IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 31, 2020, accepted August 12, 2020, date of publication August 14, 2020, date of current version August 26, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3016888

Rolling Bearing Fault Diagnosis Using Time-Frequency Analysis and Deep Transfer Convolutional Neural Network

ZHIHAO CHEN, JIAN CEN, AND JIANBIN XIONG, (Member, IEEE)
[1]School of Automation, Guangdong Polytechnic Normal University, Guangzhou 510665, China
[2]Guangzhou Key Laboratory of Intelligent Building Equipment Information Integration and Control, Guangzhou 510665, China

Corresponding author: Jian Cen (mmcjian@163.com)

This work was supported in part by the Natural Science Foundation of Guangdong Province under Grant 2019A1515010700, in part by the Guangdong Special Project in Key Field of Artificial Intelligence for Ordinary University under Grant 2019KZDZX1004, in part by the Guangzhou Key Laboratory Project of Intelligent Building Equipment Information Integration and Control under Grant 202002010003, in part by the Key Project of Ordinary University of Guangdong Province under Grant 2019KZDXM020, in part by the Guangzhou People's Livelihood Science and Technology Project under Grant 201903010059, and in part by the Guangzhou Yuexiu District Science and Technology Plan Major Project under Grant 2019-GX-010.

ABSTRACT Due to the advantage of automatically extracting features from raw data, deep learning (DL) has been increasingly favored in the field of machine fault diagnosis. However, DL exposes the problems of large sample size and long training time, and in actual working conditions, the amount of labeled fault data available is relatively small, so a DL model of good generalization and high accuracy is difficult to be trained. In order to solve these problems, a deep transfer convolutional neural network (DTCNN) is proposed in this research. ResNet-50 is selected as the pre-trained model of deep convolutional neural network, and is transferred to solve the problem of bearing fault classification based on the idea of transfer learning. Firstly, raw fault signals are converted into time-frequency images by using continuous wavelet transform (CWT). Then, the images are further converted into RGB formats, which are used as the input of DTCNN. Finally, an end-to-end fault diagnosis model based on DTCNN is designed. The proposed method is validated on two datasets collected from motor bearing and self-priming centrifugal pump, respectively. Most sub-datasets from motor bearing show the prediction accuracies near 100%, and in the self-priming centrifugal pump dataset, we achieve improvement in accuracy from 99.48%±0.1966 to 99.98%±0.0332. The experimental results demonstrate that the proposed method outperforms other DL methods and traditional machine-learning methods.

INDEX TERMS Fault diagnosis, rolling bearing, convolutional neural network, continuous wavelet transform, transfer learning.

I. INTRODUCTION

The field of machine fault diagnosis has always been the focus of attention [1]. With the rapid development of industrialization, ensuring the normal and orderly operation of machines, especially large rotating machines, plays a crucial role in industrial production and life. When the machine breaks down, if the fault type and fault location can be quickly and accurately diagnosed and the fault can be solved, property losses will be greatly reduced and safety accidents will be avoided [2], [3]. Rolling bearings are the very key components of many large rotating machines, playing the role of supporting the rotating shaft and components on the shaft. Rolling bearings are easy to wear out and these wears are not easy to be noticed. Once they are damaged, the whole machines will stop working [4]. Therefore, it is very important to find a more effective and intelligent fault diagnosis method for rolling bearing.

Since the age of big data, with the mining and application of massive data, traditional fault diagnosis methods are no

The associate editor coordinating the review of this manuscript and approving it for publication was Shiping Wen.

longer applicable. At the same time, data-driven fault diagnosis methods emerge and develop rapidly, becoming a research hotspot [5]. Machine-learning-based fault diagnosis is one of the typical data-driven methods, and its common algorithms include artificial neural network (ANN) [6], ensemble empirical mode decomposition (EEMD) [7], support vector machine(SVM) [8], extreme learning machine (ELM) [9], etc. Jiang *et al.* [10] used SVM and multi-sensor information fusion to diagnose rolling bearing fault and gear fault. Qin *et al.* [11] used EEMD to decompose raw vibration signals and used random forest to classify bearing fault features. Machine-learning-based fault diagnosis methods are almost manual to extract fault features. However, the vibration signals taken from industrial site are often non-stationary, while the manually extracted fault features largely depend on expert experience and prior knowledge [12], which brings difficulties and errors to the feature extraction. Moreover, machine-learning models tend to learn only one or two layers of data representations and fail to learn enough abundant fault information, which limits the ultimate diagnosis accuracy. Therefore, the trained models have poor performances, which can no longer meet the requirements of modern fault diagnosis in terms of rapidity and high accuracy.

In recent years, the emergence and development of deep learning (DL) has solved the above problems. DL is a new branch of machine learning, which can learn features automatically from raw data without any expert experience [12]. Different from the shallow representations of machine learning, DL tends to have dozens or even hundreds of successive layers of representations, which is a kind of hierarchical representations learning [13]. These successive layered representations are often learned through neural networks. DL has been widely applied to the areas of computer vision, video generation, speech recognition, etc. Wen *et al.* [14] designed two concatenated generative adversarial networks (GANs) to generate realistic and sharp videos. Wen *et al.* [15] proposed an end-to-end detection-segmentation fully convolutional network (FCN) to get the state-of-the-art results of detailed face labeling in HEELEN face dataset. Ren *et al.* [16] presented a new learning rate scheme for Elman neural network (ENN) to improve the convergence speed. Wen *et al.* [17] combined direct label recognition using a ResNet model and a feature label co-projection module to solve the problem of multilabel image classification.

Recently, some DL models including deep belief network (DBN) [18], convolutional neural network (CNN) [19], and stacked auto-encoder (SAE) [20] have been successfully applied in the area of fault diagnosis. Shao *et al.* [21] extracted fault features with double tree complex wavelet packet, and designed an adaptive DBN for bearing fault diagnosis, which achieved good results. Qi *et al.* [3] proposed a fault diagnosis method for rotating machinery based on stacked sparse auto-encoder (SSAE) and obtained a better performance compared with traditional machine-learning methods. CNN has great application prospects in bearing fault diagnosis due to its excellent performance in image classification. Wen *et al.* [22] converted vibration signals into grayscale images and used them in the improved CNN model, which obtained a high classification accuracy. However, DL-based fault diagnosis methods still have the following problems:

1) **Acquisition of fault samples:** In actual working conditions, the number of labeled fault samples can be acquired is relatively small, and labeling a large number of samples is a time-consuming and laborious work [23].

2) **Depth of DL model:** The performances of DL models usually depend on their depth. The model is deeper, and its performance is better [24], [25]. The benchmark CNN models, trained on ImageNet dataset of more than 1.4 million labeled images, can reach depths of 50 or even 152 layers [26], [27]. Due to the small-scale of fault datasets, the CNN model used for bearing fault diagnosis can only stack up to 5 hidden layers [28]. If the model has too many layers, it will be easy to overfit. The above problem limits the depth of the CNN model and the ultimate accuracy of fault diagnosis.

3) **Time cost of model training:** It takes a lot of time to randomly initialize weights from scratch to train a neural network [29]. Moreover, in order to obtain a deeper model, it is necessary to find a larger labeled fault dataset and input it into the neural network for training. As the number of layers increases, the number of parameters to be trained also increases greatly, which is also very laborious and time-consuming.

The emergence of transfer learning solves the above problems well. Unlike DL that directly acquires knowledge from data from scratch, transfer learning aims to transfer what has already been learned and apply it to solve a different but similar problem. In the field of fault diagnosis, there are few studies on transfer learning, which are still in the exploratory stage. Zhang *et al.* [30] studied the neural network model based on transfer learning for small data under different working conditions, which was well applied to bearing fault diagnosis. Shao *et al.* [31] realized the transfer learning of three fault datasets by fine-tuning the VGG-16 model and achieved excellent diagnosis results. By using the deep CNN model trained on a large-scale dataset in advance, and combining with transfer learning technology, the problem that small datasets cannot be trained in deep networks can be solved [32]. Inspired by this, this paper designs a deep transfer convolution neural network (DTCNN). DTCNN uses Resnet-50, one of the benchmark CNN models, as the pre-trained model. ResNet-50 model was previously trained on ImageNet dataset and further transferred to two bearing datasets for fault classification. The results show that the proposed DTCNN model can achieve near 100% prediction accuracy, indicating that the proposed model has a good performance for rolling bearing fault diagnosis.

The rest of this paper is organized as follows. Section II introduces the principles of time-frequency analysis, CNN and transfer learning. In Section III, the proposed method

is presented, including constructing time-frequency images by CWT and using DTCNN for bearing fault diagnosis. Section IV carries out the experiment verification and result analysis. Section V presents the conclusion and future work.

