational conditions, the mining sensitive feature from the measured signals is still a great challenge for intelligent fault diagnosis of spindle bearings. This paper proposed a novel energy-fluctuated multiscale feature mining approach based on wavelet packet energy (WPE) image and deep convolutional network (ConvNet) for spindle bearing fault diagnosis. Different from the vector characteristics applied in intelligent diagnosis of spindle bearings, wavelet packet transform is first combined with phase space reconstruction to rebuild a 2-D WPE image of the frequency subspaces. This special image can reconstruct the local relationship of the WP nodes and hold the energy fluctuation of the measured signal. Then, the identifiable characteristics can be further learned by a special architecture of the deep ConvNet. Other than the traditional neural network architecture, to maintain the global and local information simultaneously, deep ConvNet combines the skipping layer with the last convolutional layer as the input of the multiscale layer. The comparisons of clustering distribution and classification accuracy with six other features show that the proposed feature mining approach is quite suitable for spindle bearing fault diagnosis with multiclass classification regardless of the load fluctuation.**

***Index Terms*— Deep convolutional network (ConvNet), multiscale feature, spindle bearing fault diagnosis, wavelet packet energy (WPE) image.**

I. Introduction

SPINDLE bearings have been widely used as vital parts in intelligent manufacturing. Mechanical faults on the spindle bearings may cause significant economic losses (e.g., quality problems in manufacturing) and increase the downtime [1], [2]. However, due to complex and varying conditions in the real spindle bearing applications, systematically identifying the sensitive features without human intervention is still a great challenge for effective and complete spindle bearing fault diagnosis. During the past decades, many feature extraction methods have been available for bearing fault diagnosis and bearing health monitoring in the literature [1]–[6]. Malhi and Gao [4] proposed a Principal Component Analysis (PCA)-based feature selection scheme, and Van and Kang [5] employed wavelet kernel local fisher discriminant analysis to classify the machine defects. Soualhi *et al.* [6] used Hilbert–Huang transform, support vector machine (SVM), and regression to monitor the bearing health conditions. In these conventional machine learning techniques, sensitive features were generated and extracted for bearing fault diagnosis. Meanwhile, the artificial intelligence technique has been a useful tool for intelligent health diagnosis of machinery, such as neural networks (NNs) [7], SVM [8], [9], decision tree [10], etc. More recently, deep learning (DL), as a new research hotspot of machine learning, has been achieved state-of-art in many advanced fields, including face verification and speech recognition, and many powerful features have been proposed, such as deep face [11], deepID [12], and DeepSpeech [13]. In the learning process, a much deeper architecture can be established by DL technique, such as deep belief networks, deep NNs (DNNs), and convolutional NN (CNN). With these multiple nonlinear layers for nonlinear abstractions, DL shows its strong potential in establishing nonlinear mapping relationship between health conditions and pattern recognition.

Motivated by the strong abilities of DL in the application of feature expression and extraction, in recent years, some researchers have been trying to do some related studies about DL for bearing fault diagnosis. Jia *et al.* [14] employed DNN to extract fault characteristics for rotating machinery. Sun *et al.* [15] developed sparse autoencoder based on DNN for induction motor faults classification. Gan *et al.* [16] proposed a hierarchical diagnosis network for recognition. However, in this paper, three aspects need to be further analyzed and improved. First, the raw signal or the amplitudes of whole/integrant frequency band obtained by fast Fourier transform (FFT) are regarded as the input vector features of the NN. It can be seen that there would be serious nonstationarity in the raw signal once a possible fault happened under the varying operational conditions. On this account, the traditional features based on FFT will lose the varying information in time domain, while those based on time-domain methods will lose the varying characteristics in the frequency domain. Furthermore, the signal or features are simply used as the 1-D vectors for fault diagnosis. However, it can be foreseen that the information may be much easier to be understood and mined in a high dimension. Second, due to the noise interference and window function influence for the time–frequency distribution,

Manuscript received September 3, 2016; revised November 10, 2016; accepted January 9, 2017. Date of publication March 17, 2017; date of current version July 12, 2017. This work was supported in part by the Program for New Century Excellent Talents in University under Grant NCET-13-0539, in part by the National Natural Science Foundation of China under Grant 51475441, and in part by the Youth Innovation Promotion Association of the Chinese Academy of Sciences under Grant 2016396. The Associate Editor coordinating the review process was Dr. Ruqiang Yan. *(Corresponding author: Qingbo He.)*

The authors are with the Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China, Hefei 230026, China (e-mail: qbhe@ustc.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIM.2017.2674738

it is not easy to construct effective and robust features for classification according to time–frequency domain methods. Third, in this paper, only one layer of the DL architecture is used for the next layer construction in the network propagation, and the output only contains the information of the last layers in the feature learning process. The mined features will become robust along with the network depth. Although this will bring benefits for bearing fault diagnosis under various operating conditions in a certain degree, these features will lose the detailed information to distinguish the various bearing health states. Thus, this means that the features, directly extracted from one layer, are insufficient.

As displayed in deep face and DeepSpeech, it is easy known that CNN can directly extract the robust and identifiable characteristics from a higher dimensional structure, such as 2-D/3-D images. CNN architectures [17]–[19] employ the kernels in the convolution process (that shares the same weights) to sense the local structure of the characteristics from the 2-D/3-D inputs and use subsampling for lower resolutions in a robust way (alternately convolute and pool on the feature maps). In this view of high-dimensional input/structures, CNN can be further imported into bearing fault diagnosis to improve the feature mining in the mentioned aspects. Specifically, motivated by the representation of the raw signal in a 2-D space (image) based on time–frequency analysis, this paper provides a frequency distribution image for mining analysis by the convolutional network (ConvNet) mentioned above. In this way, the frequency energy fluctuation with the local structure of the frequency components can be built for a further feature learning. As one of the famous wavelet transform (WT) methods, wavelet packet transform (WPT) [21]–[24] can separate the signal into a series of frequency subspaces without losing the energy of the frequency components. The distribution of the wavelet packet energy (WPE) can reflect the frequency characteristics but ignores the integral energy and the physical relations among the WP nodes.

Nevertheless, in the architecture of CNN, only the feature maps in the last convolutional layer (which contains few neurons) are employed as the input of the last layer, which is mentioned in the third aspect. And, these feature maps will become more invariant and robust with the loss of the precise details from the previous ones, which is not beneficial to feature mining. Inspired by the specific multilayers of CNN, Sermanet and LeCun [20] proposed multiscale ConvNets for traffic sign recognition by the use of connections that skip layers. Sun *et al.* [12] introduced deepID for face representation by adding the bypassing connections between the last convolutional layer and the previous pooling layer in the multilayer CNN. Their studies showed that the multiscale features/deep ID were much more invariant and robust (global information) with precise details (local information) than the deep features based on the basic architecture of CNN. The global and local information can be maintained and mined well to increase the network capacity. Therefore, it can be foreseeable that much more identifiable characteristics can be established between bearing health conditions and the patterns by the use of connections between the skipping layers. In these manners, this multiscale ConvNets can effectively improve the shortcoming as mentioned in the third aspect.

