IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ADVANCES IN PROGNOSTICS AND SYSTEM HEALTH MANAGEMENT

Received November 3, 2018, accepted November 24, 2018, date of publication December 12, 2018,
date of current version January 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2886343

Deep Decoupling Convolutional Neural Network for Intelligent Compound Fault Diagnosis

RUYI HUANG[1], YIXIAO LIAO[1], SHAOHUI ZHANG[2], AND WEIHUA LI[1], (Senior Member, IEEE)

[1]School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510641, China
[2]College of Mechanical Engineering, Dongguan University of Technology, Dongguan 523808, China

Corresponding author: Weihua Li (whlee@scut.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 51475170 and Grant 51875208.

ABSTRACT Intelligent compound fault diagnosis of rotating machinery plays a crucial role for the security, high-efficiency, and reliability of modern manufacture machines, but identifying and decoupling the compound fault are still a great challenge. The traditional compound fault diagnosis methods focus on either bearing or gear fault diagnosis, where the compound fault is always regarded as an independent fault pattern in the process of fault diagnosis, and the relationship between the single fault and compound fault is not considered completely. To solve such a problem, a novel method called deep decoupling convolutional neural network is proposed for intelligent compound fault diagnosis. First, one-dimensional deep convolutional neural network is employed as the feature learning model, which can effectively learn the discriminative features from raw vibration signals. Second, multi-stack capsules are designed as the decoupling classifier to accurately identify and decouple the compound fault. Finally, the routing by agreement algorithm and the margin loss cost function are utilized to train and optimize the proposed model. The proposed method is validated by gearbox fault tests, and the experimental results demonstrate that the proposed method can effectively identify and decouple the compound fault.

INDEX TERMS Compound fault decoupling, deep decoupling convolutional neural network (DDCNN), intelligent fault diagnosis, rotating machinery, decoupling classifier.

I. INTRODUCTION

In the context of intelligent industry, modern manufacture machines are becoming more and more complex, precise, efficient and intelligent than ever before [1], [2], in which rotating machinery is among the basic and significant part. Intelligent fault diagnosis of rotating machinery plays a crucial role for the security, high-efficiency and reliability of modern manufacture machines. At the same time, as a promising and powerful technology in the field of mechanical prognostic and health management, intelligent compound fault diagnosis methods have attracted more and more attention from industry and academia and obtained some achievements in recent years [3]–[6].

However, there are still great challenges in the intelligent compound fault diagnosis methods, which can be summarized as: (1) the critical components of rotating machinery, such as bearings and gears, are usually running under the condition of high speed, heavy load and harsh working environment, which causes the vibration signals collected by the accelerometer sensor are always complicated and coupled with heavily background noise; (2) the compound fault, as the mainly causes of catastrophic failure of rotating machinery, usually occurs when several key parts of the rotating machinery are damaged simultaneously, which are almost nonlinear compound and interfered with each other, and more dangerous and harmful than a single fault.

Unlike the traditional compound diagnosis methods based on advanced signal processing techniques, such as time-domain statistical feature extraction [7], spectral analysis [8], wavelet transform [9], [10], and empirical mode decomposition method [11]–[13], intelligent compound fault diagnosis can effectively identify the faults and automatically report the results [14], [15].

In general, there are two necessary steps for intelligent fault diagnosis of rotating machinery based on the state-of-the-art methods in recent years [16]–[20]: (1) feature learning using signal preprocessing skills or training a deep neural network, and (2) fault classification using pattern recognition

techniques [21]–[23]. Among the various intelligent algorithms, the most common ones used in fault diagnosis are k-Nearest Neighbor (k-NN) [24], Support Vector Machine (SVM) [25], [26], Artificial Neural Network (ANN) [27], and Deep Learning (DL) [28], [29]. Yaqub *et al.* [30] proposed an inchoate fault detection framework based on an adaptive feature extractor and k-NN classifier. Liu *et al.* [31] proposed an intelligent diagnosis method based on a short-time matching atom decomposition technique and SVM for bearing fault diagnosis. Shen *et al.* [32] proposed a new intelligent fault diagnosis scheme based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier, and the effectiveness of the intelligent fault diagnosis scheme is validated separately using datasets from bearing and gearbox test rigs. Zhu *et al.* [33] used three classification methods including classification, regression trees and radial basis function support vector machine (RBFSVM), to test ten fault datasets of rolling bearings. Li *et al.* [34] proposed a novel dimension reduction algorithm called feature denoising and nearest-farthest distance preserving projection for rotating machine fault diagnosis, and the effectiveness of the proposed method was validated by identifying compound faults in the locomotive bearing. Shao *et al.* [35] used a deep auto-encoder feature learning method to classify electrical locomotive roller bearing faults which contains compound bearing faults with multiple categories. A novel intelligent diagnosis method combined with compressed sensing and deep learning was proposed by Sun *et al.* [36], and validated by datasets from rolling element bearings with multiple fault categories. Ma *et al.* [37] proposed a deep coupling auto-encoder model that utilizing multisource sensory data to perform accurate fault diagnosis for rotating machinery.

Although traditional intelligent diagnosis methods have been succeeding in rotating machinery fault diagnosis, they still have some inherent limitations which can be listed as follows:

(a) The techniques in the field of fault diagnosis, especially the traditional compound fault diagnosis methods, focus on either bearing or gear fault diagnosis. However, a gear unhealthily will affect the bearing vibration signature collected by data acquisition system, and vice versa [38]. It is difficult to make a precise diagnosis if only analyzes the individual bearing or gear fault separately. Furthermore, in practical industrial applications, the faults may be in both gear and bearing, so the fault diagnosis methods may generate wrong results;

(b) In the field of intelligent fault diagnosis, the single fault and the compound fault are always regarded as independent fault patterns, and the relationship between them is not considered. Moreover, traditional intelligent diagnosis methods always depend on the labeled data of compound fault collected by time-consuming and labor-intensive experiments;

(c) Since the vibration signatures are usually contaminated by the background noise, the performance of traditional methods heavily depending on whether the learned and selected features are sensitive for the different categories;

(d) The traditional classifier can only output one label for a testing sample of compound faults, rather than multiple labels, which will cause the compound faults cannot be classified into two or more single faults. It means that the compound faults cannot be decoupled as the single faults in the process of fault diagnosis. Therefore, it is necessary to investigate and develop a new intelligent compound fault diagnosis method for rotating machinery.

As one of the intelligent fault diagnosis technology, deep learning has brought brilliant progress to the state-of-the-art in the field of fault diagnosis via a variety of neural networks, which has been widely used in fault classification, condition monitoring and remaining useful life prediction of rotating machinery [3], [39]. In other words, deep learning technology has a great capacity to overcome above inherent limitations. Inspired by Deep Convolutional Neural Network [40] and Dynamic Routing Between Capsules [41], a novel method called Deep Decoupling Convolutional Neural Network (DDCNN) is developed for intelligent compound fault identification. Firstly, one-dimensional deep convolutional neural network (1D DCNN) is employed as the feature learning model, which can effectively learn and extract the discriminative features from raw vibration signals. Secondly, multi-stack capsules are designed as the decoupling classifier (DC) to accurately identify and decouple the compound fault. Finally, the algorithm named routing by agreement and the margin loss cost function are used to train and optimize the proposed model. The proposed method is validated by the gearbox fault tests. To the best knowledge of the authors, this work is the first effort to identify and decouple the compound fault for rotating machinery automatically and efficiently.