II. THE RELATED THEORY

A. TIME-FREQUENCY ANALYSIS

Time-frequency analysis provides the joint distribution information of time domain and frequency domain of signal, which is a powerful tool for processing non-stationary signals in fault diagnosis [33]. Common time-frequency analysis methods include short-time Fourier transform (STFT) and continuous wavelet transform (CWT) [34], [35]. CWT is an adaptive time-frequency analysis method. Compared with STFT, CWT can give a good balance between frequency resolution and time resolution, which has certain advantage. CWT refers to the inner product operation between signal $x(t)$ and a set of continuous basic wavelet functions $\psi_{a,b}(t)$, and then projected onto the two-dimensional (2D) time-scale phase plane [36]. The basic wavelet function $\psi_{a,b}(t)$ is obtained by the translation and telescopic transformation of the mother wavelet function $\psi(t)$, and its formula is:

$$\psi_{a,b}(t) = a^{\frac{1}{2}}\psi(\frac{t-b}{a}), \quad a > 0,\ b \in R \tag{1}$$

where b is the translation factor and a is the scale factor. Since a and b take the values of continuous transformations, y is a continuous basic wavelet function of the parameters a and b. The mother wavelet function $\psi(t)$ refers to a square integrable function whose Fourier transform $\psi(\omega)$satisfies the following conditions:

$$\int_R \frac{|\psi(\omega)|^2}{\omega} d\omega < \infty \tag{2}$$

In signal processing, common mother wavelet functions include Morlet wavelet, Haar wavelet, Gabor wavelet, etc. We carry out CWT operation on signal $x(t)$, and its transformation formula is:

$$\langle x(t), \psi_{a,b}(t)\rangle = \int_R x(t)\overline{\psi_{a,b}(t)}dt = \frac{1}{\sqrt{a}}\int_R x(t)\overline{\psi(\frac{t-b}{a})}dt \tag{3}$$

where $\langle\cdot\rangle$ denotes the inner product, and $\overline{\psi(\cdot)}$ is the complex conjugate of $\psi(\cdot)$. According to the above formula, CWT is carried out on signals to realize the transformation of 1D time series into 2D time-frequency images, which can be conducive to the extraction of features.

B. CNN AND ITS TYPICAL ARCHITECTURE

CNN is one of the most typical models of DL, which has great potential for fault diagnosis. The typical architectures of CNN include Lenet-5, AlexNet, VGGNet, GoogLeNet, and ResNet [24]–[26], [37]. This section focuses on the basic structure of CNN and one of its typical architectures, the ResNet-50.

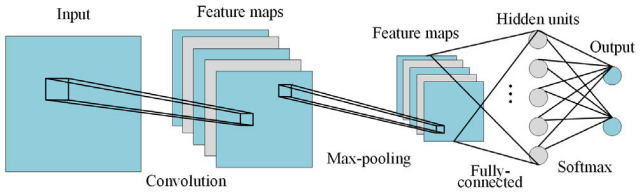

FIGURE 1. The basic structure of CNN.

1) CNN

CNN is widely used in image classification tasks. The main structure of CNN is input, convolutional layer, pooling layer, fully-connected layer, and output [14], as shown in Fig. 1. The input of 2D CNN are image tensors, which are 3D tensors of shape (height, width, channels), not including the batch dimension. The channel of grayscale image is 1, and the channel of RGB image is 3. The characteristics of CNN are local receptive field, weight sharing, and pooling. There are multiple convolution kernels inside the convolutional layers. Convolution kernels scan the input features in the local receptive field, and their neurons share the weight and bias of the convolutional layer. Each convolution kernel learns a kind of feature. The convolution operation is generally expressed as:

$$x^i = f(\sum \omega^i * x^{i-1} + b^i) \tag{4}$$

where x^{i-1} and x^i respectively denote the input and output feature map of i-th layer. ω^i and b^i respectively denote the convolution kernel and bias of i-th layer. $f(\cdot)$ represents the activation function, which provides nonlinearity ability to the convolutional layer. The activation functions usually include ReLU, sigmoid, and tangent. Among them, ReLU function can improve the nonlinearity ability and convergence speed of CNN [24], [38]. It is adopted in this paper because of its outstanding performance when applied in the proposed DTCNN. The ReLU function is defined as:

$$x^j = max(0, x^{j'}) \tag{5}$$

After the convolution layer is generally connected to the pooling layer, the pooling operation reduces the size of the output feature map, avoids the dimension disaster, and well preserves the features described by the feature map. Pooling operation is generally described as:

$$x^{m',n'} = pool(x^{m,n}) \tag{6}$$

where $x^{(m,n)}$ and $x^{(m',n')}$ represent the values of feature map before and after the pooling operation at the point (m, n), and $pool(\cdot)$ represents the pooling function. Pooling operations including average pooling and maximum pooling. The convolutional layer and the pooling layer constitute the convolution block, and the stacking of the convolutional block constitutes the deep CNN, which is conducive to learning more abundant features.

The fully-connected layer is usually used to further extract the high-level and abstract features, in which all the neuron nodes are connected to all the neuron nodes in the output

feature map of the previous layer. For the multi-classification task, the output layer is usually connected to the Softmax classifier, which takes the input from the previous linear layer and outputs the probability on a given number of samples.

For image classification tasks, we adopt the cross entropy loss function to calculate the loss value of CNN, which is defined as:

$$H(y, a) = -\sum_{i} y_i \log(a_i) \tag{7}$$

where y_i denotes the true label of class i, and a_i is the output probability of class i.

2) RESNET-50

The ResNet-50 is one of the most advanced network architectures of CNN. By introducing residual learning, ResNet-50 can effectively avoid the gradient disappearance and degradation caused by the desire to deepen the network [26]. ResNet-50 no longer simply learns the potential mapping between input x and output $H(x)$ directly, but through the residual $F(x) = H(x) - x$ between the two, adds the residual to the input to learn $F(x) + x$. The basic structure of the residual block, as shown in Fig. 2, bypasses the output of one of the previous hidden layers to connect directly to the input of the later hidden layer.

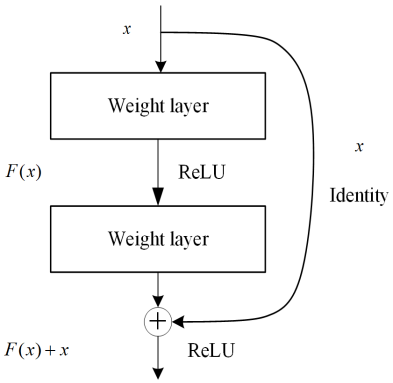

FIGURE 2. The basic structure of residual block.

Resnet-50 has two basic blocks, named Conv Block and Identity Block, as shown in Fig. 3, where Conv Block changes the dimension of the network and Identity Block makes the network deeper. The simplified structure of the ResNet-50 model is shown in Fig. 4. It consists of several Conv Blocks and Identity Blocks stacked, with a total of 49 convolutional layers and 1 Softmax layer. ResNet-50 has excellent performance to deal with image classification tasks [26].

C. TRANSFER LEARNING AND FINE-TUNING

Transfer learning refers to the transfer of knowledge from the source domain to the target domain by virtue of the similarity between the source domain and the target domain [23]. The source domain refers to the domain with knowledge and lots of annotation. The target domain is the object to be endowed with knowledge.

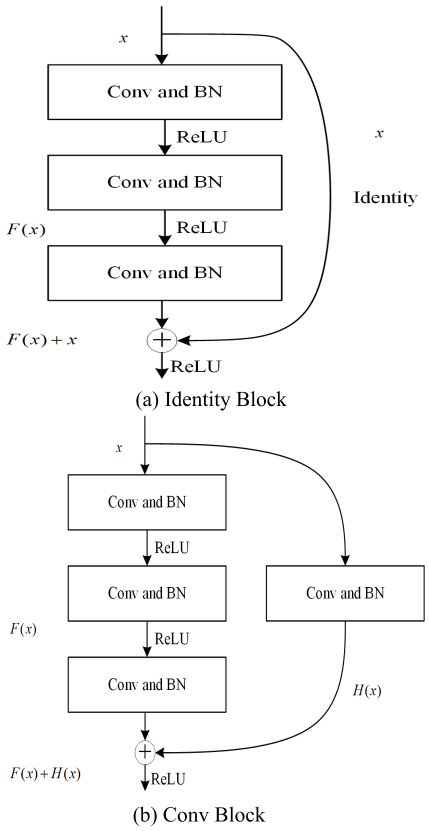

FIGURE 3. Residual block of ResNet-50.

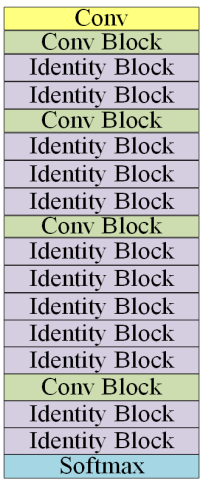

FIGURE 4. The basic structure of ResNet-50.