Inspired by the 2-D/3-D image structure and the multiscale ConvNet corresponding to the aforementioned three aspects, an energy-fluctuated multiscale feature (EFMF) learning method is proposed by combining the WPE image (WPI) and deep ConvNet for intelligent spindle bearing fault diagnosis. In this proposed method, based on dynamic structure construction via the phase space reconstruction (PSR) technique [25], a 2-D WPI is first constructed to represent the dynamic structure of the WPE distribution for different patterns under varying operating conditions. In this manner, the integral energy and the physical relations can be combined into this specific energy image with different brightness distributions. Then, a multilayer deep ConvNet is built. These WPIs are used as the input to train the deep ConvNet architecture. Finally, the multiscale features in the multiscale layer can be extracted to reveal the identifiable characteristics of measured signals. The main contribution of this paper is development of the deep ConvNet-based multiscale feature for intelligent spindle bearing fault diagnosis under varying operational conditions, e.g., load fluctuation.

The remainder of this paper is organized as follows. Section II gives the theoretical background. Section III provides the basic theoretical framework of the proposed EFMF learning method. Section IV investigates the effectiveness of the proposed EFMF for spindle bearing fault diagnosis. Finally, Section V provides the conclusion.

II. Theoretical Background

A. WPT

Various valid features have been extracted for bearing fault diagnosis. WPT is a powerful WT technique by means of decomposing a nonstationary signal into different subspaces with the multiresolution merit, which can overcome the weak resolution in the high-frequency regions of the discrete WT. In brief, with the low-pass filter $h(k)$ and the high-pass filter $g(k)$, the raw signal $x(t)$ is decomposed iteratively as [23], [24]

$$\begin{aligned} d_{i+1}^{2n}(\tau) &= \sum_k h(k-2\tau) d_i^n(k) \\ d_{i+1}^{2n+1}(\tau) &= \sum_k g(k-2\tau) d_i^n(k) \end{aligned} \tag{1}$$

where the signal $x(t)$ with the sample length N is $\{d_0^1(k), k = 1, 2, \ldots, N\}$ and $d_i^n(k)$ is the WP decomposition coefficients of the nth node at the ith level. Moreover, $d_{i+1}^{2n}(\tau)$ and $d_{i+1}^{2n+1}(\tau)$ denote the wavelet coefficients of the $2n$th and $(2n+1)$th nodes at the $(i+1)$th level, which, respectively, correspond to the approximation coefficients and detail coefficients. That means WPT can divide the full frequency range of the raw signal into the comprehensive frequency band for an ideal signal analysis, which is realized through filtering with a set of low-pass and high-pass wavelet filters recursively. At the ith level, 2^i nodes (frequency bands) can be obtained and the node can be numbered as (i, p) $(p = 0, 1, 2, \ldots, 2^i - 1)$ in a binary decomposition tree, as shown in Fig. 1

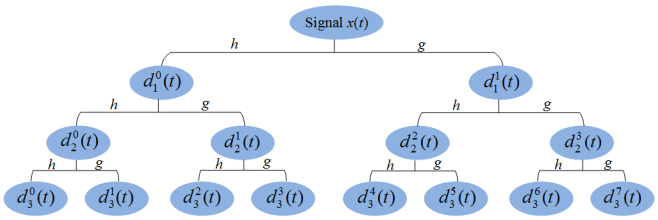

Fig. 1. Illustration of the WPT.

(where a three-level WPT with eight different subspaces are demonstrated).

Contrary to the recursive splitting operations of (1), a reconstruction process can be represented based on the WP coefficients as

$$d_i^n(\tau) = \sum_k H(k-2\tau)\tilde{d}_{i+1}^{2n}(k) + \sum_k G(k-2\tau)\tilde{d}_{i+1}^{2n+1}(k) \tag{2}$$

where $\tilde{d}$ expresses inserting a zero beside each point of d (corresponding to the down-sampling operation in the decomposition process). In this paper, except for the node (i, p), coefficients of all WP nodes at ith level are set to zero. Therefore, a series of signals (2^i) with the same length of the raw signal $x(t)$ can be reconstructed, which retains the frequency information of the reconstructed node.

B. CNN

The architecture of CNN can be generally regarded as a multiple combination of a convolutional layer (feature extraction layer), a nonlinear transform layer (feature mapping layer), and a pooling layer. In the convolutional layer, the input of each neuron connects to the local receptive field of the preceding layer and extracts the local feature by the kernels/filter band, which can represent features of the frequency maps in a limited frequency range. Through a series of convolutions between the kernels and the feature maps of lower layer, the desired features can be regrouped and extracted based on the similar statistical property among the feature image. In each nonlinear transform layer, the multiple lower inputs further compose the feature maps of higher layer through a series of activation functions. The convolution operation with nonlinear transformation can be expressed as

$$h_n^{r,k} = \psi\left(\sum_m v_m^{r-1} * K_n^r + b_n^r\right) \tag{3}$$

where v_m^{r-1} and $h_n^{r,k}$ are, respectively, the *mth* input of level $r-1$ and *nth* output feature maps of level r in the convolution process. And K_n^r is the convolution kernel between the *mth* input feature map and the *nth* output feature map and k is the number of the kernels. b_n is the bias of the *nth* output feature maps. It can be seen that the different feature maps as input share the same kernels and bias. And ψ $(\bullet)$ is a nonlinear activation function, which is set to be "*sigmoid*" in the whole net structure, including the convolutional layers and the soft-max layer.

Specifically, a spatial pooling (subsampling) layer follows on the heels of the top of each convolutional layer to reduce the resolution representation of feature structure, which can make the feature maps much more robustly to small variance. Here, the max-pooling is applied as

$$h_m(a,b) = \max_{0\le p,q<s}\{v_m(a\cdot s+p, b\cdot s+q)\} \tag{4}$$

where each neuron in the *mth* output feature map h_m subsamples from a local region with the size of $s\times s$ nonoverlapped in the *mth* input feature map v_m. Finally, the output of the last convolutional layer is unfolded and used as the input of the soft layer to predict the health condition. In this process, the loss function (mean square error) is defined as

$$J(\theta; f, y) = \frac{1}{2}||h_\theta(f) - y||^2 \tag{5}$$

where f is the expanding features in the last layer (convolutional or pooling layer) and y is the desired output. h_θ is the last regression function for results predicting. $\theta = \{W, b\}$ are the parameters of the regression function. Thus, the structure/parameters can be fine-tuned by the gradient-based supervised training process to minimize the loss function based on backpropagation (BP) algorithm.

