IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 14, 2022, accepted April 30, 2022, date of publication May 9, 2022, date of current version May 17, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3173444

Intelligent Machine Fault Diagnosis Using Convolutional Neural Networks and Transfer Learning

WENTAO ZHANG, TING ZHANG, GUOHUA CUI, AND YING PAN
School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

Corresponding author: Ting Zhang (zhangt@sues.edu.cn)

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 11702168.

ABSTRACT With the development of automated and integrated large-scale industrial systems, accurate and effective fault diagnosis methods are required to ensure the security and reliability of running mechanical equipment. Due to the time consumption and poor generalization performance of conventional machine learning-based methods, deep learning (DL)-based methods have wider application prospects due to their end-to-end architectural properties. However, in the DL models, problems such as a large number of trainable parameters, complicated hyperparameter tuning, and initialization instability increase the difficulty of model training and limit higher performance. To address these disadvantages of the DL method, we proposed a novel DL framework by applying convolutional neural networks (CNNs) based on the optimization of transfer learning (TL). TL can help the model achieve higher precision with less computational cost by transferring low-level features and fine-tuning high-level layers. In addition, data processing was implemented using continuous wavelet transformation (CWT) to convert vibration signals into 2-D images, and support vector machines (SVM) were employed to replace the fully connected layers for better classification. As a result, the proposed method was superior to the classical deep architecture trained from scratch. The performance of the proposed method is analyzed by presenting testing reports, convergence curves, and confusion matrixes. Moreover, experiments comprised of cross-domain diagnosis, simulated composite fault detection, and performance comparison on seven mechanical datasets, including bearings, gearboxes, and rotors, are presented. Based on these results, it can be observed that our method achieved the highest accuracy under various conditions.

INDEX TERMS Convolutional neural network, fault diagnosis, deep learning, continuous wavelet transformation, transfer learning, support vector machine.

I. INTRODUCTION

Nowadays, developments in mechanical and electrical equipment are focused on large-scale, automation and integration. Coupled with the complexity and polytropic properties of the operating conditions and working environment in the mechanical and electrical fields, the probability of failure increases gradually, so there is an urgent need for accurate and effective fault diagnosis for complex equipment to improve system security and reliability.

Mechanical fault diagnosis is a comprehensive technology that crosses multiple disciplines to monitor, diagnose and predict the running state and ensure the safe operation of machine equipment, and can essentially be considered pattern recognition and classification issues. Traditionally, the process of fault diagnosis can be divided into three key stages: data acquisition, feature extraction, and health state recognition. Data acquisition usually refers to the employment of multiple sensors that are installed on machines to collect data such as vibration, current, and instantaneous speed. Feature extraction involves the extraction of some sensitive features from collected data by converting the data to a low-dimensional feature vector representation, generally including time-domain analysis, frequency-domain analysis, and time-frequency-domain analysis. For health state recognition, machine learning-based diagnosis models are used to establish mapping relationships between extracted features and corresponding health statuses.

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

Researchers have been actively exploring the field of signal processing, feature extraction, and intelligent fault diagnosis, and many diagnosis methods have been proposed for certain problems. Gu *et al.* [1] proposed a method based on the filtering algorithm, Hilbert-Huang transform (HHT), and energy entropy to extract the fault characteristic of the rotor bearing system, and then used support vector machines (SVMs) to defect fault types. This signal processing method was validated to be effective through experiments. Wang and Chan [2] combined wavelet packet transform (WPT), local weighted scatter smoothing method (LOWESS), and least square support vector machine (LSSVM) to detect gear wear degree, and the final diagnosis accuracy reaches 98.33%. Yang *et al.* [3] used complete-information-based principal component analysis (CIPCA) to reduce data dimensionality, and then used a back-propagation neural network (BPNN) to predict the failure of unmanned aerial vehicles. The proposed CIPCA-BPNN method can make accurate predictions before the failure occurred.

Even though these methods have achieved high accuracy, there are still several limitations. Artificial feature extraction greatly relies on expert knowledge, which means that it usually needs complex mathematical operations and some understanding of the signal to be processed. For some complex systems with external environmental interference and nonlinear inner-coupling, the shallow structure of conventional methods is not sufficient to mine features sensitive to all types of faults [4]. Furthermore, the designed diagnosis methods are usually applied to carry out some specialized tasks but are not applicable to others. Therefore, it is difficult to design a method that provides reliable precision in all situations. To improve the generalization and robustness of the algorithm, deep learning (DL)-based methods with self-adaptability have been widely used.

As an important branch of machine learning, DL has expanded the field of artificial intelligence and been successfully applied in many other research fields, such as object detection [5], image segmentation [6], natural language processing [7], fault diagnosis [8], visual tracking [9], and smart manufacturing [10]. In DL, which is derived from the research of neural networks, hierarchical representations from original data are learned in deep architectures with multiple hidden layers. As a result, abstract features can be extracted automatically, and the uncertainty caused by human interference can be reduced. In general, DL-based methods have end-to-end characteristics that can be used to spontaneously complete the whole process of feature extraction, data dimension reduction, and health status recognition. Moreover, deep architectures can represent the complex mapping relationship between signals and health status well and are appropriate for fault diagnosis tasks that have diverse, nonlinear, and high-dimensional characteristics. Therefore, DL-based methods can overcome the limitations of conventional diagnosis methods and provide a novel alternative for intelligent fault diagnosis.