The main contributions of the proposed method can be summarized as follows:

(1) To decouple the compound fault containing both bearing fault and gear fault, a decoupling classifier is designed to output multiple labels for samples belonging to a compound fault, which overcomes the limitations of **(b)** and **(d)**;

(2) The margin loss function and the routing by agreement algorithm are utilized to train and optimize the proposed model, which effectively enlarges interclass differences and reduces intraclass variations. The experimental results illustrate that the margin loss function and the routing by agreement algorithm enable the Deep Convolutional Neural Network (DCNN) to learn more sensitive and discriminative features from the raw vibration signals, which is effective to overcome the limitations of **(a)** and **(c)**;

(3) The most creative characteristic is that the proposed method can decouple the compound fault into single faults even though the model is trained on the dataset only containing normal and single fault samples. It can be effective to solve the problem where the labeled data of compound fault is insufficient or totally unavailable.

The rest of this paper is organized as follows. In section II, the basic theory of the standard DCNN is briefly introduced. Section III describes the proposed method in detail.

The experimental results are discussed and analyzed in section IV. Finally, a conclusion is drawn in section V.

II. DEEP CONVOLUTIONAL NEURAL NETWORK

The architecture of deep convolutional neural network (DCNN), a state-of-the-art method for fault diagnosis and classification of rotating machinery, is briefly introduced in this section. Taking the 1D DCNN shown in Figure 1 as example, it typically consisted of three kinds of layers: convolutional layer, pooling layer and fully connected layer. The 1D DCNN has been tremendously succeed in feature learning and classification from raw vibration signals due to its shared-weights architecture and translation-invariant characteristics [40]. Specifically, the 1D DCNN is a multi-step neural network consisting of two key steps: feature learning and softmax classification. The details about each step of 1D DCNN will be described as follows.

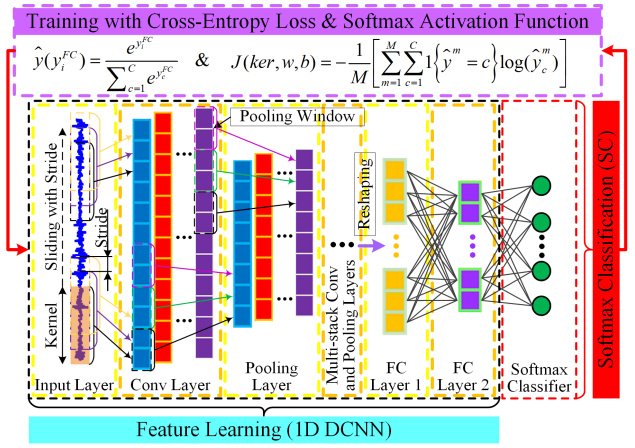

FIGURE 1. The structure of 1D DCNN.

A. FEATURE LEARNING

Assuming there is a sample $\boldsymbol{x} = [x_1, x_2, \cdots, x_N]^T \in \Re^{1\times N}$, the convolutional layer convolves the input vector x to the hidden feature maps $\boldsymbol{h} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_K] \in \Re^{K\times D}$ with the filter kernels and activation function (ReLU, Rectified Linear Unit). The output features $\boldsymbol{y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_K] \in \Re^{K\times D}$ can be described as follows:

$$\boldsymbol{h}_i = \boldsymbol{Ker}_i * \boldsymbol{x} + \boldsymbol{b}_i = \sum_{j=1}^{D} \boldsymbol{Ker}_i * \boldsymbol{x}(j) + \boldsymbol{b}_i \quad (1)$$

$$\boldsymbol{y}_i = ReLU(\boldsymbol{h}_i) = \max\{0, \boldsymbol{h}_i\} \quad (2)$$

where $\boldsymbol{Ker}_i$ and b_i denotes the weights and bias of the i-th filter kernel (i =1,2,..., K, K is the number of kernels), respectively, $\boldsymbol{h}_i$ denotes the i-th feature map, $\boldsymbol{x}(j)$ denotes the j-th local region signal (j=1,2,..., D, D is the dimension of the output feature), the notation $*$ denotes the dot product of the kernel and the local region signal, and $\boldsymbol{y}_i$ is the$i$-th output feature.

The output of convolutional layer $\boldsymbol{y}$ is then modified by the next pooling layer. In this paper, the average pooling is used to conduct a form of downsampling, and the output feature of average pooling layer $\boldsymbol{p} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_K] \in \Re^{K\times R}$ can be expressed as

$$\boldsymbol{p}_i = \sum_{r=1}^{R} \{\frac{1}{W} \sum_{t=(t-1)W+1}^{tW} \boldsymbol{y}_i(t)\} \quad (3)$$

where $\boldsymbol{y}_i(t)$ denotes the value of the t-th neuron in the i-th output feature of the former convolutional layer, W is the width of pooling window, $\boldsymbol{p}_i$ denotes the i-th pooling feature map, and R is determined by the width of pooling window.

The convolutional and pooling layers can be stacked layer by layer to learn the discriminative deep features and representations. After the multi-stack convolutional and pooling layers, the learned feature vectors are reshaped into one vector and such vector is used as the input of the fully connected layers. The process can be described as bellow:

$$\boldsymbol{y}^{FC} = ReLU(\boldsymbol{w}^{FC}(\boldsymbol{y}^{FC-1})^T + b^{FC}) \quad (4)$$

where $\boldsymbol{w}^{FC}$ and b^{FC} denote the weight matrix and bias of the fully connected layer, respectively, and $\boldsymbol{y}^{FC-1}$ is the output feature of the previous fully connected layer or the vector reshaped from the output feature of last pooling layer.

B. SOFTMAX CLASSIFICATION

Trained by the cross-entropy loss function, softmax classifier is widely used to make a classification in deep learning neural networks. Given a training set $\{\boldsymbol{x}^i, y^i_{true}\}_{i=1}^{M}$, where x^i and $y^i_{true} \in \{1, 2, \cdots, C\}$ is the$i$-th sample and label, respectively, and M is the number of samples. The softmax function $softmax(\cdot)$ can be represented as

$$\widehat{\boldsymbol{y}}_i = softmax(\boldsymbol{y}_i^{FC}) = \exp(\boldsymbol{y}_i^{FC}) / \sum_{c=1}^{C} \exp(\boldsymbol{y}_c^{FC}) \quad (5)$$

where $\boldsymbol{y}_i^{FC-1}$ is the output of the last fully connected layer, $\widehat{\boldsymbol{y}}_i$ is the output value of i-th neuron in the last fully connected layer, and C is the number of classes. The label outputted by softmax classifier can be calculated by

$$label_{softmax} = \arg\max(\widehat{\boldsymbol{y}}) \quad (6)$$

The cost function of DCNN is a cross-entropy cost function which shows stronger global optimization capability and faster convergence speed than other cost functions [42]. The model parameters are trained to minimize the cross-entropy cost function:

$$J(\boldsymbol{ker}, \boldsymbol{w}, \boldsymbol{b}) = -\frac{1}{M}\left[\sum_{m=1}^{M}\sum_{c=1}^{C} 1\left\{y^m_{true} = c\right\} \log(\widehat{\boldsymbol{y}}^m_c)\right] \quad (7)$$

where $1\{\cdot\}$ is the indicator function returning 1 if the statement is true, and 0 if the statement is false. More details about DCNN can be found in [43].