III. Multiscale Feature Learning With Deep ConvNet

Considering various health conditions under varying operational conditions, this paper proposed an EFMF learning method based on WPI and deep ConvNet for intelligent spindle bearing fault diagnosis. The flowchart of the proposed method is drawn in Fig. 2. The first step is to acquire the data set. After the vibration signal is measured from the spindle bearings, WPI is constructed based on the PSR. This WP phase space image can reflect the integral energy and the physical relations of the WP nodes in a 2-D space, which is quite different from the conventional vector feature. Then these acquired WPIs are used as the input of the deep ConvNet, where the BP algorithm is employed to fine-tune the architecture by minimizing the log-likelihood function. In this process, the frequency energy fluctuation and local relationship of the frequency energy pixels can be deeply sensed by the convolution kernel function. Once the network is trained, it can be easily known that the output of the deep ConvNet for each sample (the proposed multiscale features are put into the soft-max layer for a probability distribution of the bearing fault) can be directly employed for bearing fault diagnosis. Actually, there are two kinds of application here: one is the probability distribution for class identification and the other is the multiscale feature for feature mining. In this paper, the multiscale features, which possess invariant but precise details, are extracted to effectively illustrate the feature discrimination and clustering. Here, the nearest mean classifier is simply applied to measure the classification accuracy, whose principle is based on the closest Euclidean distance. More details about the proposed system framework are presented in the following sections.

A. WPI

It is foreseeable that there is a wealth of class information among the reconstructed coefficients from the WP nodes and

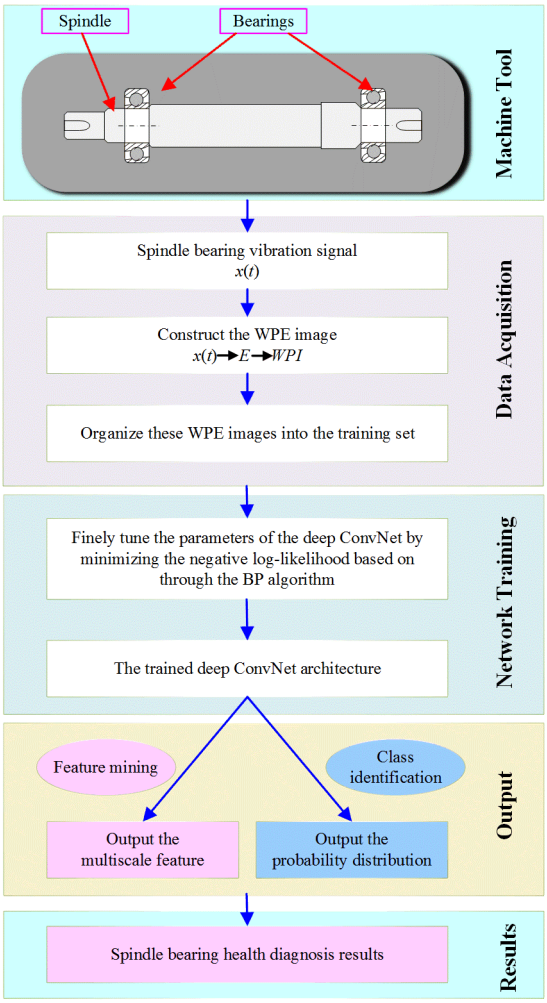

Fig. 2. Flowchart of the proposed EFMF mining method.

the coefficients of some nodes will have a significant change once a fault happens. As an intuitive characteristic, the energy is first calculated from the WP reconstructed coefficients. This paper reengineers the energy characteristics distributed on the WP nodes and explores a new energy image among different WP nodes.

According to (1) and (2) in Section II, the reconstructed coefficients of WP node $(i,\ p)$ are expressed as $\{Rc_i^p(n),\ p = 0, 1, \ldots, (2^i - 1)\}$, and the corresponding WPE characteristics can be calculated as

$$E_i^p = \sum_{n=1}^{N} \left[Rc_i^p(n)\right]^2. \tag{6}$$

The number of this energy feature vector is totally 2^i. Based on the PSR technique [25], the energy feature vector is changed into a new WP phase space. In this manner, the dynamic structure of the frequency distribution can be rebuilt in WPI. The jth phase vector in an m-dimensional phase space is set as $\text{WPI}_i|_j^m = [E_i^p(j), E_i^p(j+\tau), \ldots, E_i^p(j+(m-1)\tau)\ ]$. In this paper, the time delay τ is set to be equal to the embedding dimension m. Therefore, this WPI can be reengineered and

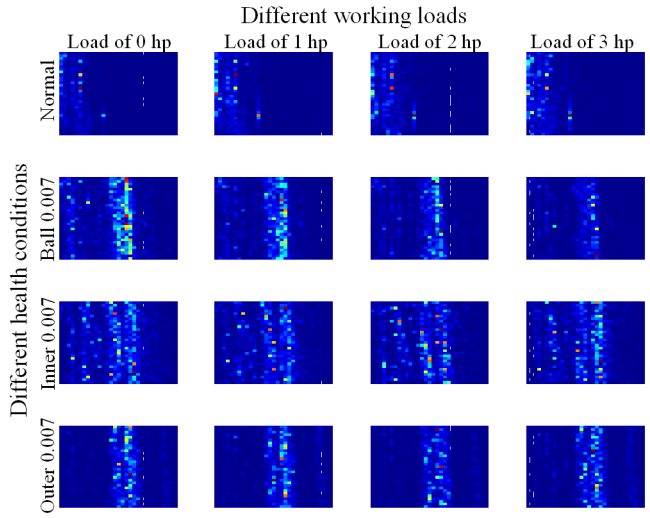

Fig. 3. WPI illustration for different bearing health conditions (normal, ball: 0.007 in, inner: 0.007 in, outer: 0.007 in) under different working loads (0, 1, 2, and 3 hp).

mathematically defined as follows:

$$\text{WPI}_i = \begin{bmatrix} E_i^0 & E_i^{q+1} & \cdots & E_i^{q(q-1)+1} \\ E_i^1 & E_i^{q+2} & \cdots & E_i^{q(q-1)+2} \\ \vdots & \vdots & \vdots & \vdots \\ E_i^q & E_i^{2q} & \cdots & E_i^{q^2} \end{bmatrix} \tag{7}$$

where i is the decomposition level and q is the dimension of the WPI with the size of $2^{i/2} \times 2^{i/2}$. In this multidimensional data construction process, the dynamic structure of WPI_i (including the frequency energy fluctuation and local relationship of the frequency energy pixels) can reflect the discriminant information for different classes, which is beneficial to fault diagnosis and performance degradation assessment in complex working conditions. Fig. 3 draws the WPI (WPI_{10} with the size of 32×32 pixels) illustration for different bearing health conditions (Normal, Ball 0.007 in, Inner 0.007 in, Outer 0.007 in) under different working loads (0, 1, 2, and 3 hp). It can be seen that these images have different brightness distributions. For each WPI, the horizontal axis stands for a global frequency distribution, while the vertical axis is a local frequency distribution. A special dynamic structure can be constructed from the local–global relationship. From the point of view of different working loads, it can be seen that WPI_{10} in the same health conditions shows a similar energy distribution although they are in four different load conditions. From the point of view of the different health conditions (different fault types), it can be also found that WPI_{10} in different health conditions have a different energy distribution. Moreover, all the WPI_{10} images show a sparse energy distribution. For example, the energy of WPI_{10} in the normal condition mainly concentrates in the left regions, while that in the inner race fault condition has a strip area. These WPI shows different local structures with textural features for different bearing health conditions. The energy distribution map is an intuitive description for the whole frequency bands. However, due to the complex working conditions, the local structures