In DL, the deep CNN model is often selected to transfer knowledge, or to share the parameters between models, which is a learning method combining feature transfer and model transfer. The source domain is usually a large annotated dataset, such as ImageNet, which consists of more than 1.4 million annotated natural images. We call the deep CNN model that has been trained in the source domain the pre-trained model. In the bottom layers of deep CNN, only some lines, textures, and contours are learned, which are general, while the higher layers closer to the output layer

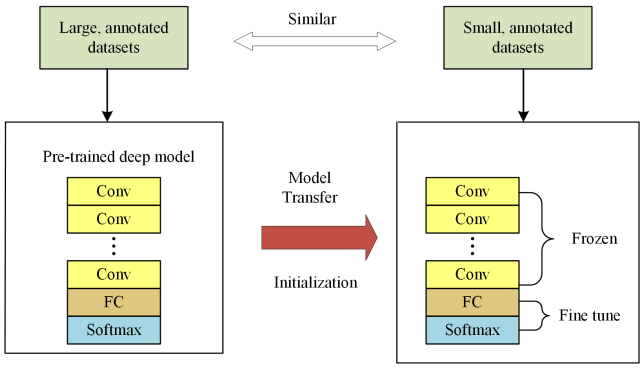

FIGURE 5. Transfer learning by using fine-tuning.

are more abstract, and the more advanced the features are learned [32]. With this character, we can share the parameters of bottom layers between models. It is assumed that the target domain and the source domain satisfy a certain similarity, and then the parameters of bottom layers of the pre-trained model are shared with the target domain model, and only the parameters of higher layers of the target domain model are trained. This transferring, which is called fine-tuning, can be realized by freezing the weights of bottom layers and only updating the weights of higher layers. The process of fine-tuning is shown in Fig. 5. It can be said that fine-tuning is the simplest form of deep CNN transfer. For those with high similarity between the source domain and the target domain, only the fully-connected layers need to be fine-tuned, while for those with low similarity, several more convolution blocks adjacent to the fully-connected layers need to be fine-tuned. It is worth noting that for different classification tasks in the source domain and target domain, the number of neurons in the fully-connected layer and output layer of the model should be changed accordingly. Fine-tuning uses the pre-trained network to adjust according to its own task, which can effectively overcome the differences between two datasets.

III. THE PROPOSED METHOD

A. CONSTRUCTION OF TIME-FREQUENCY IMAGES

2D CNN is the most commonly used neural network in image processing, and its input is usually grayscale images or RGB images. Therefore, it is necessary to convert the raw fault signals into 2D time-frequency images abundant in fault information by CWT. In this paper, cmor3-3 wavelet is used for CWT, which is a complex Morlet wavelet with excellent time-frequency analysis ability. For the raw vibration signals collected by sensors, every 1024 signal sampling points are taken as a time series, and CWT is carried out for this time series, as shown in Fig. 6.

B. FAULT DIAGNOSIS BASED ON DTCNN

To solve the problem that the annotated fault datasets are not large enough, a DTCNN is presented for fault diagnosis in this paper. DTCNN refers to the deep CNN model that has been trained in ImageNet dataset for transfer learning, such as ResNet-50, VGG-16, Inception v3, etc. In this paper,

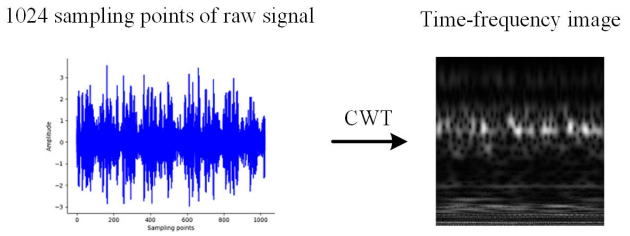

FIGURE 6. Construction of time-frequency image by using CWT.

the sufficiently strong generalization ability of ResNet-50 is selected as the DTCNN model. We assumed in advance that the fault time-frequency images could realize transfer learning on DTCNN(ResNet-50). Due to certain differences between the fault time-frequency images and the natural images on ImageNet, fine-tuning of DTCNN(ResNet-50) is needed to further overcome the differences.

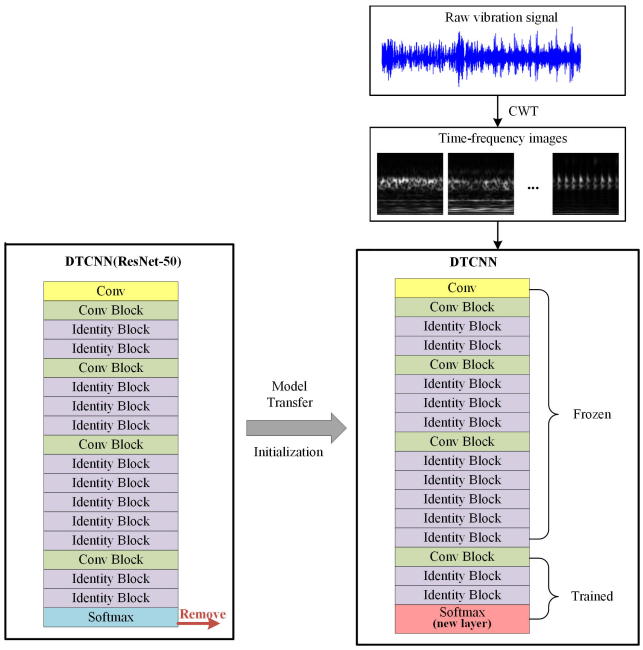

FIGURE 7. The flowchart of fault diagnosis based on the proposed DTCNN and time-frequency analysis.

The flowchart of rolling bearing fault diagnosis based on the proposed DTCNN and time-frequency analysis is shown in Fig. 7. The diagnosis processes are divided into data preprocessing, building and fine-tuning of DTCNN model, training of DTCNN, and application of DTCNN model.

1) DATA PREPROCESSING

Raw fault signals are converted into time-frequency images by CWT. Since the input of DTCNN are RGB images in $224 \times 224 \times 3$ formats, and the time-frequency images are grayscale images with 1 channel, it is necessary to copy the time-frequency images into 3 channels to form R=G=B false-colored images. Then the time-frequency images are

compressed into 224 × 224 images and normalized. Finally, the dataset is divided into training dataset and testing dataset.

2) BUILDING AND FINE-TUNING OF DTCNN MODEL

Since the task of the trained ResNet-50 is to classify 1,000 classes of natural images, it is necessary to remove the Softmax layer of DTCNN(ResNet-50) and add a new one, the number of neurons in the new Softmax layer depends on the number of fault classes. Due to some differences between the fault time-frequency images and the natural images, the last three residual blocks(1 Conv Block and 2 Identity Blocks) near the output layer need to be fine-tuned to further extract the advanced fault features.

3) TRAINING OF DTCNN

We only update the parameters of the last three residual blocks and the output layer of the model, while the remaining layers still use the parameters trained in ImageNet.

4) APPLICATION OF DTCNN MODEL

The trained model is applied to rolling bearing fault diagnosis.

IV. EXPERIMENTAL VERIFICATION

In order to evaluate the effectiveness of this proposed method, the proposed DTCNN diagnosis model is tested on the motor bearing dataset and the self-priming centrifugal pump dataset. In addition, the proposed method is compared with some existing methods, including traditional machine-learning methods and other DL methods. The proposed method is implemented by Python 3.6 and Keras with Tensorflow as the backend, in which the trained ResNet-50 model can be imported from *keras.applications*, and all the experiments are run on Window 10, using NVIDIA GTX960 GPU.

A. CASE 1: MOTOR BEARING DATASET

1) DATASET AND EXPERIMENTAL SETUP

In this case, the proposed DTCNN model is conducted on the standard fault dataset of motor rolling bearing published by Case Western Reserve University (CWRU) [39]. The bearing experimental platform of CWRU is shown in Fig. 8. The acceleration sensor is used to collect bearing vibration signals of the motor's drive end and fan end respectively. The former has sampling frequencies of 12kHz and 48kHz, while the latter has a sampling frequency of 12kHz. This paper only studies the drive end bearing with a sampling frequency of 12kHz. The vibration data of the bearing was measured under four different motor load conditions of 0, 1, 2, and 3 hp. Rolling bearing faults can be divided into ball fault (BF), inner race fault (IF), and outer race fault (OF). Their faults are all single point faults caused by electro-discharge machine (EDM), and there are three different fault diameters of 7 mils, 14 mils, and 21 mils for each fault type (1 mil = 0.001 inches). In particular, there are three damage points in the outer race: 3 o 'clock, 6 o 'clock, and 12 o 'clock, and only 6 o 'clock is selected in this paper. In this way, the bearings studied in this paper have a total of 10 health conditions, including 9 fault conditions(BF-07, BF-14, BF-21, IF-07, IF-14, IF-21, OF-07, OF-14, OF-21) and 1 normal condition (NO).

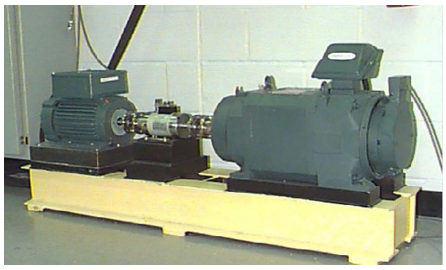

FIGURE 8. The bearing experiment platform of CWRU in Case 1.