III. THE PROPOSED METHOD

In this paper, a new intelligent compound fault diagnosis technique named Deep Decoupling Convolutional Neural Network (DDCNN) is proposed for rotating machinery. This

method includes two key parts: feature learning and decoupling classification. The theory of the feature learning model designed in this method is almost the same as the previous one (DCNN) introduced in section II except that the latter reshapes the output features of last pooling layer into a vector used as the input of the softmax classifier and the former reshapes the output features into a matrix used as the input of the decoupling classifier. In the process of decoupling classification, a novel decoupling classifier is constructed to identify the single fault or decouple the compound fault. Furthermore, in order to learn more sensitive features from the raw vibration signals, the margin loss function and the routing by agreement algorithm are used to train and optimize the proposed model.

A. DECOUPLING CLASSIFICATION

To decouple the compound fault into different single faults, an urgent is required to design an intelligent decoupling classifier for rotating machinery. Inspired by Dynamic Routing Between Capsules [41], the algorithm of routing by agreement is applied to construct the decoupling classifier. The theories and details of the decoupling classifier are described as follows.

Firstly, for a matrix $\boldsymbol{y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_{K_l}]^T \in \Re^{K_l \times R_l}$ obtained from reshaping the output feature of the last pooling layer, where $\boldsymbol{y}_i \in \Re^{R_l \times 1}$ is the i-th feature map (a vector) of the last pooling layer, K_l denotes the kernel number of the former layer, and R_l is determined by the width of pooling window. The output tensor of decoupling classifier $\boldsymbol{d} \in \Re^{C \times R_d}$ is a weighted sum over all middle prediction vectors $\widehat{\boldsymbol{y}} \in \Re^{C \times K_l \times R_d \times 1}$, and this process can be described as

$$\boldsymbol{d}_j = \sum_{i=1}^{K_l} c_{ij} \widehat{\boldsymbol{y}}_{j|i} \tag{8}$$

$$\widehat{\boldsymbol{y}}_{j|i} = \boldsymbol{W}_{ij} \boldsymbol{y}_i \tag{9}$$

where $i = 1, 2, \cdots, K_l$, $j = 1, 2, \cdots, C$, $\boldsymbol{W}_{ij} \in \Re^{R_d \times R_l}$ is weight matrix between $\widehat{\boldsymbol{y}}_{j|i} \in \Re^{R_d \times 1}$ and y_i, all the weight matrix consist the weight tensor $\boldsymbol{W} \in \Re^{C \times K_l \times R_d \times R_l}$, R_d is the length of $\boldsymbol{d}_j$, $\boldsymbol{d}_j$ denotes the j-th output vector of decoupling classifier, and $\boldsymbol{c}_{ij}$ are coupling coefficients that are determined by the iterative dynamic routing process.

The coupling coefficients c_{ij} between $\boldsymbol{y}_i$ and all the outputs of decoupling classifier sum to 1 and are updated by a routing softmax whose initial logits b_{ij} are the log prior probabilities. Such log prior probabilities can be discriminatively learned as the other weights at the same time. The function of routing softmax can be expressed as

$$c_{ij} = softmax(b_{ij}) = \exp(b_{ij}) / \sum_{c=1}^{C} \exp(b_{ic}) \tag{10}$$

A non-linear squash function is designed as the activation function to transform the short vectors to almost zero length and long vectors to a length close 1, which is described as follow

$$\boldsymbol{v}_j = squash(\boldsymbol{d}_j) = \frac{\|\boldsymbol{d}_j\|^2}{1 + \|\boldsymbol{d}_j\|^2} \frac{\boldsymbol{d}_j}{\|\boldsymbol{d}_j\|} \tag{11}$$

The initial coupling coefficients are then iteratively updated by measuring the agreement which is the scalar product $\boldsymbol{a}_{ij} = < \boldsymbol{v}_j, \widehat{\boldsymbol{y}}_{j|i} >$. This agreement is added to the initial logit b_{ij} before computing the new values for all the coupling coefficients linking $\boldsymbol{y}_i$ to $\boldsymbol{d}_j$.

Secondly, the L_2 norm is used to convert the output tensor of decoupling classifier $\boldsymbol{d}$ into a final prediction vector $\boldsymbol{y}_{pred} = [y^1, y^2, \cdots, y^C] \in \Re^{C \times 1}$. Each magnitude of y^i in $\boldsymbol{y}_{pred}$ represents the probability whose input sample belongs to the i-th category. The probability of the predicted category becomes larger when the magnitude y^i is closer to 1.

Finally, a threshold φ is selected to restrict the number of predicted labels for classification. The classifier will output the label of the i-th class to 1 if the y^i is larger than the selected threshold, and 0 for the others. For a reliable classification, a good prediction should be made at the situation that the corresponding y^i of actual classes are almost reached to 1, and others to 0. The procedure of the decoupling classifier algorithm are summarized in Table 1 and demonstrated in Figure 2.

TABLE 1. The algorithm of decoupling classifier.

Algorithm: The decoupling classifier

Input: The reshaped feature $\boldsymbol{y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_{R_l}]^T \in \Re^{K_l \times R_l}$ obtained from the last pooling layer, the number of classes C, the number of iteration r, the threshold φ, the weight tensor $\boldsymbol{W} \in \Re^{C \times K_l \times R_d \times R_l}$ which can be optimized by the backpropagation algorithm, $i = 1, 2, \cdots, K_l$, $j = 1, 2, \cdots, C$.

Output: The predicted labels

Initialization: for all $b_{ij} \leftarrow 0$

For r iterations **do**

for i in $[1, 2, \cdots, K_l]$: $\boldsymbol{c}_i \leftarrow softmax(\boldsymbol{b}_i)$ (10)

for i in $[1, 2, \cdots, K_l]$ and j in $[1, 2, \cdots, C]$: $\widehat{\boldsymbol{y}}_{j|i} \leftarrow \boldsymbol{W}_{ij} \boldsymbol{y}_i$ (9)

for i in $[1, 2, \cdots, K_l]$ and j in $[1, 2, \cdots, C]$: $\boldsymbol{d}_j \leftarrow \sum_i^{K_l} c_{ij} \widehat{\boldsymbol{y}}_{j|i}$ (8)

for j in $[1, 2, \cdots, C]$: $\boldsymbol{v}_j \leftarrow squash(\boldsymbol{d}_j)$ (11)

for i in $[1, 2, \cdots, K_l]$ and j in $[1, 2, \cdots, C]$: $b_{ij} \leftarrow b_{ij} + < \boldsymbol{v}_j, \widehat{\boldsymbol{y}}_{j|i} >$