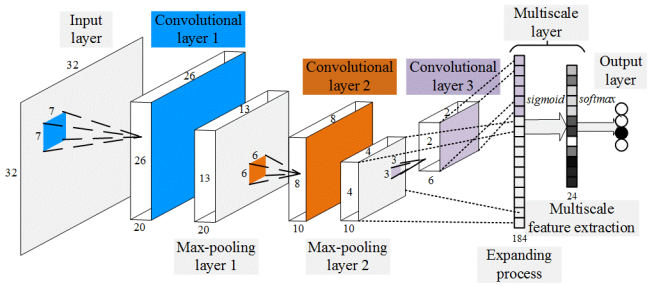

Fig. 4. Illustration of deep ConvNet architecture, which contains six hidden layers, including three convolutional layers, two max-pooling layers, and a multiscale layer. The multiscale layer contains unfold process and multiscale feature extraction. The width, height, and length of each cube represent the dimension of each feature map and the map number.

are not always the same and there will be evident difference for the same health condition under different loads. This will bring difficulties in bearing fault classification and diagnosis. Thus, the remaining work is to employ 2-D DL to obtain the invariant and discriminated features from the images.

B. Deep ConvNet for Multiscale Feature Extraction

In the related studies about bearing fault diagnosis based on NNs, only the neurons in the last or just one layer of the architecture are used as the final features. In this robust manner, these extracted features, maybe, lose many detailed pieces of information that existed in the interlayers. As seen in multiscale features [20] and deepID [12], the deep features in the last convolutional layer are much more invariant and robust (global information) than those in the low-level layers, which are beneficial to "big data" under complex operational conditions. However, the high-level features will lose some precise details and this will limit the information propagation in the nets, while the low-level features contain much precise details (local information) that are sensitive to the interference. Therefore, combining the skipping layer with the last convolutional layer as the input of the fully connected layer can maintain the global and local features simultaneously for more accurate classification.

In this paper, the deep ConvNet is employed as illustrated in Fig. 4, which contains six hidden layers, including three convolutional layers, two max-pooling layers, and a multiscale layer (including unfolded process and multiscale feature extraction). The features in the third convolutional layer and those in the second max-pooling layer are unfolded together as the input of the multiscale layer for multiscale feature extraction through an activation function. The WPI is used as the input of the deep ConvNet. Thus, the multiscale layer takes the function

$$z_j = \psi\left(\sum_i f_i^c \cdot w_{i,j}^c + \sum_i f_i^s \cdot w_{i,j}^s + b_j\right) \tag{8}$$

where f_i^c, $w_{i,j}^c$, f_i^s, and $w_{i,j}^s$ express the neurons and weights in the third convolutional layer and the second max-pooling layer, respectively. Finally, the output of the multiscale layer is put into the soft-max layer to predict the condition of the data, and the probability distribution of a given sample belonged to the lth class can be predicted as follows [12], [20], [26]:

$$y_l = \frac{\exp\ (\tilde{y}_l)}{\sum_{l=1}^{c} \exp(\tilde{y}_l)}, \quad \tilde{y}_l = \sum_{j=1}^{24} z_j\, w_{jl} + b_l \tag{9}$$

where 24 multiscale features are put into the neuron l of soft-max layer ($l = 1, 2, \ldots, c$, c is the dimension of the output layer, which is determined by the number of the bearing health conditions). The loss function (negative log-likelihood) is described as

$$J(\theta; f, y) = -\log h_\theta(f)_y. \tag{10}$$

By minimizing the error, the deep ConvNet can be also learned by the BP algorithm with gradient descent.

Especially, according to the reverse flow path of the error in the ConvNet, the feedback error E^{s2} in the second max-pooling layer contains two error terms: the indirect error $E^{m(c3)}$ and the direct error $E^{m(s2)}$ from the feedback error $E^m = E^{m(c3)} + E^{m(s2)}$ in the multiscale layer. Thus, the error of ith feature map in the second max-pooling layer, contributing to its feedback error E^{s2}, can be expressed as

$$\begin{aligned} E_i^{s2} &= Ib(E^{m(c3)})_i + b(E^{m(s2)})_i \\ &= \sum_{j=1} E_j^{c3} * K_{ij} + E_{(i)}^{m(s2)} \cdot \bullet\, \psi'_{s2} \\ &= \sum_{j=1} \left[\left(E_{(j)}^{m(c3)} \cdot \bullet\, \psi'_{c3} \cdot \bullet\, f'\right) * K_{ij}\right] + E_{(i)}^{m(s2)} \cdot \bullet\, \psi'_{s2} \end{aligned} \tag{11}$$

where $Ib(\bullet)/b(\bullet)$ stands for the indirect/direct error inverse transfer function, ψ' is the derivative of the activation function in the multiscale layer [as defined in (8)], f' is the derivative of the activation function in the third convolutional layer, and K_{ij} is the corresponding kernel function with 180° rotated. $E_{(j)}^{m(c3)}$, $E_{(i)}^{m(s2)}$ express the errors that belong to the jth feature map in the third convolutional layer and the ith feature map in the second max-pooling layer, respectively. Then, the error E^{s2} will be used to calculate the stochastic gradient of the weighs/basis in the second max-pooling layer and sequentially transferred into a lower level. Therefore, the architecture of the deep ConvNet can be learned by BP with the gradient descent strategy. As illustrated in Fig. 4, the WPI with the size of 32 × 32 is put into the trained deep ConvNet, a vector with the length of 184 is unfolded based on the combination of the previous two layers. Finally, a vector with the dimension of 24 in the multiscale layer (called as multiscale features) is further extracted to diagnose the bearing health condition.

C. Procedure of the Proposed EFMF Learning

The proposed EFMF learning is based on the deep ConvNet, which will simultaneously blend the global and local features with the robust invariants and precise details maintained. Specifically, the EFMF can capture the distribution variation and local structure among the WPI for much more robust and discriminant information although the multiple health conditions are under varying operational conditions. This is

TABLE I
DESCRIPTION OF DATA SETS FOR FAULT DIAGNOSIS

Dataset	Loads (hp) of training/testing data	Number of training/testing samples	Fault type	Fault size (inches)	Abbreviation
A	0/0	50/50	Normal	0	Normal
B	1/1	50/50	Ball	0.007	B/0.007
C	2/2	50/50	Inner	0.007	I/0.007
D	3/3	50/50	Outer	0.007	O/0.007
E	(0-3)/(0-3)	50/50	Ball	0.014	B/0.014
F	(0-2)/3	50/50	Inner	0.014	I/0.014
		50/50	Outer	0.014	O/0.014
		50/50	Ball	0.021	B/0.021
		50/50	Inner	0.021	I/0.021
		50/50	Outer	0.021	O/0.021

just a novel feature mining method in intelligent bearing fault diagnosis. In summary, the EFMF learning based on deep ConvNet contains the following three steps.