The bearing vibration signals under the condition of 0 hp load are selected to construct sub-dataset A. For each health condition, 400 samples are randomly selected to construct the training dataset and 200 samples to construct the testing dataset, each sample containing 1024 signal sampling points. There are ten health conditions, so the training dataset has a total of 4000 training samples, and the testing dataset has a total of 2000 testing samples. In addition, 25% of the training dataset is used as the validation dataset randomly, so the training dataset is divided into 3000 samples for training and 1000 samples for validation.

CWT is used to convert the raw data into time-frequency images, and then the images are compressed to the size of 224 × 224, and the single-channel time-frequency images are copied into three channels to construct RGB images for the input of DTCNN(ResNet-50). The faults of motor bearing are classified as 10, so we need to remove the Softmax layer of DTCNN(ResNet-50) and add a new one with 10 neurons. The last three residual blocks of DTCNN together with the new Softmax layer are set to be trainable and the remaining layers are frozen. The remaining layers use weights trained in ImageNet, while the last three residual blocks and Softmax layer need to be initialized randomly. When training DTCNN, Adam optimizer is used, and the learning rate is set to 0.0001, batch size is 32, and epoch is 5.

2) RESULTS AND DISCUSSION

The CWT results of raw vibration signals are shown in Fig. 9 (load = 0 hp). As can be seen from Fig. 9, it is difficult to distinguish the corresponding fault types according to the raw time domain signals, but after CWT, the differences between the time-frequency images of each fault type can be distinguished, which is suitable for further input of these time-frequency images into CNN for feature extraction.

The time-frequency images are input into DTCNN for training. After 5 epochs, the training accuracy curve and loss curve are shown in Fig. 10. As can be seen from Fig. 10, after only 2 epochs, accuracy and loss values of the training and validation datasets have been very stable, and the model has started to converge, indicating the strong convergence ability of the model.

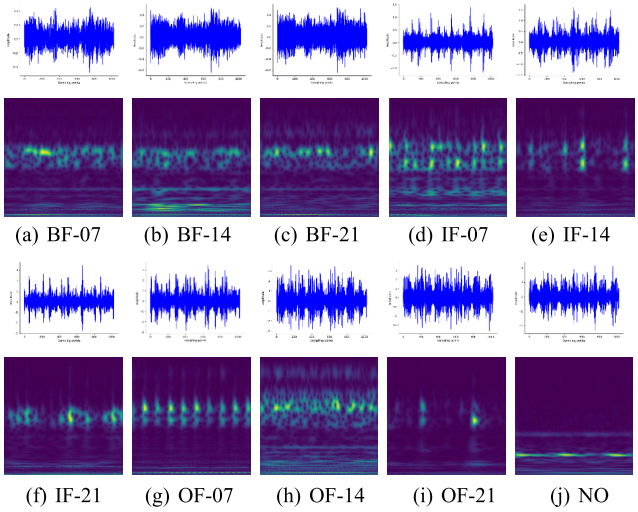

FIGURE 9. The results of CWT under 0 hp load condition.

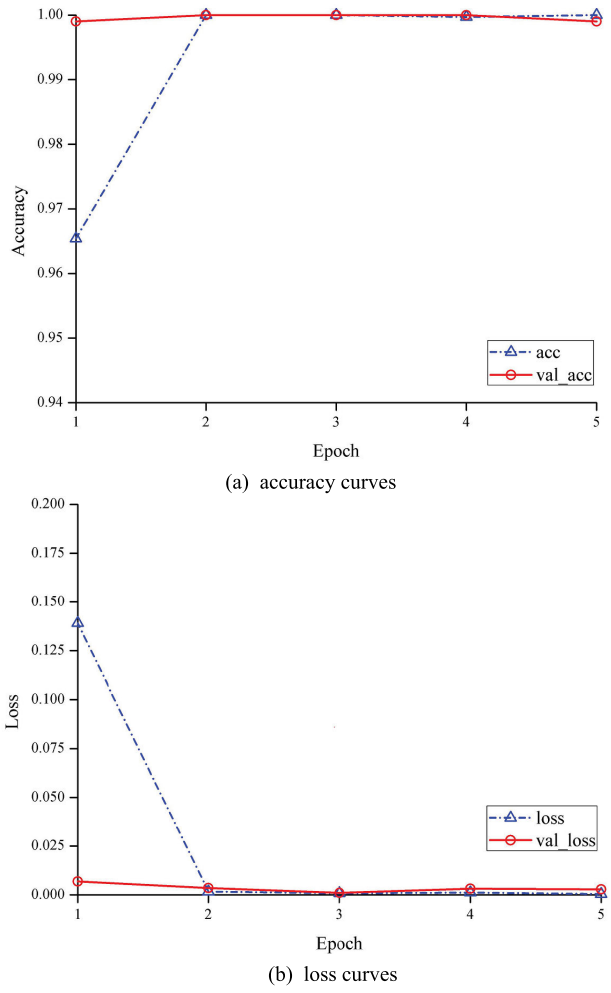

FIGURE 10. The accuracy and loss curves of the proposed DTCNN in sub-dataset A.

The trained model is used to classify the testing dataset to obtain classification accuracy, and the work is repeated for 10 times under the same conditions, and the maximum accuracy, minimum accuracy, mean accuracy, standard deviation (std), and computation time will be counted. Note that the computation time consist of training time and testing time. The results are listed in Table 1. As can be seen from the table, DTCNN model has achieved 99.97%±0.0332 accuracy in sub-dataset A, indicating that the model has a very high prediction accuracy, and the average computation time is 287 seconds, which demonstrates that the propose method is fast in diagnosis.

TABLE 1. The result of DTCNN model in sub-dataset A.

Sub-dataset	Max accuracy	Min accuracy	Mean accuracy	Std	Computation time
A	100%	99.90%	99.97%	0.0332	287s

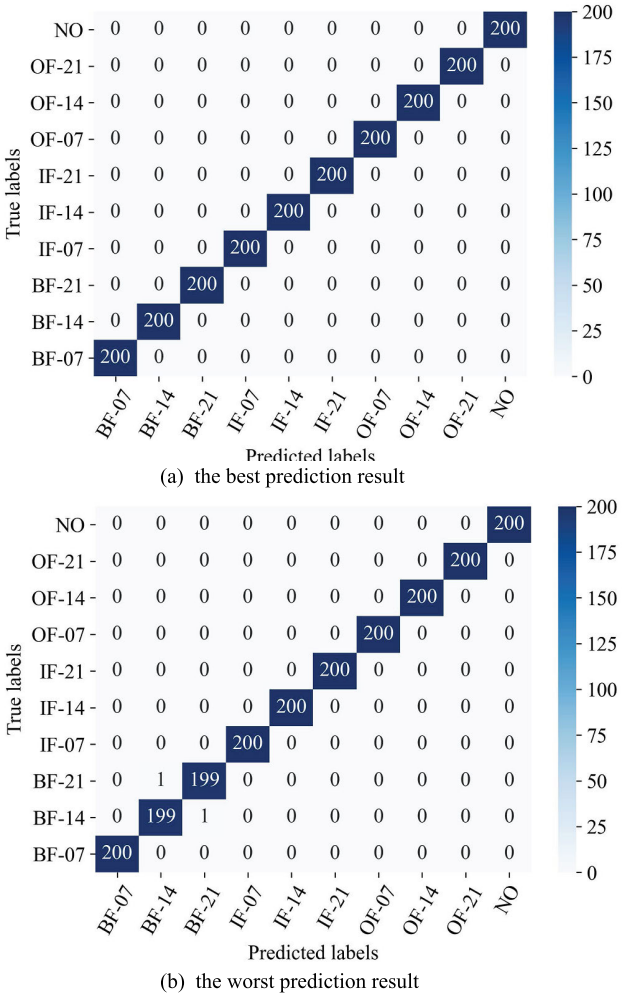

FIGURE 11. The confusion matrix of the best and worst prediction results in sub-dataset A.

Fig. 11 is the confusion matrix representations of the best and worst results, which obtain the prediction accuracies of 100% and 99.90% (the maximum and minimum values in Table 1). In Fig. 11, the columns represent the predicted

label for each health condition, and the rows represent the actual label for each health condition. It can be seen from the confusion matrix of the best prediction result that 100% of the prediction accuracy is achieved for each health condition. It can be seen from the confusion matrix of the worst prediction result that 100% prediction accuracies have been achieved in all health conditions except BF-14 and BF-21, one BF-14 is misclassified as BF-21 and one BF-21 is misclassified as BF-14. More importantly, no any faulty condition was misclassified as normal (NO).

We perform sensitivity and specificity analysis on the confusion matrix of the worst prediction result. The analysis results are listed in Table 2. From the results, except for the sensitivity of BF-14 and BF-21 of 99.50%, the sensitivity and specificity of all other health conditions can reach more than 99.90%, which demonstrates the effectiveness of this proposed method.