Return v_i

L2 norm: $\boldsymbol{y}_{pred} \leftarrow \|\boldsymbol{v}\|_2$

If $y^i_{pred} \geq \varphi$

return i-th class label as 1

else

return i-th class label as 0

Return The predicted labels

Note that the main difference between softmax classifier and decoupling classifier is the number of the label that the classifier can output. The softmax classifier can only output a single label due to using the softmax activation function and the arguments of the maxima (argmax) algorithm. Instead, decoupling classifier can output single or multiple labels for a testing sample because of multi-stack capsules and the algorithm of routing by agreement. Assuming the square, circle and triangle indicate three different single faults, and the overlapping pattern of square and triangle represents the com-

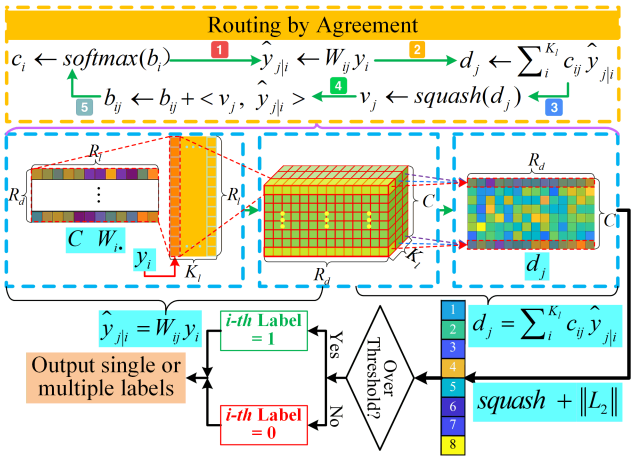

FIGURE 2. The flowchart of decoupling classifier.

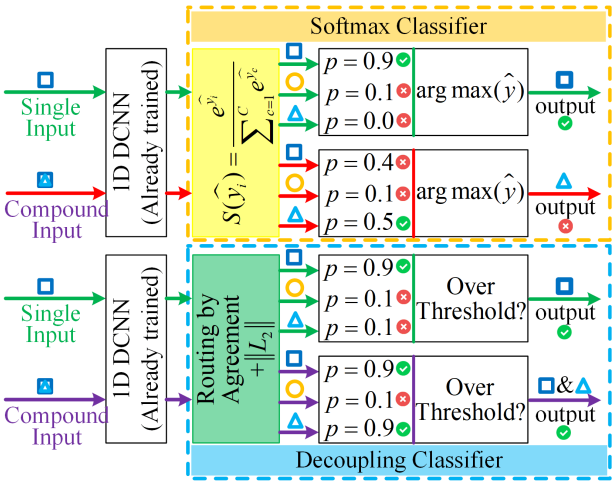

FIGURE 3. The difference between softmax and decoupling classifier.

pound fault coupled by two different single faults. As illustrated in Figure 3, the decoupling classifier can identify and decouple the compound fault accurately via outputting two different labels which represents the two fault types coupling the compound fault, but the softmax classifier identifies the compound fault to a false category because it cannot decouple the compound fault correctly.

B. MARGIN LOSS FUNCTION

Based on the decoupling classifier algorithm and inspired by [44], the cost function named margin loss function is adopted to design the DDCNN cost function for the multi-label prediction. Different from cross-entropy cost function, margin loss function is based on Euclidean distances that directly measures the similarity of different categories. This new loss function enlarges interclass differences and reduces intraclass variations effectively, which can easily implement the DDCNN training with multi-label. Hence, the model is trained by minimizing the cost function

$$J = \sum_{c=1}^{C} L_c = \sum_{c=1}^{C} \{T_c \max(0, m^+ - \widehat{\mathbf{y}}_c)^2 + \lambda(1 - T_c)\max(0, \widehat{\mathbf{y}}_c - m^-)^2\} \quad (12)$$

where T_c is an indicator function, T_c=1 means that an object of class c is present and T_c=0 means not, m^+ and m^- denotes the lower and upper boundary of $\|v_c\|$, respectively, and λ is regular parameter down-weighting of the loss for absent object classes. In this paper, m^+=0.9, m^-=0.1 and $\lambda = 0.25$, which means that if an object of class c is present, then $\|v_c\|$ should be no less than 0.9, and if not, $\|v_c\|$ no more than 0.1.

C. THE GENERAL PROCEDURE OF DDCNN

The architecture of the proposed method is shown in Figure 4 and the procedure is listed as bellows.

Step 1: The raw vibration data of rotating machinery are acquired by BBM-PAK test system with acceleration sensors.

Step 2: The original vibration data collected in the normal condition and the single fault condition are sliced to obtain huge amounts of training samples, each with a length of 8192 and 0.5 overlap rate. Meanwhile, the original vibration data collected in the compound fault condition are sliced with the same length and overlap rate to obtain huge amounts of testing samples.

Step 3: DDCNN is constructed and trained by the training dataset which only contains the normal samples and the single fault samples.

Step 4: All of the testing samples including the normal samples, the compound fault samples and the single fault samples are employed to validate the superiority and effectivity of the proposed method, and the decoupled result of compound faults will be reported automatically.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. DESCRIPTION OF EXPERIMENTAL DATASET

Bearing and gear are the most common and critical machine components in condition monitoring of rotating machinery, such as gearboxes of all automobiles, high-speed rails, aircrafts, machine tools, wind turbines, which are easy to degrade during the operation. In this experiment, a gearbox dataset consists of bearing fault (inner race defect), gear fault (chipped tooth defect) and a compound fault (bearing and gear fault) have been collected on a five-speed automobile transmission and used to verify the accuracy and efficiency of the proposed method. The automobile transmission is operated at 1000rpm of shaft speeds and 50N • m of the output shaft load torque. The original vibration signals are collected with a sampling frequency of 24 kHz by an accelerometer mounted on the gearbox case near the output shaft. Figure 5 is the test rig of the automobile transmission and the sensor configuration. The raw vibration data collected by BBM-PAK test system are separated into training samples, including normal samples, gear fault samples and bearing fault samples, and testing samples containing both single gear or bearing fault samples and compound fault samples. Details about the gearbox dataset are listed in Table 2.