1) *Construct the Multidimensional Data:* Obtain the WPI of bearings under different health conditions according to (3). The WPIs compose the training set $\{x^i_{eimg}, y^i\}^L_{i=1}$, where x^i_{eimg} is the ith WPI of the training sample (with the sample length L) and y^i is the corresponding health condition label.
2) Build a deep ConvNet with multiple hidden layers, including three convolutional layers, two max-pooling layers and a multiscale layer. The multiscale layer combines the output of the last convolutional layer with the previous pooling layer. The decentered label training set with the size of $2^{i/2}\times 2^{i/2}\times L$ is used as the input of the net. Then based on the BP algorithm, (8)–(11) are used to minimize the loss function $-\log h_\theta(f)_y$ and fine-tune the architecture.
3) *Extract the EFMF:* Employ the multiscale feature in the multiscale layer of the trained deep ConvNet to diagnose the conditions of the bearings.

IV. PERFORMANCE VERIFICATION

A. Data Description

To evaluate the effectiveness of the proposed feature extraction scheme for machine fault diagnosis, several experiments were studied in this paper and the experimental data with multiple faults are provided by the Case Western Reserve University Bearing Data Center [27]. Single point faults of size 0.007, 0.014, and 0.021 in were set, respectively, on rolling element, inner raceway, and outer raceway of the drive-end bearings. The resulting vibration was measured at a sampling frequency of 12 kHz under four different operational conditions (0, 1, 2, and 3 horse hp). In this paper, six data sets (A–F, as listed in Table I) of the bearings are employed to validate the proposed method. There are 100 samples (50 for training and 50 for testing) for each health condition under each load. The signal length of each sample is 2048. A ten-level WPT (with Daubechies 8 wavelet [23], [24]) was employed on each signal to obtain 2^{10} WP reconstructed coefficients, which can construct the WPI with the size of $2^5\times 2^5$. Data sets A–D contain ten bearing health conditions under loads of 0, 1, 2, and 3 hp, respectively (there is a total of 1000 samples for each data set). There are four different loads for training data of data set E with 2000 samples. For data set F, the training data are acquired under three loads (0/1/2 hp), and the testing data are sampled from the 3 hp load.

In this paper, the designed deep ConvNet has six layers, including three convolutional layers, two max-pooling layers and a multiscale layer. For DL in the application of bearing fault diagnosis, Jia *et al.* [14] give a simple rule that the unit number of each layer becomes smaller with the layer increasing so that the information/characteristics can be compressed and mined in this dimension reduction process. According to this simple rule, the neuron number of the first convolutional layer, the second convolutional layer, and the third convolutional layer are 20, 10, and 6, respectively. Meanwhile, the corresponding kernel size of these three convolutional layers are, respectively, 7×7, 6×6, and 3×3. The unit number of the output layer is set to be ten according to the number of the health conditions. All the active functions are sigmoid functions, the number of training epoch is 100, the batch size is 25, and the learning rate is 0.2. As introduced in the procedure of Section III-C, $2^5\times 2^5$ WPIs are first extracted for each sample. Fifty percent of the samples are used to fine-tune the structure of the deep ConvNet, and the rest samples are used to verify the identifiability and discrimination of the deep features (multiscale features) extracted by the proposed scheme. Hence, there are totally 24 ($2\times 2\times 6$) features for each sample, which is employed as the input of the soft-max layer. And it should be noted that the following experiments were tested under the computation scenario of MATLAB 2015a on a desktop computer (64-b operating system, 3.4-GHz CPU, and 6-GB memory). Due to the complex networks, the computation time of the proposed ConvNet is nearly half an hour for data sets A–D and two hours for data set E (the size of data set E is four times those of the former data sets).

B. Multiscale Feature Learning Under Different Loads

For convenience and visual analysis, the dimension of the EFMF is later reduced to two by the PCA. Therefore, the achieved final feature results for the testing data of the data sets A–D are shown in Fig. 5. As a note, deep ConvNet-PC1 and deep ConvNet-PC2 stand for the first two principal components (PCs) extracted by PCA from the EFMF of the ConvNet. From Fig. 5, it can be seen that the feature distributions for the same health conditions are relatively concentrated, while those with different health conditions are effectively separated for these four different data sets. The ConvNet-PCs show a strong identifiability and high robustness for bearing health condition diagnosis under different working conditions. These mean that the extracted EFMF by the proposed scheme is quite suitable for bearing fault diagnosis with multiclass classification.

To verify the applicability of the proposed EFMF for complex multiclass classification, a ten-class classification, under various operating conditions (data set E), was analyzed in this

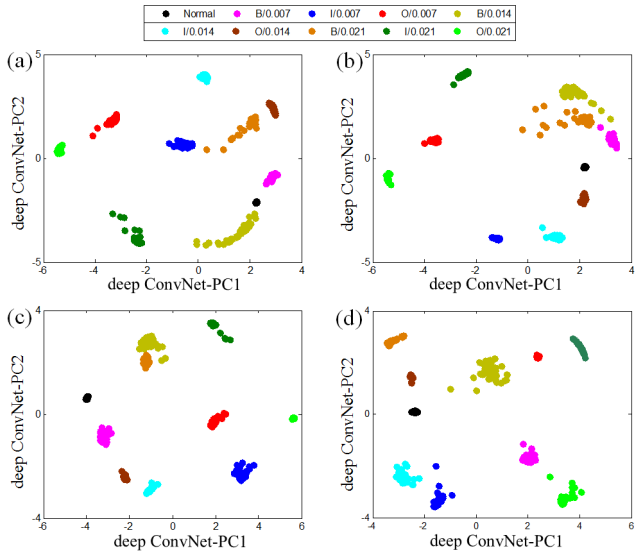

Fig. 5. Scatter plots of testing samples by the EFMFs based on deep ConvNet for ten different bearing health conditions under four different working loads (0, 1, 2, and 3 hp). (a) Data set A. (b) Data set B. (c) Data set C. (d) Data set D.