TABLE 2. The analysis results of the confusion matrix of the worst prediction accuracy in sub-dataset A.

Condition type	Sensitivity	Specificity	Accuracy
BF-07	100%	100%	
BF-14	99.50%	99.94%	
BF-21	99.50%	99.94%	
IF-07	100%	100%	
IF-14	100%	100%	99.90%
IF-21	100%	100%	
OF-07	100%	100%	
OF-14	100%	100%	
OF-21	100%	100%	
NO	100%	100%	

TABLE 3. The details of fine-tuned layers of various DTCNN models in Case 1.

Model	Fine-tuned layers
DTCNN1 (the proposed)	The last 3 residual blocks and Softmax
DTCNN2	The last 2 residual blocks and Softmax
DTCNN3	The last 1 residual block and Softmax
DTCNN4	Only the Softmax

3) INFLUENCES OF DIFFERENT FINE-TUNED LAYERS

In order to study the influence of the number of fine-tuned layers on the performance of the proposed DTCNN model, different DTCNN models are constructed corresponding to different fine-tuned layers, as shown in Table 3. The four models are trained on sub-dataset A, and the trained epoch is set to 10. Other settings are the same as those set above. The average time required for each DTCNN model to conduct one epoch training is calculated, and the results are represented in Fig. 12. The accuracy and loss curves of the training and validation datasets of each DTCNN model are shown in Fig. 13.

Fig. 12 shows that the time required by DTCNN1 (the proposed) model to train one epoch is 5s, 8s, and 11s longer than that required by DTCNN2, DTCNN3, and DTCNN4, respectively. It indicates that the more layers of fine-tuning, the more parameters need to be updated, and the training time is relatively increased. Fig. 13 shows that DTCNN1 converges faster than other DTCNN models, with higher accuracy and lower loss. It indicates that the more top layers of fine-tuning, the more abstract and specific features the model can learn from the dataset, and the faster the model converges and the higher the accuracy.

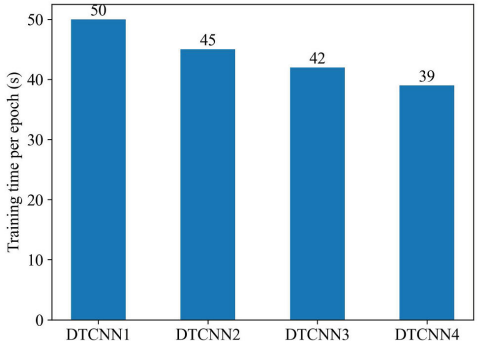

FIGURE 12. The average training time of various DTCNN models (1 epoch) in sub-dataset A.

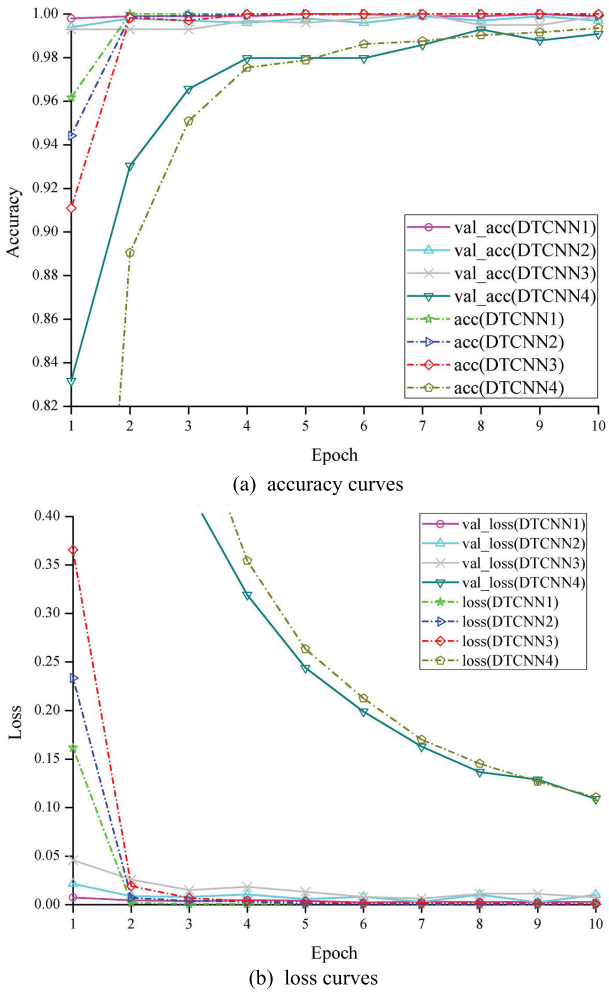

FIGURE 13. The accuracy and loss curves of various DTCNN models in sub-dataset A.

The number of fine-tuned layers is increased to overcome the difference between the source domain and the target domain within a certain range. Although DTCNN1 takes the most time to train one epoch, it converges faster, requires fewer epochs, and is more accurate. So DTCNN1, the model used in this paper, is the best model. Note that the DTCNN model proposed in this paper has nearly 100% prediction accuracy, and there is no need to fine-tune another residual block, so as to prevent too many parameters that need to be updated and risk of overfitting for small datasets.

4) GENERALIZATION VALIDATION

In order to validate the generalization of the proposed model, several sub-datasets under various working conditions are constructed as follows:

Sub-dataset B: For each health condition, 400 samples are randomly selected under 1 hp load as the training dataset, and 200 samples are selected as testing dataset in the same manner. Therefore, there are 4000 training samples and 2000 testing samples with 10 conditions.

Sub-dataset C: 4000 samples with 10 health conditions are randomly selected under 2 hp load as the training dataset, and 2000 samples are selected as testing dataset in the same manner.

Sub-dataset D: 4000 samples with 10 health conditions are randomly selected under 3 hp load as the training dataset, and 2000 samples are selected as testing dataset in the same manner.

Sub-dataset E: 100 samples are randomly selected for each health condition as the training dataset under each load condition (0, 1, 2, and 3 hp), and 50 samples are selected as testing dataset in the same manner. Therefore, there are 4000 training samples and 2000 testing samples with 10 conditions.

Sub-dataset F: 150 samples are randomly selected for each health condition as the training dataset under each load condition (0, 1, and 2 hp), while 200 samples are selected as testing dataset under 3hp load. Therefore, there are 4500 training samples and 2000 testing samples with 10 conditions.

In addition, for each sub-dataset, 25% of the training datasets above are used as the validation datasets during training. The proposed DTCNN model is used to train the sub-datasets B-F respectively. The settings of training are the same as that of the previous sub-dataset A. The model is run for 10 times under the same conditions, and the experimental results are shown in Table 4.

Results show that the proposed model has respectively achieved 99.99%±0.0200 accuracy in sub-dataset B, 100%±0 accuracy in sub-dataset C, 99.98%±0.0335 accuracy in sub-dataset D, 99.95%±0.0387 accuracy in sub-dataset E, and 99.72%±0.1163 accuracy in sub-dataset F. It shows that the proposed DTCNN model can achieve very good diagnosis accuracy for fault data under various working conditions. Especially in sub-dataset F, model is trained from 0-2 hp training dataset, and is tested from 3 hp testing dataset which is a brand-new working condition for the model. However, the model can still well identify the bearing faults, indicating that the model has good generalization and robustness. In addition, from the computation time in Table 4, the proposed method has a fast diagnosis in sub-datasets B-F.

TABLE 4. The results of DTCNN model under various working conditions for the motor bearing dataset.

Sub-dataset	Max accuracy	Min accuracy	Mean accuracy	Std	Computation time
B	100%	99.95%	99.99%	0.0200	291s
C	100%	100%	100%	0	293s
D	100%	99.90%	99.98%	0.0335	298s
E	100%	99.90%	99.95%	0.0387	294s
F	99.90%	99.50%	99.72%	0.1163	318s

5) COMPARISONS WITH OTHER DTCNN(INCEPTION V3, VGG-16) MODELS

To evaluate the performance of the DTCNN(ResNet-50) model presented in this paper, DTCNN(ResNet-50) is compared with other DTCNN(Inception v3, VGG-16) models. DTCNN(Inception v3) removes the Softmax layer and fine-tunes the last five Inception modules and the newly added Softmax layer, in which the new Softmax layer has 10 neurons. DTCNN(VGG-16) removes two fully-connected layers and one Softmax layer, and fine-tuned the newly added two fully- connected layers and one Softmax layers. The number of neurons in both fully-connected layers is 512, and the number of neurons in the Softmax layer is 10. Three different DTCNN models are trained on sub-dataset A to calculate the average time required for each DTCNN model to conduct one epoch training, and the results are shown in Fig. 14. The accuracy curves of the training and validation datasets for each DTCNN model are shown in Fig. 15.