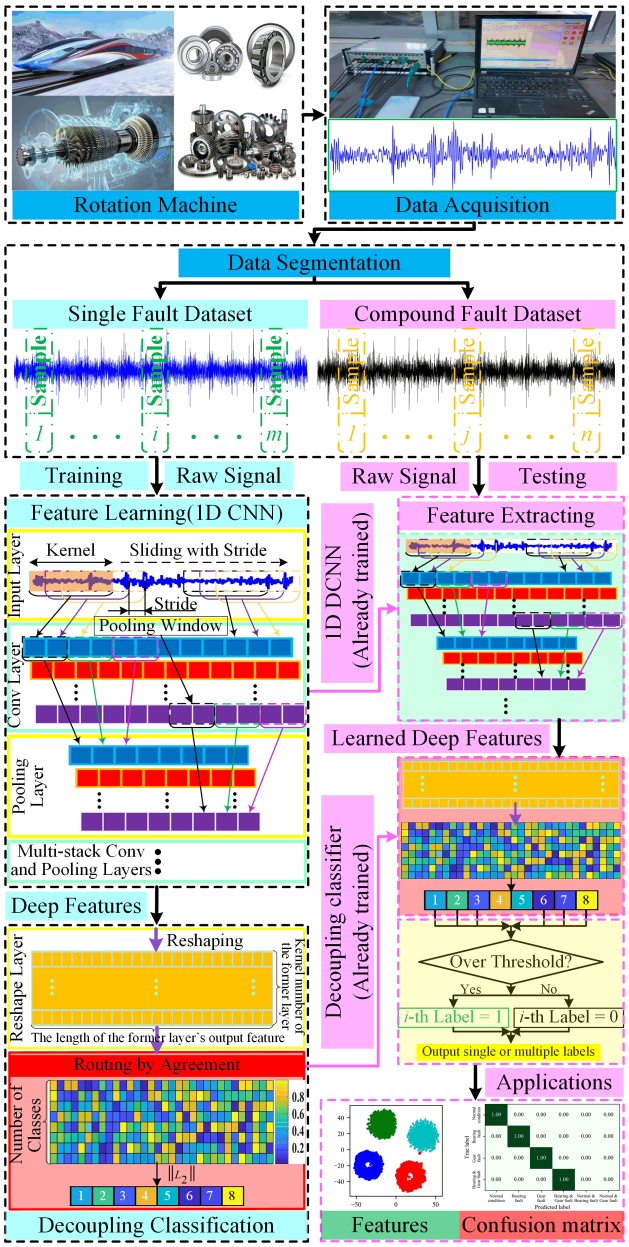

FIGURE 4. The architecture of DDCNN.

It is worthy to note that the training dataset only contains single faults and normal samples, which means that the proposed model is trained without the compound fault. After the DDCNN model being trained, the compound fault samples will be used to test the performance of DDCNN model. All the labels of training dataset are single labels which can be used as the supervised indicator to calculate the margin loss and supervise the model to learn the discriminative features. To indicate the compound fault coupled by single faults, multiple labels are used to represent the compound fault in the testing dataset which would be used to validate the performance of the proposed method in the task of decoupling compound fault.

(a) (b)

FIGURE 5. (a) The test rig of the automobile transmission. (b) sensor configuration.

TABLE 2. Description of the gearbox dataset.

Datasets	Health conditions			
	Normal condition	Bearing fault	Gear fault	Compound fault
Training	715	715	715	
Testing	715	715	715	715
Labels	1	2	3	2 & 3

B. PARAMETER OF THE PROPOSED METHOD

As illustrated in section III, part *A*, the first key step is feature learning. The model of 1D DCNN is designed to learn and extract the deep discriminative and sensitive features from the original vibration signals. In the second step, the learned features are used as the input of decoupling classifier for fault classification. The hyper-parameters of the proposed method are determined by grid-search in scikit-learn which is a model hyper-parameter optimization technique. As shown in Table 3, the architecture of the proposed DDCNN model utilized in this experiment contains two convolutional-pooling layers followed by a reshape layer and two decoupling classification layers. The first convolutional kernel size of the first convolutional-pooling layer is 256×64×16, and the second one is 256×3×1. The size of convolutional kernels in the second convolutional-pooling layer is 128×3×1. The activation function is ReLU or squash function, and the first and second pooling layers are the average pooling and max pooling layers, respectively. After two convolutional-pooling layers, a reshape layer and two

TABLE 3. The parameters of DDCNN.

Layer Type	Activation Function	Parameter Name	Parameter Size	Output Size
Input	/	/	/	(8192,1)
Conv1d_1	ReLU	Kernels	256×64×16	(256,512,1)
Conv1d_2	ReLU	Kernels	256×3×1	(256,512,1)
Average Pooling	/	Pooling size	2	(256,256,1)
Conv1d_3	ReLU	Kernels	128×3×1	(128,256,1)
Conv1d_4	ReLU	Kernels	128×3×1	(128,256,1)
Max Pooling	/	Pooling size	2	(128,128,1)
Reshape	/	/	/	(128,128)
DC_1	squash	Vectors	32×128	(32,128)
DC_2	squash	Vectors	3×32	(3,32)
L2 norm	/	/	/	(3,1)

decoupling classification layers are used for fault classification. The algorithm is implemented using the Keras toolbox. The max training epoch is 20 with a batch-size of 64. In order to minimize the cost function, the Adam optimizer [45] with default parameter settings is applied to train and optimize the DDCNN model.

After selected the hyper-parameters mentioned above, there are still two significant parameters for implementation of proposed method. The first one is the threshold φ. The higher threshold means a higher likelihood for the classification to be right. However, a larger threshold brings less predictions, as well as an increased probability of error rate. Therefore, there is a tradeoff between the accuracy of the classification and the reliability of the prediction result. $\varphi = 0.7$ is recommended in this paper. The another is the number of iterations r. According to [41], r can be selected to 3 because the decoupling classifier algorithm with 3 iterations of dynamic routing process can perform the faster convergence and lower loss.

C. EXPERIMENTAL RESULTS AND ANALYSIS

After choosing the hyper-parameters of the DDCNN, the superiority and effectivity of the proposed model and the features learned and extracted by the feature learning model are investigated and compared with the state-of-the-art intelligent fault diagnosis method (DCNN) here. The cost function and classifier of DDCNN are the margin loss function and decoupling classifier, respectively, whereas those of DCNN are cross-entropy function and softmax classifier respectively. Besides these function and classifier, the other parameters are the same for a fair comparison.

1) DIAGNOSIS RESULTS AND ANALYSIS

To illustrate the superiority of the DDCNN, the gearbox dataset which contains the normal, bearing fault, gear fault and compound fault samples is used to train and test the performance of the DDCNN and DCNN. It is necessary to point out that the proposed method uses the training dataset only consisting of the normal samples and the single fault samples to train the DDCNN and DCNN, but uses the testing dataset contains both the training and the compound samples to test the DDCNN and DCNN.

In this experiment, ten trials are carried out for each method, and the evaluation metric is classification accuracy. The multi-class confusion matrix contains classification accuracy and misclassification error, which is a significant way to visualize the classification result of all conditions. The horizontal axis of the confusion matrix represents the predicted labels of samples, while the longitudinal axis represents the true labels of samples. The right color bar in the figure indicates the correspondence between values and colors. In Figures 6 (a) and (b) presents the multi-class confusion matrix of the traditional DCNN and the proposed method. Figure 6 (a) illustrates that the signal fault samples can be accurately classified into the actual categories by the traditional DCNN, but the compound fault samples

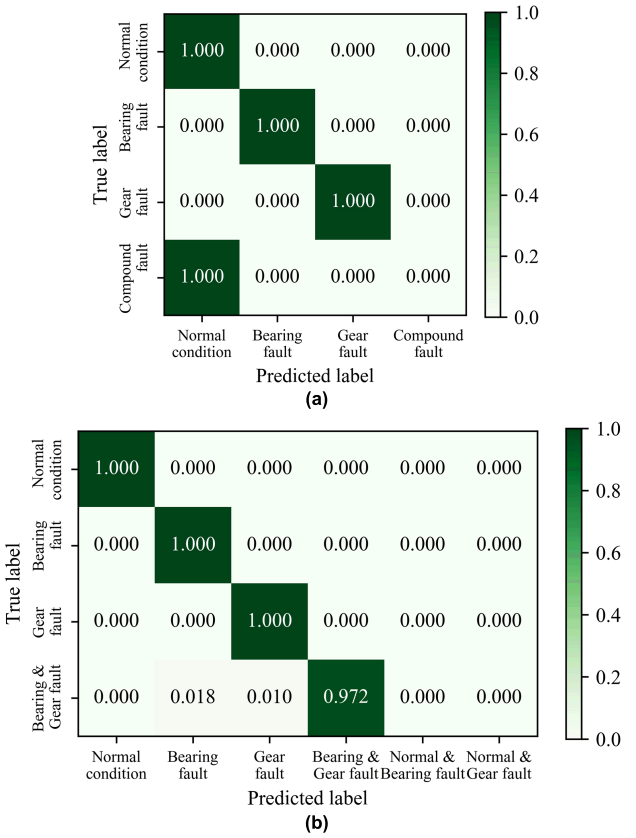

FIGURE 6. The confusion matrix (a) DCNN; (b) DDCNN.