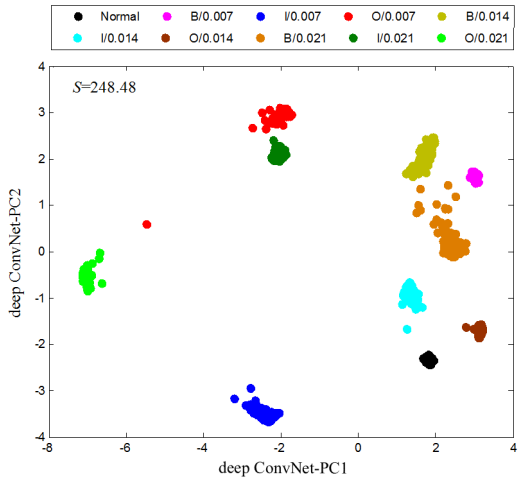

Fig. 6. Scatter plots of testing samples by the EFMFs ($S = 248.48$) based on deep ConvNet for data set E [including ten different bearing health conditions under four different working loads (0, 1, 2, and 3 hp)].

case. Besides the batch size of deep ConvNet that is set to 50, the other parameters of the deep ConvNet are the same with those mentioned above. As displayed in Fig. 6, it can be seen that the deep ConvNet-PCs also show an excellent clustering distribution among different classes. The samples can be well gathered together into the same pattern although there are four different operating conditions for each health condition. By the contrast analysis between Figs. 5 and 6, there is also an interesting phenomenon that deep ConvNet-PCs can have a relatively optimal and robust clustering effect along with the increase of the quantity and category of the original data, although the operational condition of the samples becomes complex. Meanwhile, another challenging task was tested here by using data set F. As shown in Fig. 7, it can be seen that

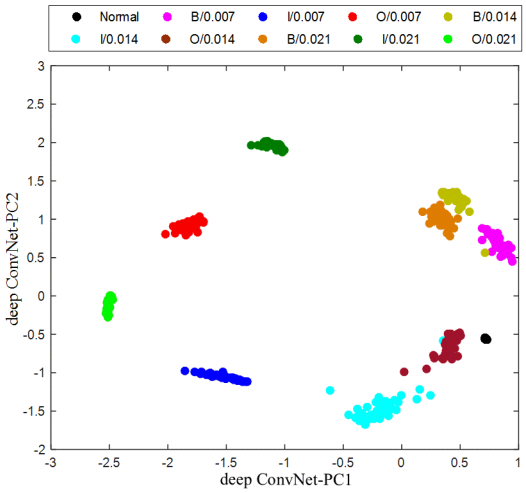

Fig. 7. Scatter plots of testing samples by the EFMFs based on deep ConvNet for data set F [including ten different bearing health conditions under one working loads (3 hp)].

the samples can be well identified although the samples are completely unknown for the ConvNet model. Based on the proposed model, the robust features regardless of loads can be effectively learned for an intelligent classification of new unknown samples. These robust features contain abundantly identifiable information among the massive fault data. This indicates that the deep ConvNet has a strong ability in robust feature representation for fault pattern classification.

C. Comparison With Other Features Under Various Loads

Furthermore, to demonstrate the merits of the proposed EFMF extraction method based on deep ConvNet, the results analyzed above have been compared with six other feature extraction methods, including PCA, linear discriminant analysis (LDA), BPNN, CNN, deep fusion feature I (combination of Conv3+Conv2 features in the architecture of CNN), and deep fusion feature II (combination of Conv3+Pool2 features in the architecture of CNN). Here, similar to the previous explanation in Section IV-B, the dimensions of all these features are also reduced to two for these six methods. Specifically, PCA is applied to reduce the dimension of the raw features before using LDA. The BPNN architecture has five layers, where the neuron numbers of three hidden layers are set to be 600, 200, and 100, respectively. The three other CNN-based feature extraction methods have the same parameters as the deep ConvNet except the multiscale layer in the deep ConvNet architecture. The fusion features I/II are employed here to further validate the advantage of concatenation network learning compared with that of this simple combination without network relearning.

The clustering distributions of the first two principal components based on the above six methods are displayed in Fig. 8. By observing from Figs. 6 and 8, it can be seen that the ConvNet-PCs (multiscale features) show a better clustering distribution among different classes while the samples of some classes overlap in the scatter plots for other methods. Specifically, for the feature distribution based on PCA [Fig. 8(a)], due to the change of the load, it can be seen that some

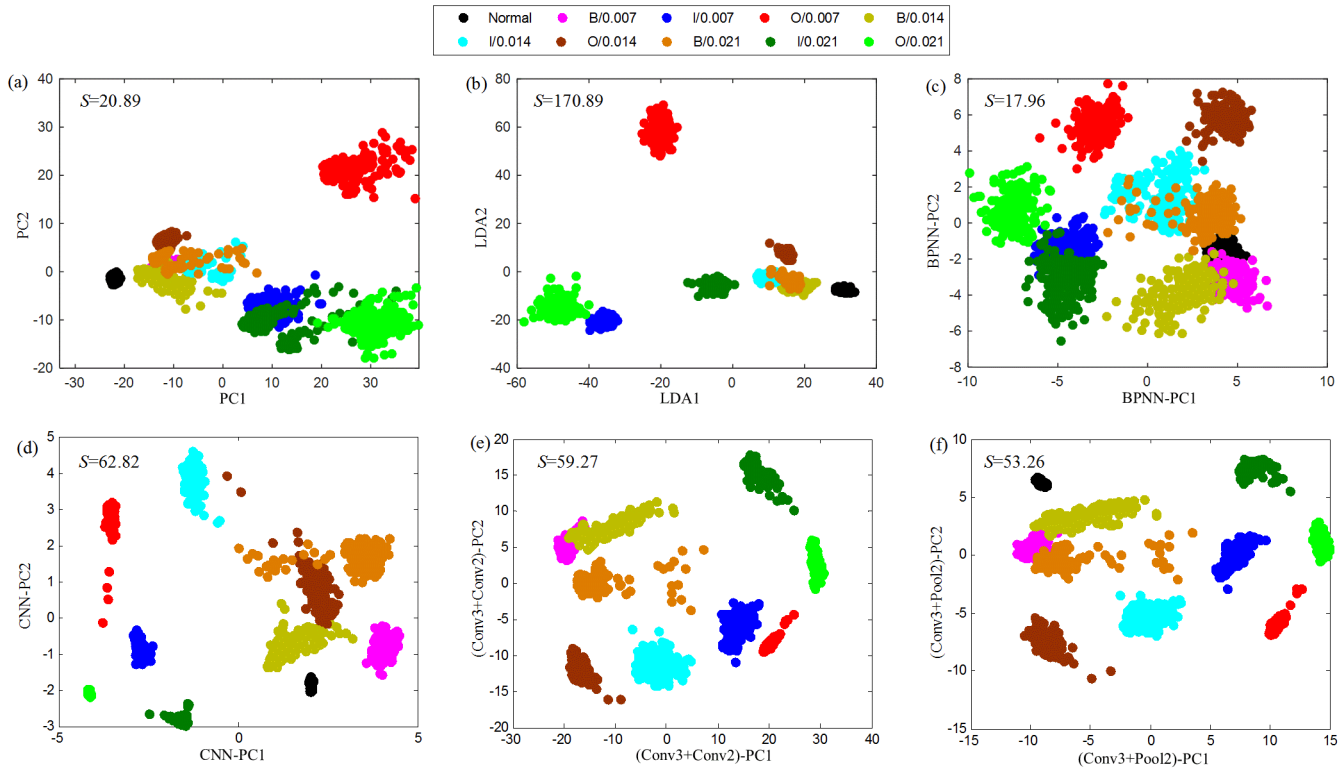

Fig. 8. Scatter plots of the extracted features for data set E by different methods. (a) PCA ($S = 20.89$). (b) LDA ($S = 170.89$). (c) BPNN ($S = 17.96$). (d) CNN. (e) Conv3+Conv2 features of CNN ($S = 59.27$). (f) Conv3+Pool2 features of CNN ($S = 53.26$).