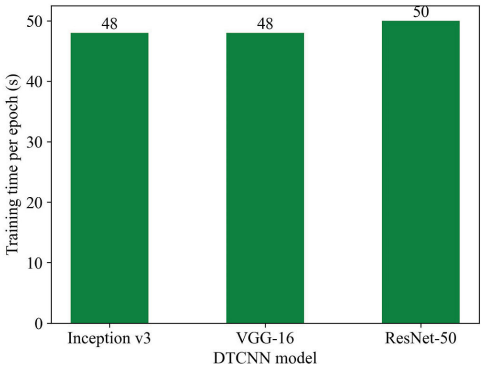

FIGURE 14. The average training time of various DTCNN models (1 epoch) in sub-dataset A.

Fig. 14 and Fig. 15 show that the time required for DTCNN(Inception v3), DTCNN(VGG-16), and DTCNN (ResNet-50) to undergo one epoch training is approximately the same, but DTCNN(ResNet-50) converges faster than DTCNN(Inception v3) and DTCNN(VGG-16). In order to further verify the superiority of DTCNN(ResNet-50),

TABLE 5. The comparison results of DTCNN models under various working condition in Case 1.

DTCNN model	Sub-dataset A		Sub-dataset B		Sub-dataset C		Sub-dataset D		Sub-dataset E		Sub-dataset F	
	Acc(%)	Time(s)	Acc(%)	Time(s)	Acc(%)	Time(s)	Acc(%)	Time(s)	Acc(%)	Time(s)	Acc(%)	Time(s)
Inception v3	99.94	**279**	99.92	284	99.96	**278**	99.97	**280**	99.79	**279**	98.45	313
VGG-16	99.90	282	99.97	**280**	**100**	279	99.96	281	99.89	283	98.31	**310**
ResNet-50	**99.97**	287	**99.99**	291	**100**	293	**99.98**	298	**99.95**	294	**99.72**	318

TABLE 6. The comparison results for the motor bearing dataset in Case 1(%).

Method	Sub-dataset under various working conditions					
	A	B	C	D	E	F
Wavelet features + SVM [40]	88.90	-	-	-	-	-
ELM [41]	97.50	-	-	-	-	-
EEMD + SVM [42]	97.64	99.12	99.64	97.56	97.91	-
DWAE + ELM [43]	95.20	-	-	-	-	-
DBN [44]	99.03	-	-	-	-	-
SSAE + DNN [45]	-	-	-	-	99.10	-
CNN [22]	-	-	-	-	99.79	-
CNN + RF [46]	-	-	-	-	99.73	99.08
Proposed method	**99.97**	**99.99**	**100**	**99.98**	**99.95**	**99.72**

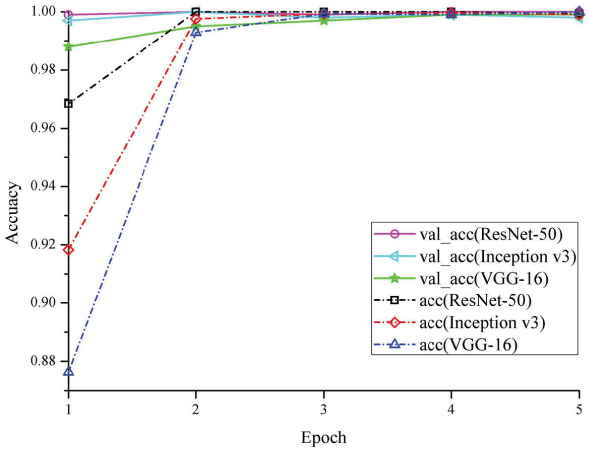

FIGURE 15. The accuracy curves of various DTCNN models in sub-dataset A.

DTCNN(Inception v3), DTCNN(VGG-16), and DTCNN (ResNet-50) are trained on sub-datasets A-F respectively, with mean accuracy (Acc) and computation time (Time) as the measurement items. The results are shown in Table 5.

Table 5 shows that DTCNN(ResNet-50) has higher diagnosis accuracy, especially in sub-dataset F, the diagnosis accuracy of DTCNN(ResNet-50) is 1.27% and 1.41% higher than that of DTCNN(Inception v3) and DTCNN(VGG-16), indicating better generalization of DTCNN(ResNet-50). From the average computation time in Table 5, the three DTCNN models are not much difference in computation time. In conclusion, DTCNN(ResNet-50) is slightly better than DTCNN(Inception v3) and DTCNN(VGG-16).

6) COMPARISONS WITH TRADITIONAL MACHINE-LEARNING METHODS AND OTHER DL METHODS

In order to further evaluate the performance of the proposed DTCNN method, the proposed method is compared with the traditional machine learning method and other DL methods. The methods are listed as follow: [40] is based on wavelet leaders multifractal features and SVM, [41] uses ELM for bearing fault diagnosis, [42] is based on EEMD and optimized SVM, [43] uses deep wavelet auto-encoder (DWAE) and ELM for fault diagnosis, [44] uses DBN to solve the problem of hierarchical identification of machine, [45] is based on SSAE and deep neural network (DNN), [22] is based on 2D images and CNN, [46] uses CNN and random forest (RF) ensemble learning for fault diagnosis. Mean accuracy is taken as the measurement item, and the results are listed in Table 6.

Table 6 shows that the method presented in this paper is superior to traditional machine-learning methods and other DL methods. Compared with classical method [40] and [41], the accuracy of proposed method is significantly improved by 11.07% and 2.47%, respectively. Compared with the machine-learning based method [42], the proposed method can automatically learn the features and achieve higher prediction accuracy. Compared to [43], the accuracy of proposed method is improved by 5.77%. Compared with other DL based methods [44] and [45], the accuracy of proposed method is improved by 0.94% and 0.85% respectively. Compared with the method based on CNN from scratch [22] and [46], the accuracy of proposed method is improved by 0.16% and 0.22% respectively. In particular, compared to [46], the accuracy of sub-dataset F is improved by 0.64%, indicating that the proposed model has better generalization.

B. CASE 2: SELF-PRIMING CENTRIFUGAL PUMP DATASET

1) DATASET AND EXPERIMENTAL SETUP

In order to verify the universality of the proposed method, the DTCNN model is also tested on the self-priming centrifugal pump dataset, which is provided from Lu *et al.* [47]. The self-priming centrifugal pump data acquisition system is shown in Fig. 16. The acceleration sensor is installed above the motor housing to collect vibration data with the sampling frequency of 10239 Hz, and the rotation speed of motor is 2900 rpm. The dataset contains 4 fault conditions and

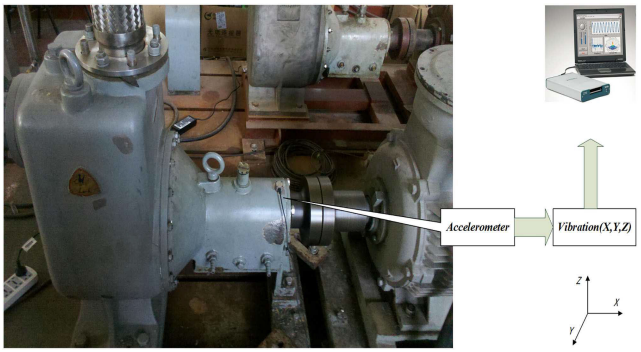

FIGURE 16. The self-priming centrifugal pump data acquisition system in Case 2.

1 normal condition (NO). The fault conditions consist of bearing inner race wearing (IR), bearing outer race wearing (OR), bearing roller wearing (RW), and impeller wearing (IW). In this case, for each condition, 1000 samples with 1024 data points are randomly selected as the training dataset, and 400 samples are selected as testing dataset in the same way. Therefore, there are 5000 training samples and 2000 testing samples with 5 conditions. In addition, 25% of the training dataset is used as the validation dataset during training, so the training dataset is divided into 3750 samples for training and 1250 samples for validation. Fault diagnosis for the self-priming centrifugal pump dataset are considered as a 5-class task, so the Softmax output layer of DTCNN(ResNet-50) is replaced with a new Softmax layer with 5 neurons.

2) RESULTS AND DISCUSSION

In this experiment, the time-frequency images with RGB formats are input into DTCNN for training. After 5 epochs, the accuracy curve and loss curve are shown in Fig. 17. As can be seen from Fig. 17, after only 2 epochs, the accuracy and loss curves have been very stable, and the model has started to converge, indicating the strong convergence ability of the proposed model.

The trained model is used to classify the testing dataset, and the work is executed 10 times under the same conditions. The maximum accuracy, minimum accuracy, mean accuracy, standard deviation, and computation time will be counted, and the results are listed in Table 7. As can be seen from the table, the proposed model has achieved 99.98%±0.0332 accuracy, indicating that the proposed model has a very high prediction accuracy. The average computation time is 361s, indicating that the proposed model is fast in diagnosis.

The confusion matrix representation of the worst result (the minimum values in Table 7) is shown in Fig. 18. From the result, one IR is misclassified as NO and one RW is misclassified as OR. We perform sensitivity and specificity analysis on the confusion matrix. The analysis results are listed in Table 8. As can be seen from the table, the sensitivity and specificity of all health conditions can reach 99.75% and above, which demonstrates the effectiveness of this proposed method.