are totally confused with the normal samples. On the contrary, Figure 6 (b) shows that the accuracy of the proposed method for the compound fault identification and decoupling is 97.2%, which is completely higher than that of DCNN. In other words, the DDCNN model can make a precise diagnosis for the compound fault. Specifically, 2.8% of errors are caused by the false classification (1.8%) of the compound fault with the bearing fault and gear fault (1.0%), respectively. Comparing Figure 6 (a) and 6 (b), a conclusion can be drawn that the proposed method achieves the superior and competitive results for the gearbox fault diagnosis.

In addition, from the comparison of DCNN and DDCNN, a conclusion can also be drawn that the compound fault can be decoupled into multiple-single faults by the proposed method. In order to further illustrate the powerful ability of DDCNN on decoupling compound fault, the predicted and actual labels of the two methods are shown in Figures 7 (a) and (b). It can be found that DDCNN completely decouples the compound fault coupled by bearing and gear faults into two single faults. However, DCNN classifies the all of compound samples into the normal category because of the DCNN inherent limitations mentioned in section I. These results demonstrate that the proposed method has a powerful ability for decoupling the compound fault for rotating machinery. Moreover, it might be a great progress for the compound fault diagnosis using artificial intelligent skills.

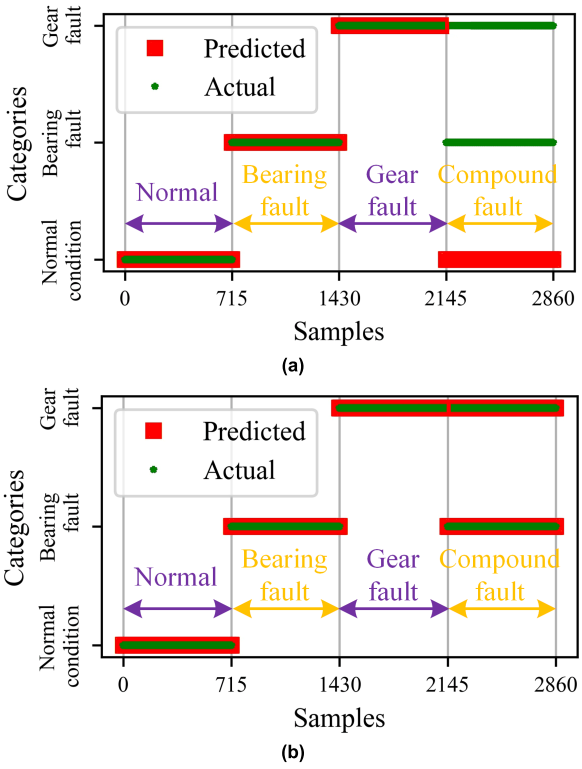

FIGURE 7. The predicted and actual label of (a) DCNN; (b) DDCNN.

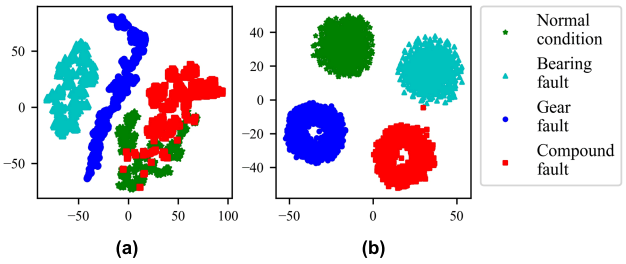

FIGURE 8. The clustering results using T-SNE (a) DCNN. (b) DDCNN.

2) THE LEARNED DEEP FEATURES EVALUATION

It is a critical step to learn and obtain discriminative features for intelligent compound fault identification. The quality of the deep features learned by DCNN and DDCNN is compared and evaluated in this part. A new dimensionality reduction technique called t-Distributed Stochastic Neighbor Embedding (t-SNE) is used to visualize the high-dimensional learned features [46]. Taking the last trial as an example, Figures 8 (a) and (b) show the clustering results of DCNN and DDCNN respectively. The deep features learned by the DCNN show some overlaps between the normal condition and the compound fault samples. Fortunately, the clustering result of the features learned by DDCNN is obviously separable, which is better than that of DCNN. Thus, it could be easily obtained a conclusion that the deep features learned by the DDCNN can represent the original data in a more discriminative way than that of learned by the DCNN. This is mainly because the margin loss function, rather than cross-entropy loss function, is used as the cost function to train and optimize the proposed model. Margin loss function is able to make the centers of interclass distributed separately with large margins by adding the interclass constraint, which also facilitates the DDCNN model converging to a generalization optimal result accurately.

In addition, six deep features learned by the two methods are randomly selected from the last layer of the feature learning model and displayed in Figures 9 (a) and (b) respectively. The horizontal axis of each sub-figure represents the number of samples, and the longitudinal axis represents the value of features. Comparing these deep features, it can be seen that the normal condition and the compound fault features learned by DCNN are always overlapped. Moreover, as shown in the last two features of Figure 9 (a), the features which are totally overlapped with all categories are completely indistinguishable. This is reason for that the DCNN model classifies the compound fault samples into the normal samples and generates wrong results. By contrast, the deep features learned by DDCNN, as shown in Figure 9 (b), are more discriminative than that of DCNN. On one hand, there are also some features indistinguishable, such as the second and the last features, possibly since the vibration signatures are usually influenced by the heavy background noise or the common features of different categories have been learned. On the other hand, the value of compound fault features either close to the bearing fault features or the gear fault features, such as the first, third, fourth and fifth feature, which indicates that the common features between the bearing fault and compound fault, as well as the gear fault and compound fault, are learned by the DDCNN. From the above analysis and discussion above, it can be found that DDCNN can effectively

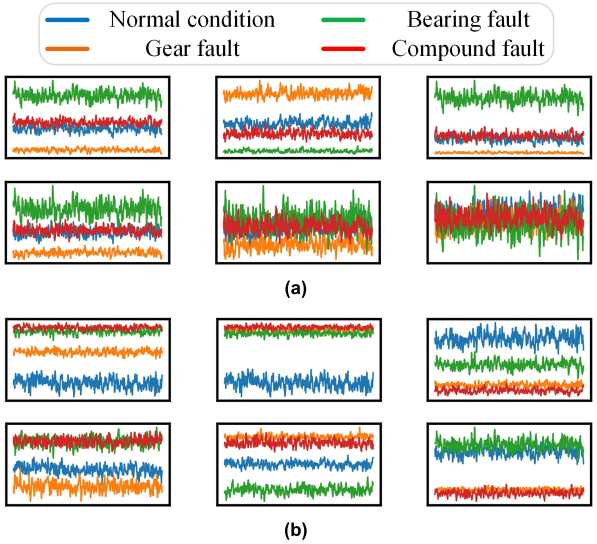

FIGURE 9. The deep features (a) DCNN; (b) DDCNN.

learn the decoupled features from the raw compound fault vibration signals.