of the samples from the same health condition cannot be gathered together and some of the samples from different health conditions cannot be separated. As a shallow layer NN algorithm compared with the DL methods, BPNN can only finitely mine the identifiable information for classification. Thus, the clustering distribution for some classes is overlapped with many classes as shown in Fig. 8(c). The CNN, two other deep fusion features I/II and the deep ConvNet are all based on the architecture of the ConvNet. According to Fig. 8(d)–(f), it can be seen that there also exist some overlapped samples among the classes. The deep fusion features I/II weaken the clustering distribution compared with those based on the basic CNN, which means that direct combinations of the output for these two layers cannot bring effective improvement for cluster analysis. Comparing Fig. 8(d) and (f) with Fig. 6, it can be also known that although both Conv3 and Pool2 layers are used in the proposed deep ConvNet and fusion feature I, concatenation with network relearning shows much better ability in feature mining. Furthermore, to numerically demonstrate the feature ability for clustering, we calculated a comprehensive evaluation factor S [23], [24], which combines between-class scatter and within-class scatter. Basically, a larger evaluation factor means that the separation of classes is clearer and the concentration of each class is better, which would benefit pattern classification. As pointed out in Figs. 6 and 8, it can be easily seen that the factor S from the multiscale feature has the biggest value 248.48 among these seven methods. Therefore, the proposed EFMF can effectively mine the identifiable characteristics of the bearings under varying operational conditions.

TABLE II

PERCENT ACCURACY OF THE DIAGNOSIS RESULTS FOR SIX DATA SETS BY SEVEN DIFFERENT FEATURE EXTRACTION METHODS BASED ON THE NEAREST MEAN CLASSIFIER

Methods	Dataset					
	A	B	C	D	E	F
PCA	85.6	75.8	94.2	86.8	84.75	70.2
LDA	95.4	91.2	99.6	98.4	92.75	46
BPNN	85.86	83.2	88.16	89.3	89.08	65.2
CNN	98.6	97.6	98.2	96.8	98.2	75
(Conv3+Conv2) feature I	94.8	92	88.6	97	91.55	80
(Conv3+Pool2) feature II	87	93.2	91.6	95.8	96.15	78.8
Deep ConvNet	98.8	98.8	99.4	99.4	99.8	96.8

D. Classification Accuracy Verification

To further evaluate the clustering effectiveness of the proposed EFMF for bearing fault diagnosis, the nearest mean classifier was simply applied. The accuracy of the diagnosis results of six data sets based on seven feature extraction methods are listed in Table II. It can be found that the results based on the multiscale features have the best accuracy for the testing samples and relatively stable effect for the five data sets. And it can be also seen that the features based on BPNN, CNN, feature II, and deep ConvNet for data set E have a relatively higher accuracies than the ones for data sets A–D. The classification accuracy based on deep ConvNet is concordant to the clustering performance of the features as mentioned

in Section III-B. Although the ten fault patterns of data set E are under various (four) operating loads, the network learning shows a potential in feature mining from these massive fault data instead of those simple data. Meanwhile, the proposed ConvNet can also achieve the highest identification accuracy for the unknown testing samples in data set F, which has the same performance with classification tasks in data sets A–E. The features mined by deep ConvNet can effectively ignore the influence of the operating conditions. These features are based on the architecture of the ConvNet, e.g., CNN, two other deep fusion features I/II, and the deep ConvNet. They show much better fault diagnosis effect than the BPNN, which further indicates that the DL method is much more effective than the shallow learning methods for the complex data. Moreover, for the same data set E, the maximal false alarm rate of one class can be, respectively, recalculated as 58% (PCA), 38% (LDA), 24% (BPNN), 10% (CNN), 36% (feature I), 23% (feature II), and 2% (deep ConvnNet) for seven features. This means that the proposed multiscale feature based on ConvNet can realize a much smaller error rate for each class. And the ConvNet can achieve a stable and high identification accuracy for these five data sets under different working loads. Therefore, it can be foreseen that these excellent performances of the proposed EFMF based on deep ConvNet in fault type identification is also beneficial to spindle bearing health monitoring and condition maintenance.

E. Discussion

1) Different from traditional vector characteristics and simple image characteristics, this paper rearranges energy vector of the WP nodes into an energy image by the PSR, and the features work well for spindle bearing fault classification. As we know, there will be no difference for feature extraction once the characteristic order of the vector is changed. That means except for the amplitude of the raw features, there is no other relevancy attribute among these features and some attribute of these features may be lost directly. For instance, there exists obvious frequency amplitude/energy attribute/distribution among the nearby frequency points/nodes in the FFT spectrum [14], [15] and WP nodes [16]. Nevertheless, this relevancy attribute among the features has been ignored. Thus, in this paper, an energy-fluctuated image in a WP phase space is first built to depict the brightness (frequency energy) fluctuation according to the health condition. This establishes a new local relationship of the nearby WP nodes, which cannot be expressed in other traditional vector features. These images are beneficial to 2-D improved CNN in mining deep features through the convolution kernel. The size selection of the energy-fluctuated image has not been studied in this paper. Here, we followed the experience that a square image with the size of 32×32 is chosen, which is common in image application. Therefore, this novel energy-fluctuated image construction is suitable for 2-D DL.
2) By grouping the convolutional layer 3, pooling layer 2 and "*sigmoid*" function together as a new multiscale layer, it can be seen that this layer is able to combine the global and local information for a much more robust feature with precise details. As for other fusion features of CNN, including feature I (Conv3+Conv2) and feature II (Conv3+Pool2), no more efficient effect can be reached but the performance decreased compared with the feature of CNN. This means that the fusion feature of multilayers in CNN may not contribute to the improvement of classification. Much more invalid and interference characteristics are mixed in this direct fusion. However, the multilayers (combination with the previous layers) are deeply integrated and mined again in the improved architecture of ConvNet. Certainly, many hidden layers and different concatenations can be used to improve the ability of the ConvNet. In this paper, we have not studied the optimal multilayer combination and architecture selection for an improved architecture of CNN. First, we just choose the last two layers, representing the global information (by the convolutional layer) and the local information (by the pooling layer), so that these two kinds of information can be deeply mined. Second, when building the architecture of CNN, we also followed the simple rule in [14]. These two problems are still hot topics in the architecture of NNs and we will do some related research in further study.
3) According to the shortcomings of NNs, including generalization ability and operating parameters selection, it can be foreseen that this deep ConvNet architecture maybe also suffer from the overfitting problem in some situations. To overcome this issue, many related techniques have been successfully imported in the architecture, such as rectified linear units (ReLU), tanh, Leaky ReLUs, maxout function, dropout technique, and neurons selective activation in statistics. Specifically, like the sparse brain NNs of animals and human, the sparse technique was introduced into the NN architecture for a more sparse and robust sensing. Meanwhile, these techniques can also be of benefit to reduce the number of activated neurons so that the computation time will be shortened. And regularization is also another appropriate choice. In this manner, this deep ConvNet architecture can be further enhanced to adapt measured data recognition accuracy in the real bearing application under various and complex situations.