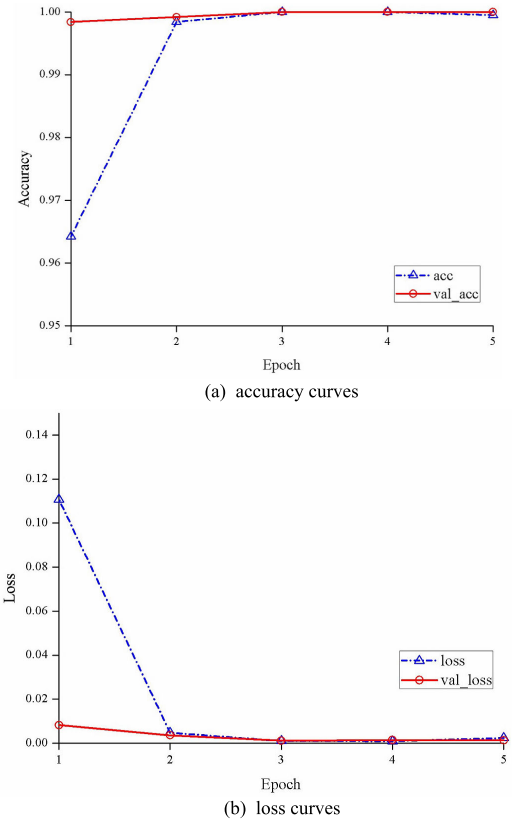

FIGURE 17. The accuracy and loss curves of the proposed DTCNN in Case 2.

TABLE 7. The result of the proposed DTCNN model in Case 2.

Max accuracy	Min accuracy	Mean accuracy	Std	Computation time
100%	99.90%	99.98%	0.0332	361s

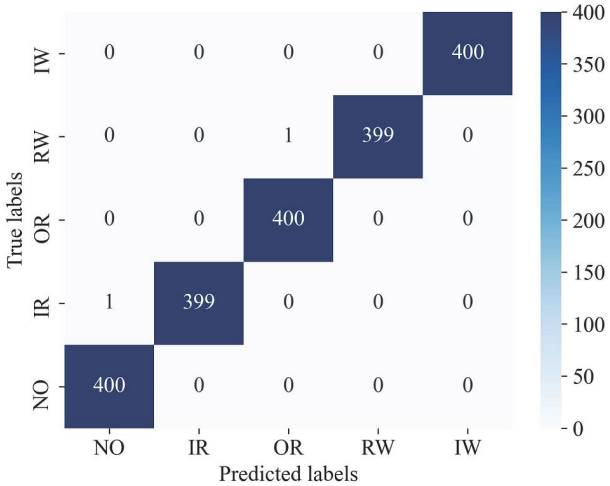

FIGURE 18. The confusion matrix of the worst prediction results in Case 2.

3) COMPARISONS WITH OTHER DTCNN(INCEPTION V3, VGG-16) MODELS

DTCNN(Inception v3, VGG-16, and ResNet-50) models are trained on the self-priming centrifugal pump dataset, and the

TABLE 8. The analysis results of the confusion matrix of the worst prediction accuracy in Case 2.

Condition type	Sensitivity	Specificity	Accuracy
NO	100%	99.94%	
IR	99.75%	100%	
OR	100%	99.94%	99.90%
RW	99.75%	100%	
IW	100%	100%	

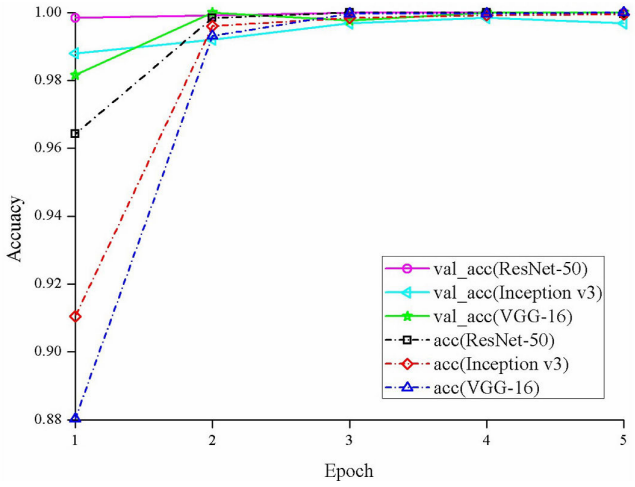

FIGURE 19. The accuracy curves of various DTCNN models in Case 2.

TABLE 9. The comparison results of various DTCNN models in Case 2.

	Inception v3	VGG-16	**ResNet-50**
Max accuracy	99.95%	**100%**	**100%**
Min accuracy	99.55%	99.55%	**99.90%**
Mean accuracy	99.78%	99.94%	**99.98%**
Std	0.1470	0.1338	**0.0332**
Computation time	354s	**342s**	361s

parameters of three DTCNN models are same with Case 1. The accuracy curves are shown in Fig. 19, the maximum accuracy, minimum accuracy, mean accuracy, standard deviation, and computation time are counted and the results are listed in Table 9. As can be seen from Fig. 19, the training and validation accuracy curves of DTCNN(ResNet-50) converges faster than DTCNN(Inception v3) and DTCNN(VGG-16), and DTCNN(Inception v3) has the worst validation accuracy values. As can be seen from Table 9, the computation time of the three different DTCNN models are not much difference. DTCNN(ResNet-50) achieves 99.98% ±0.0332 accuracy, while DTCNN(Inception v3) achieves 99.78%±0.1470 accuracy and DTCNN(VGG-16) achieves 99.94%±0.1338 accuracy, which indicates that DTCNN (ResNet-50) is slightly better than DTCNN(VGG-16) and DTCNN(Inception v3).

4) COMPARISONS WITH OTHER METHODS

The proposed method is compared with the methods in [47] and [22], where [47] is based on speeded up robust features (SURF) and probabilistic neural network (PNN) and [22] is based on 2D images and CNN. The maximum, minimum, mean and standard deviation of the prediction accuracies are taken as the comparison terms, and the results are listed in Table 10. From the table, compared with the method [47], the proposed method achieves significant improvement in accuracy from 98.33%±1.7164 to 99.98%±0.0332, and compared with the method [22], the accuracy of proposed method is improved by 0.50%, which demonstrates that the proposed method has the potential in fault diagnosis.

TABLE 10. The comparison results in Case 2 (%).

Methods	Max accuracy	Min accuracy	Mean accuracy	Std
SURF + PNN [47]	**100**	96.00	98.33	1.7164
CNN [22]	99.74	99.11	99.48	0.1966
Proposed method	**100**	**99.90**	**99.98**	**0.0332**

V. CONCLUSION AND FUTURE WORK

This paper presents a fault diagnosis method for rolling bearing based on time-frequency analysis and DTCNN. The contributions of this paper are summarized as follows:

1) A time-frequency analysis method based on CWT is proposed to convert fault signals into time-frequency images suitable for CNN training. The images are further converted into RGB images in 224 × 224 × 3 formats for input to the ResNet-50.

2) Based on the idea of transfer learning, an end-to-end fault diagnosis DTCNN(ResNet-50) model is proposed. By fine-tuning the last three residual blocks and the new Softmax layer, this model solves the problem that small datasets cannot be trained in a very deep CNN. In addition, DTCNN shares the pre-trained weights in the ImageNet dataset instead of training network from scratch, greatly reducing the number of parameters that need to be updated and the training time is greatly reduced, which are the advantages of the proposed method in this paper.

3) The proposed model is tested on two famous datasets, and they achieve near 100% prediction accuracies, which are superior to other traditional machine-learning methods and other DL methods. We also compare the performance of three different DTCNN(ResNet-50, Inception v3, and VGG-16), and the results demonstrate that the proposed DTCNN (ResNet-50) is the best among them. All these results show that the proposed method has the good performance in fault diagnosis.

In practical applications, the limitations of proposed method are as follows: Firstly, the proposed method still cannot solve the problem of different distributions, because fine-tuning is also based on the assumption that training samples and test samples are identically distributed. Secondly, online fault diagnosis is not achieved in this paper. The future research directions of proposed method are as follow: Firstly, we can improve the proposed method by introducing adaptation layers to solve the problem of different data distribution.

Secondly, we can further study online fault diagnosis based on the proposed DTCNN.

REFERENCES

[1] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part II: Fault diagnosis with knowledge-based and hybrid/active approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3768–3774, Jun. 2015, doi: 10.1109/TIE.2015.2419013.

[2] M. He and D. He, "Deep learning based approach for bearing fault diagnosis," *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 3057–3065, May 2017, doi: 10.1109/TIA.2017.2661250.

[3] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu, "Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery," *IEEE Access*, vol. 5, pp. 15066–15079, 2017, doi: 10.1109/ACCESS.2017.2728010.