From the figures 8 and 9, the performance of the DDCNN on learning and extracting the deep features is presented in qualitative analysis. To further evaluate the discriminability of deep features in quantitative analysis, two parameters including between-class covariance S_b and within-class covariance S_w are calculated. Mathematically, assuming there is a reduced feature matrix $\boldsymbol{f} = [\boldsymbol{f}_1, \boldsymbol{f}_2, \cdots, \boldsymbol{f}_N]$, where N is the total number of sample, the formulas of the two parameters are described as follows [47]:

$$S_b = \sum_{c=1}^{C} N_c(\boldsymbol{m}_c - \boldsymbol{m})(\boldsymbol{m}_c - \boldsymbol{m})^T \tag{13}$$

$$S_w = \sum_{c=1}^{C} \sum_{n \in C_c} (\boldsymbol{f}_n - \boldsymbol{m}_c)(\boldsymbol{f}_n - \boldsymbol{m}_c)^T \tag{14}$$

with

$$\boldsymbol{m}_c = 1/N_c \sum_{n \in C_c} \boldsymbol{f}_n \tag{15}$$

$$\boldsymbol{m} = 1/N \sum_{n=1}^{N} \boldsymbol{f}_n \tag{16}$$

where C is the number of class, N_c is the number of samples in class C_c, m_c is the mean for samples in c th class, and m is the mean for total samples.

The between-class covariance is used to indicate the scattered degree among different classes, while the within-class covariance is used to describe the concentrated degree in the same class. Specifically, a larger between-class covariance denotes the separation of different classes is clearer and a smaller within-class covariance indicate that the concentration of each category is stronger [48]. In this paper, four evaluation indexes are calculated combined with the between-class and within-class covariances, which are used to quantitatively describe the quality of the learned features. The formulas of these indexes are defined as

$$J_1 = Tr[S_w^{-1} S_b] \tag{17}$$

$$J_2 = |S_b| \,/\, |S_w| \tag{18}$$

$$J_3 = Tr[S_b]/Tr[S_w] \tag{19}$$

$$J_4 = |S_w + S_b| \,/\, |S_w| \tag{20}$$

where $Tr(A)$ means the trace of the matrix A.

In the testing stage, the four evaluation indexes are calculated based on (17) - (20). As shown in Table 4, the four evaluation indexes of the deep features learned by DDCNN are 9.79, 2.32, 2.58 and 2.53, respectively, whereas those of the deep features learned by DCNN are only 5.33, 1.98, 1.90 and 2.23. According to the evaluation criterion, the classification is better when a larger J_i is obtained. The above results demonstrate that the proposed method has a powerful ability in learning the discriminative representations from the raw vibration data.

Through the aforementioned comparison and discussion, there are some conclusions could be clearly obtained. Firstly, the compound fault can be decoupled into multiple-single faults by the proposed method, even without the compound fault training samples. Secondly, using the margin loss function and routing by agreement algorithm to train and optimize the proposed model, the discriminative deep features are learned and obtained for intelligent compound fault identification. Finally, compared with DCNN, the effectiveness and superior performance of the proposed method is demonstrated by the experiment.

TABLE 4. Quantitative evaluation of the features learned by the two kinds of method.

Methods	Evaluation index			
	J_1	J_2	J_3	J_4
DCNN	5.33	1.98	1.90	2.23
DDCNN	**9.79**	**2.32**	**2.58**	**2.53**

V. CONCLUSIONS

This paper proposes a novel method, DDCNN, to address the task of compound fault identification and decoupling in the field of intelligent fault diagnosis. In accordance with the demands from practical industrial applications, DDCNN aims to learn deep discriminative features and decouple the compound fault via only normal and single-fault training samples to train and optimize the model. The experimental results and the compared analyses illustrate that the proposed method can effectively identify and decouple the compound fault for rotating machinery. Both the learned deep features and the diagnosis accuracy of DDCNN are more powerful than that of traditional intelligent fault diagnosis methods. In other words, the proposed method outperforms the state-of-the-art methods in the field of intelligent fault diagnosis. In the future, decoupling the compound fault in a semi-supervised or unsupervised manner may be a great challenge. Therefore, investigating the application of a semi- or unsupervised neural network for compound fault diagnosis will be of importance, which is the direction that the authors would make further efforts.

REFERENCES

[1] W. Qiao and D. Lu, "A survey on wind turbine condition monitoring and fault diagnosis—Part I: Components and subsystems," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6536–6545, Oct. 2015.

[2] Y. Lei, J. Lin, M. J. Zuo, and Z. He, "Condition monitoring and fault diagnosis of planetary gearboxes: A review," *Measurement*, vol. 48, pp. 292–305, Feb. 2014.

[3] C. Li, J. L. V. de Oliveira, M. C. Lozada, D. Cabrera, V. Sanchez, and G. Zurita, "A systematic review of fuzzy formalisms for bearing fault diagnosis," *IEEE Trans. Fuzzy Syst.*, to be published, doi: 10.1109/TFUZZ.2018.2878200.

[4] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, Aug. 2018.

[5] M. Ma, C. Sun, and X. Chen, "Discriminative deep belief networks with ant colony optimization for health status assessment of machine," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 12, pp. 3115–3125, Dec. 2017.

[6] C. Sun, P. Wang, R. Yan, R. X. Gao, and X. Chen, "Machine health monitoring based on locally linear embedding with kernel sparse representation for neighborhood optimization," *Mech. Syst. Signal Process.*, vol. 114, pp. 25–34, Jan. 2019.

[7] D. Wang and K.-L. Tsui, "Statistical modeling of bearing degradation signals," *IEEE Trans. Rel.*, vol. 66, no. 4, pp. 1331–1344, Dec. 2017.