V. Conclusion

This paper proposes a novel EFMF mining method for spindle bearing fault diagnosis. The major contribution of this new multiscale feature is that the WP energy is first rebuilt into a 2-D image space by the PSR, which can reconstruct a local relationship of the WP nodes and reveal the energy fluctuation in a new WP phase space. This dynamic structure builds an energy-fluctuated image and imports a local position attribute of the frequency pixels for DL in the application of 2-D image classification. According to brightness (frequency energy) fluctuation in these energy-fluctuated images, a deep ConvNet is further employed to mine similar statistical property among them. Different from the traditional CNN architecture,

the deep ConvNet combines the skipping layer with the last convolutional layer as the input of the multiscale layer. In this manner, the mined multiscale features are much more invariant and robust (global information) with precise details (local information). This means that these multiscale features can take advantage of the global and local features simultaneously and further reveal the essential identifiable characteristics from the energy-fluctuated image. The effectiveness of the EFMF is verified by six data sets from the spindle bearings that contain ten-class health conditions under four different loads. Through the comparison with six other features, the clustering distribution and classification accuracy of the proposed EFMF show an outstanding performance, which indicates that the proposed feature mining method is quite suitable for spindle bearing fault diagnosis with multiclass classification regardless of the load fluctuation.

REFERENCES

[1] N. Tandon and A. Choudhury, "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings," *Tribol. Int.*, vol. 33, no. 8, pp. 469–480, 1999.

[2] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, 2006.

[3] Y. Liu, B. He, F. Liu, S. Lu, and Y. Zhao, "Feature fusion using kernel joint approximate diagonalization of eigen-matrices for rolling bearing fault identification," *J. Sound Vibrat.*, vol. 385, pp. 389–401, Dec. 2016.

[4] A. Malhi and R. X. Gao, "PCA-based feature selection scheme for machine defect classification," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 6, pp. 1517–1525, Dec. 2004.

[5] M. Van and H.-J. Kang, "Wavelet kernel local fisher discriminant analysis with particle swarm optimization algorithm for bearing defect classification," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 12, pp. 3588–3600, Dec. 2015.

[6] A. Soualhi, K. Medjaher, and N. Zerhouni, "Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 52–62, Jan. 2015.

[7] L. Yuan, Y. He, J. Huang, and Y. Sun, "A new neural-network-based fault diagnosis approach for analog circuits by using kurtosis and entropy as a preprocessor," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 3, pp. 586–595, Mar. 2010.

[8] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, "Fault prognostic of bearings by using support vector data description," in *Proc. IEEE Conf. Prognostics Health Manage. (PHM)*, Jun. 2012, pp. 1–7.

[9] L. Ren, W. Lv, S. Jiang, and Y. Xiao, "Fault diagnosis using a joint model based on sparse representation and SVM," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 10, pp. 2313–2320, Oct. 2016, doi: 10.1109/TIM.2016.2575318.

[10] V. Sugumaran and K. I. Ramachandran, "Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing," *Mech. Syst. Signal Process.*, vol. 21, no. 5, pp. 2237–2247, 2007.

[11] H. Fan, Z. Cao, Y. Jiang, Q. Yin, and C. Doudou. (Mar. 2014). "Learning deep face representation." [Online]. Available: https://arxiv.org/abs/1403.2802

[12] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2014, pp. 1891–1898.

[13] A. Hannun *et al.* (Dec. 2014). "Deep speech: Scaling up end-to-end speech recognition." [Online]. Available: https://arxiv.org/abs/1412.5567

[14] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mech. Syst. Signal Process.*, vol. 72, pp. 303–315, May 2016.

[15] W. Sun, S. Shaoa, R. Zhaob, R. Yana, X. Zhangc, and X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, vol. 89, pp. 171–178, Jul. 2016, doi: 10.1016/j.measurement.2016.04.007.

[16] M. Gan, C. Wang, and C. Zhu, "Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 72, pp. 92–104, May 2016.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[18] D. C. Ciresan *et al.*, "Flexible, high performance convolutional neural networks for image classification," in *Proc.-Int. Joint Conf. Artif. Intell. (IJCAI)*, vol. 22. no. 1, 2011, p. 1237.

[19] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1725–1732.

[20] P. Sermanet and Y. LeCun, "Traffic sign recognition with multi-scale convolutional networks," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2011, pp. 2809–2813.

[21] E. C. C. Lau and H. W. Ngan, "Detection of motor bearing outer raceway defect by wavelet packet transformed motor current signature analysis," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 10, pp. 2683–2690, Oct. 2010.

[22] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. San Diego, CA, USA: Academic, 1999.

[23] Q. He, "Vibration signal classification by wavelet packet energy flow manifold learning," *J. Sound Vibrat.*, vol. 332, no. 7, pp. 1881–1894, 2013.

[24] X. Ding, Q. He, and N. Luo, "A fusion feature and its improvement based on locality preserving projections for rolling element bearing fault classification," *J. Sound Vibrat.*, vol. 335, pp. 367–383, Jan. 2015.

[25] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimensionfor phase-space reconstruction using a geometrical construction," *Phys. Rev. A, Gen. Phys.*, vol. 45, no. 6, pp. 3403–3411, 1992.

[26] N. Kalchbrenner, E. Grefenstette, and P. Blunsom. (Apr. 2014). "A convolutional neural network for modelling sentences." [Online]. Available: https://arxiv.org/abs/1404.2188

[27] *Bearing Data Center*, accessed on Mar. 2017. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/pages/download-data-file

Xiaoxi Ding received the B.S. degree in mechanical engineering from the University of Science and Technology of China, Hefei, China, in 2012, where he is currently pursuing the Ph.D. degree in instrumentation science and technology with the Department of Precision Machinery and Precision Instrumentation.

His current research interests include signal processing and data mining for health monitoring and fault diagnosis of mechanical systems.

Qingbo He (M'11) received the B.S. and Ph.D. degrees in mechanical engineering from the University of Science and Technology of China, Hefei, China, in 2002 and 2007, respectively.

He was a Research Associate with The Chinese University of Hong Kong, Hong Kong, a Post-Doctoral Research Associate with the University of Massachusetts, Amherst, MA, USA, a Post-Doctoral Fellow with the University of Connecticut, Storrs, CT, USA, and a Visiting Scholar with the University of Wisconsin, Madison, WI, USA. He is currently an Associate Professor with the Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China. His current research interests include a combination of machinery dynamics, signal processing, and data mining for intelligent monitoring, diagnosis, and control in precision machines, systems, or processes.

Dr. He was selected as a member of the Youth Innovation Promotion Association of the Chinese Academy of Sciences in 2016. He received the New Century Excellent Talents in University Award from the Ministry of Education of China in 2013.