[4] B. Li, M.-Y. Chow, Y. Tipsuwan, and J. C. Hung, "Neural-network-based motor rolling bearing fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 47, no. 5, pp. 1060–1069, Oct. 2000, doi: 10.1109/41.873214.

[5] X. Dai and Z. Gao, "From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis," *IEEE Trans. Ind. Informat.*, vol. 9, no. 4, pp. 2226–2238, Nov. 2013, doi: 10.1109/TII.2013.2243743.

[6] M. Arif, K. A. Alam, and M. Hussain, "Application of data mining using artificial neural network: Survey," *Int. J. Database Theory Appl.*, vol. 8, no. 1, pp. 245–270, Feb. 2015, doi: 10.14257/ijdta.2015.8.1.25.

[7] Z. Feng, M. J. Zuo, R. Hao, F. Chu, and J. Lee, "Ensemble empirical mode decomposition-based teager energy spectrum for bearing fault diagnosis," *J. Vib. Acoust.*, vol. 135, no. 3, p. 31013, Jun. 2013, doi: 10.1115/1.4023814.

[8] 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, Aug. 2007, doi: 10.1016/j.ymssp.2006.12.007.

[9] W.-J. Niu, Z.-K. Feng, Y.-B. Chen, H.-R. Zhang, and C.-T. Cheng, "Annual streamflow time series prediction using extreme learning machine based on gravitational search algorithm and variational mode decomposition," *J. Hydrologic Eng.*, vol. 25, no. 5, 2020, Art. no. 4020008, doi: 10.1061/(ASCE)HE.1943-5584.0001902.

[10] L.-L. Jiang, H.-K. Yin, X.-J. Li, and S.-W. Tang, "Fault diagnosis of rotating machinery based on multisensor information fusion using SVM and time-domain features," *Shock Vib.*, vol. 2014, pp. 1–8, Apr. 2014, doi: 10.1155/2014/418178.

[11] X. Qin, Q. Li, X. Dong, and S. Lv, "The fault diagnosis of rolling bearing based on ensemble empirical mode decomposition and random forest," *Shock Vib.*, vol. 2017, pp. 1–9, Aug. 2017, doi: 10.1155/2017/2623081.

[12] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings," *Shock Vib.*, vol. 2017, pp. 1–17, Oct. 2017, doi: 10.1155/2017/5067651.

[13] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "A survey on deep learning for big data," *Inf. Fusion*, vol. 42, pp. 146–157, Jul. 2018, doi: 10.1016/j.inffus.2017.10.006.

[14] S. Wen, W. Liu, Y. Yang, T. Huang, and Z. Zeng, "Generating realistic videos from keyframes with concatenated GANs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 8, pp. 2337–2348, Aug. 2019, doi: 10.1109/TCSVT.2018.2867934.

[15] S. Wen, M. Dong, Y. Yang, P. Zhou, T. Huang, and Y. Chen, "End-to-end detection-segmentation system for face labeling," *IEEE Trans. Emerg. Topics Comput. Intell.*, early access, Nov. 6, 2019, doi: 10.1109/TETCI.2019.2947319.

[16] G. Ren, Y. Cao, S. Wen, T. Huang, and Z. Zeng, "A modified elman neural network with a new learning rate scheme," *Neurocomputing*, vol. 286, pp. 11–18, Apr. 2018, doi: 10.1016/j.neucom.2018.01.046.

[17] S. Wen, W. Liu, Y. Yang, P. Zhou, Z. Guo, Z. Yan, Y. Chen, and T. Huang, "Multilabel image classification via feature/label co-projection," *IEEE Trans. Syst., Man, Cybern. Syst.*, early access, Feb. 6, 2020, doi: 10.1109/TSMC.2020.2967071.

[18] H. Shao, H. Jiang, X. Zhang, and M. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Meas. Sci. Technol.*, vol. 26, no. 11, Nov. 2015, Art. no. 115002, doi: 10.1088/0957-0233/26/11/115002.

[19] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018, doi: 10.1016/j.patcog.2017.10.013.

[20] J. Gehring, Y. Miao, F. Metze, and A. Waibel, "Extracting deep bottleneck features using stacked auto-encoders," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 3377–3381, doi: 10.1109/icassp.2013.6638284.

[21] H. Shao, H. Jiang, F. Wang, and Y. Wang, "Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet," *ISA Trans.*, vol. 69, pp. 187–201, Jul. 2017, doi: 10.1016/j.isatra.2017.03.017.

[22] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018, doi: 10.1109/TIE.2017.2774777.

[23] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–14.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.

[28] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019, doi: 10.1016/j.ymssp.2018.05.050.

[29] 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, doi: 10.1016/j.ymssp.2018.02.016.

[30] R. Zhang, H. Tao, L. Wu, and Y. Guan, "Transfer learning with neural networks for bearing fault diagnosis in changing working conditions," *IEEE Access*, vol. 5, pp. 14347–14357, 2017, doi: 10.1109/ACCESS.2017.2720965.

[31] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446–2455, Apr. 2019, doi: 10.1109/TII.2018.2864759.

[32] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 3320–3328.

[33] Z. Feng, M. Liang, and F. Chu, "Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples," *Mech. Syst. Signal Process.*, vol. 38, no. 1, pp. 165–205, Jul. 2013, doi: 10.1016/j.ymssp.2013.01.017.

[34] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Process. Mag.*, vol. 9, no. 2, pp. 21–67, Apr. 1992, doi: 10.1109/79.127284.

[35] W. C. Lang and K. Forinash, "Time-frequency analysis with the continuous wavelet transform," *Amer. J. Phys.*, vol. 66, no. 9, pp. 794–797, Sep. 1998, doi: 10.1119/1.18959.

[36] J. Tezcan, J. Cheng, and Q. Cheng, "Modeling and prediction of nonstationary ground motions as time–frequency images," *IEEE Trans. Geosci. Remote Sens.*, early access, Sep. 4, 2014, doi: 10.1109/TGRS.2014.2347335.

[37] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[38] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, Haifa, Israel, 2010, pp. 1–8.

[39] *Case Western Reserve University Bearing Data Center*. Accessed: Feb. 5, 2020. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website

[40] W. Du, J. Tao, Y. Li, and C. Liu, "Wavelet leaders multifractal features based fault diagnosis of rotating mechanism," *Mech. Syst. Signal Process.*, vol. 43, nos. 1–2, pp. 57–75, Feb. 2014, doi: 10.1016/j.ymssp.2013.09.003.

[41] Y. Li, X. Wang, and J. Wu, "Fault diagnosis of rolling bearing based on permutation entropy and extreme learning machine," in *Proc. Chin. Control Decis. Conf. (CCDC)*, May 2016, pp. 2966–2971, doi: 10.1109/CCDC.2016.7531490.

[42] X. Zhang, Y. Liang, J. Zhou, and Y. Zang, "A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM," *Measurement*, vol. 69, pp. 164–179, Jun. 2015, doi: 10.1016/j.measurement.2015.03.017.
[43] 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, doi: 10.1016/j.knosys.2017.10.024.
[44] 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.*, vols. 72–73, pp. 92–104, May 2016, doi: 10.1016/j.ymssp.2015.11.014.
[45] M. Sohaib, C.-H. Kim, and J.-M. Kim, "A hybrid feature model and deep-learning-based bearing fault diagnosis," *Sensors*, vol. 17, no. 12, p. 2876, Dec. 2017, doi: 10.3390/s17122876.
[46] G. Xu, M. Liu, Z. Jiang, D. Söffker, and W. Shen, "Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning," *Sensors*, vol. 19, no. 5, p. 1088, Mar. 2019, doi: 10.3390/s19051088.
[47] C. Lu, Y. Wang, M. Ragulskis, and Y. Cheng, "Fault diagnosis for rotating machinery: A method based on image processing," *PLoS ONE*, vol. 11, no. 10, Oct. 2016, Art. no. e0164111, doi: 10.1371/journal.pone.0164111.

ZHIHAO CHEN received the B.S. degree from the School of Automation, Guangdong Polytechnic Normal University, where he is currently pursuing the M.S. degree in control theory and control engineering. His current research interests include fault diagnosis and image processing.

JIAN CEN received the B.S. degree in computer application from the China University of Petroleum, in 1996, and the M.S. and Ph.D. degrees in control theory and control engineering from the South China University of Technology, in 2006 and 2010, respectively. She is currently a Professor with the School of Automation, Guangdong Polytechnic Normal University. Her research interests include sensor networks, fault diagnosis, and other fields.

JIANBIN XIONG (Member, IEEE) received the B.E., M.A., and Ph.D. degrees from the Guangdong University of Technology, China. He is currently a Professor with the School of Automation, Guangdong Polytechnic Normal University. He has authored or coauthored over 40 papers in related international conferences, journals, and books. He has served as a reviewer of over 20 journals. His research interests include fault diagnosis, signal processing, image processing, information fusion, and computer applications.