[8] W. Li, S. Zhang, and G. He, "Semisupervised distance-preserving self-organizing map for machine-defect detection and classification," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 5, pp. 869–879, May 2013.
[9] L. Song, H. Wang, and P. Chen, "Vibration-based intelligent fault diagnosis for roller bearings in low-speed rotating machinery," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 8, pp. 1887–1899, Aug. 2018.
[10] Z. Huo, Y. Zhang, P. Francq, L. Shu, and J. Huang, "Incipient fault diagnosis of roller bearing using optimized wavelet transform based multi-speed vibration signatures," *IEEE Access*, vol. 5, pp. 19442–19456, 2017.
[11] C. Li, Y. Tao, W. Ao, S. Yang, and Y. Bai, "Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition," *Energy*, vol. 165, pp. 1220–1227, Dec. 2018.
[12] X. Yu, F. Dong, E. J. Ding, S. P. Wu, and C. Y. Fan, "Rolling bearing fault diagnosis using modified LFDA and EMD with sensitive feature selection," *IEEE Access*, vol. 6, pp. 3715–3730, 2018.
[13] H. Darong, K. Lanyan, M. Bo, Z. Ling, and S. Guoxi, "A new incipient fault diagnosis method combining improved RLS and LMD algorithm for rolling bearings with strong background noise," *IEEE Access*, vol. 6, pp. 26001–26010, Jun. 2018.
[14] S. Singh and D. N. Vishwakarma, "Intelligent techniques for fault diagnosis in transmission lines—An overview," in *Proc. Int. Conf. Recent Develop. Control, Automat. Power Eng. (RDCAPE)*, Noida, India, Mar. 2015, pp. 280–285.
[15] L. Song, H. Wang, and P. Chen, "Step-by-step fuzzy diagnosis method for equipment based on symptom extraction and trivalent logic fuzzy diagnosis theory," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 6, pp. 3467–3478, Dec. 2018, doi: 10.1109/TFUZZ.2018.2833820.
[16] Y. Li, X. Liang, Y. Yang, M. Xu, and W. Huang, "Early fault diagnosis of rotating machinery by combining differential rational spline-based LMD and K–L divergence," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 11, pp. 3077–3090, Nov. 2017.
[17] Y. Xu, Y. Sun, J. Wan, X. Liu, and Z. Song, "Industrial big data for fault diagnosis: Taxonomy, review, and applications," *IEEE Access*, vol. 5, pp. 17368–17380, Sep. 2017.
[18] C. Li, W. Zhang, G. Peng, and S. Liu, "Bearing fault diagnosis using fully-connected winner-take-all autoencoder," *IEEE Access*, vol. 6, pp. 6103–6115, Mar. 2018.
[19] C. Sun, M. Ma, Z. Zhao, and X. Chen, "Sparse deep stacking network for fault diagnosis of motor," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3261–3270, Jul. 2018.
[20] L. Cui, J. Huang, and F. Zhang, "Quantitative and localization diagnosis of a defective ball bearing based on vertical-horizontal synchronization signal analysis," *IEEE Trans. Ind. Electron.*, vol. 64, no. 11, pp. 8695–8706, Nov. 2017.
[21] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, May 2016.
[22] 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.
[23] Y. Liao, L. Zhang, and W. Li, "Regrouping particle swarm optimization based variable neural network for gearbox fault diagnosis," *J. Intell. Fuzzy Syst.*, vol. 34, no. 6, pp. 3671–3680, Jun. 2018.
[24] S. A. Dudani, "The distance-weighted k-nearest-neighbor rule," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, no. 4, pp. 325–327, Apr. 1976.
[25] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2560–2574, 2007.
[26] 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.
[27] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer*, vol. 29, no. 3, pp. 31–44, Mar. 1996.
[28] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
[29] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[30] M. F. Yaqub, I. Gondal, and J. Kamruzzaman, "Inchoate fault detection framework: Adaptive selection of wavelet nodes and cumulant orders," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 3, pp. 685–695, Mar. 2012.
[31] R. Liu, B. Yang, X. Zhang, S. Wang, and X. Chen, "Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis," *Mech. Syst. Signal Process.*, vol. 75, pp. 345–370, Jun. 2016.
[32] C. Shen, D. Wang, F. Kong, and W. T. Peter, "Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier," *Measurement*, vol. 46, no. 4, pp. 1551–1564, May 2013.
[33] X. Zhu, Y. Zhang, and Y. Zhu, "Intelligent fault diagnosis of rolling bearing based on kernel neighborhood rough sets and statistical features," *J. Mech. Sci. Technol.*, vol. 26, no. 9, pp. 2649–2657, Sep. 2013.
[34] W. Li, S. Zhang, and S. Rakheja, "Feature denoising and nearest-farthest distance preserving projection for machine fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 12, no. 1, pp. 393–404, Feb. 2016.
[35] 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.
[36] J. Sun, C. Yan, and J. Wen, "Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 1, pp. 185–195, Jan. 2018.
[37] M. Ma, C. Sun, and X. Chen, "Deep coupling autoencoder for fault diagnosis with multimodal sensory data," *IEEE Trans. Ind. Informat.*, vol. 14, no. 3, pp. 1137–1145, Mar. 2018.
[38] L. S. Dhamande and M. B. Chaudhari, "Detection of combined gear-bearing fault in single stage spur gear box using artificial neural network," *Procedia Eng.*, vol. 144, pp. 759–766, Jan. 2016.
[39] D. Wang and K.-L. Tsui, "Brownian motion with adaptive drift for remaining useful life prediction: Revisited," *Mech. Syst. Signal Process.*, vol. 99, pp. 691–701, Jan. 2018.
[40] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, Feb. 2017.
[41] S. Sabour, N. Frosst, and G. E. Hinton. (2017). "Dynamic routing between capsules." [Online]. Available: https://arxiv.org/abs/1710.09829
[42] H. Schulza, K. Chob, T. Raikob, and S. Behnkea, "Two-layer contractive encodings for learning stable nonlinear features," *Neural Netw.*, vol. 64, pp. 4–11, Apr. 2015.
[43] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2012, pp. 1097–1105.
[44] R. Gao, F. Yang, W. Yang, and Q. Liao, "Margin loss: Making faces more separable," *IEEE Signal Process. Lett.*, vol. 25, no. 2, pp. 308–312, Feb. 2018.
[45] D. Kingma and J. Ba. (2014). "Adam: A method for stochastic optimization." [Online]. Available: https://arxiv.org/abs/1412.6980
[46] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[47] S. Haidong, J. Hongkai, L. Xingqiu, and W. Shuaipeng, "Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine," *Knowl.-Based Syst.*, vol. 140, pp. 1–14, Jan. 2018.
[48] S. Zhang and W. Li, "Variable nearest neighbor locally linear embedding and applications in bearing condition recognition," *Chin. J. Mech. Eng.*, vol. 49, pp. 81–87, Jan. 2013.

RUYI HUANG was born in Yueyang, China in 1992. He received the B.S. degree in mechanical engineering from Qingdao University, Qingdao, China, in 2014. He is currently pursuing the Ph.D. degree in mechanical engineering with the South China University of Technology, Guangzhou, China.

His current research interests include deep learning methods for intelligent fault diagnostics and prognostics of rotating machinery.

YIXIAO LIAO was born in Shanwei, China in 1992. He received the B.S. degree in automotive engineering from the South China University of Technology, Guangzhou, China, in 2015, where he is currently pursuing the Ph.D. degree in mechanical engineering.

His current research interests include deep learning and transfer learning methods for machinery fault diagnosis and prognostics.

SHAOHUI ZHANG received the Ph.D. degree in mechatronic engineering from the South China University of Technology, Guangzhou, China, in 2014.

He is currently a Teacher with the College of Mechatronic Engineering, Dongguan University of Technology, Dongguan, China. His research interests include mechanical vibration, signal processing, mechanical equipment condition monitoring, and fault diagnosis.

WEIHUA LI (M'12–SM'18) received the Ph.D. degree in mechanical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2003.

He is currently a Professor with the School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou, China. His current research interests include nonlinear time series analysis, dynamic signal processing, and machine learning methods for condition monitoring and health diagnosis of complex dynamical systems.