DL-based models have been successfully utilized for machine fault diagnosis tasks. For instance, Ma *et al.* [11] presented an information fusion method based on the variational autoencoder (VAE) and random forest (RF) for fault diagnosis of rolling bearings, achieving a classification rate of 98.19%. Han *et al.* [12] developed a novel diagnosis framework that combines the spatiotemporal pattern network (STPN) and convolutional neural networks (CNNs), and the performance of this hybrid scheme was validated on the wind turbine and bearing data sets. Jang and Cho [13] extracted features through short-time Fourier transform (STFT) and classified the fault type of rotating machinery by using an attentional autoencoder (AE) and a 1D CNN LSTM, which realized fault diagnosis under different working conditions. Zhao *et al.* [14] designed the multilabel cycle translating adversarial network (MCTAN) to address the deficiency of fault data in industrial applications, which truly improves the performance of DL-based approaches.

Nevertheless, deep learning models also have some deficiencies. Because of the large size of hidden layers in deep architecture, the number of trainable parameters increases rapidly, which greatly increases the calculated amount, time consumption, and training difficulty. In addition, training a network from scratch usually means random initialization of weights and bias, whose indeterminacies could cause lower efficiency and even influence the final results. Moreover, training deep architectures require an abundance of hyperparameter tuning, which is generally determined subjectively but greatly affects performance. According to the abovementioned deficiencies, the application of DL-based methods could be a laborious and difficult task.

To make the DL-based algorithm more efficient and easier to implement, the transfer learning (TL) strategy is adopted in this paper. Instead of random initialization to train the network from scratch, a deep model trained from sufficient data as a start point is introduced using TL. Then, knowledge acquired from previous related issues is applied to solve the immediate problem. As a result, TL optimizes a deep architecture with much better results at a lower cost, improves the efficiency of the training process, and increases the enablement and operability of DL-based methods [15].

TL has also been successful in cases of fault diagnosis research. Shao *et al.* [16] were able to achieve high accuracy on three main mechanical datasets by taking time-frequency images as inputs and utilizing TL to accelerate the training of CNN. Zhang *et al.* [17] proposed a fault diagnosis approach combining DCNN and TL to detect faults in a timely and accurate manner. Wen *et al.* [18] applied a new TL-based sparse autoencoder for fault diagnosis across different working conditions, and a test accuracy as high as 99.82% was achieved. He *et al.* [19] used ensemble transfer CNNs to analyze multi-channel signals of rotating machinery cross working conditions, which shows superiorities by fully combining the properties of DL, TL, and ensemble learning (EL).

In this study, a novel intelligent fault diagnosis method using continuous wavelet transformation (CWT), CNN, SVM, and TL strategies are proposed for the detection of rotating machinery health status. The main contributions of this study are summarized as follows:

1) An effective fault diagnosis method is presented for rotating machinery. CWT is used to process input vibration signals and convert them into RGB images. The diagnosis framework is based on CNN and SVM, where CNN adaptively extracts high-level abstract features, and SVM conducts fault category recognition to replace fully connected layers.
2) The TL strategy is adopted to accelerate the training process of CNN. All weight values in the diagnosis architecture are transferred from a pretrained model trained on ImageNet. Lower-level layers of the pretrained model have common knowledge on image recognition and their weights are frozen. High-level layers are applied to specific tasks and are fine-tuned based on the machinal fault dataset.
3) We conduct a series of comparative experiments. First, the performance of CNN models trained from scratch is compared with using a pretrained model. Second, a classifier SVM and fully connected layers with softmax regression are compared with accuracy. Finally, our proposed method is comprehensively compared with other intelligent fault diagnosis algorithms, including single point diagnosis, compound failure detection, and comparisons on multiple fault datasets.

The remaining content of this paper is organized as follows. Section II briefly introduces the theoretical knowledge, including CWT, CNN, SVM, and TL. The overall workflow of the diagnosis process is introduced in Section III. In Section IV, we introduce related experimental settings and carry out some preliminary research. More comparative experiments are presented in Section V. Finally, conclusions and future work are presented in Section VI.

II. THEORETICAL BACKGROUND

A. CONTINUOUS WAVELET TRANSFORM

Joint time-frequency analysis (JTFA) is a useful technique to process nonstationary signals and provides joint distribution information about time- and frequency-domains. We utilize JTFA to obtain time-frequency images from one-dimensional signals since CNN requires image data as input. Common methods of JTFA include STFT, HHT, and CWT. Among these methods, CWT is desirable for signal time-frequency analysis and processing.

CWT decomposes a signal into components at different scales and gradually refines the signal at multiple scales. High frequency resolution in the low frequency range and high time resolution in the high frequency range are achieved through CWT, which is automatically adapted to the requirements of time-frequency signal analysis [20]. Inner product operation of the raw signal and wavelet functions is conducted by wavelet transform. The wavelet function family is obtained from the temporal telescopic and translational operation of the mother wavelet, which is shown as:

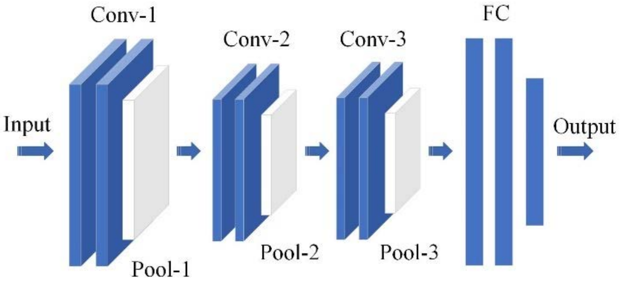


FIGURE 1. The basic architecture of CNN.

$$\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-\tau}{a}\right) \tag{1}$$

where a, τ are the scale factor and translation factor, respectively, and $\psi(t)$ is the mother wavelet. The wavelet basis *Morlet* was selected in this study. CWT executes convolutional operation of signal $f(t)$ and wavelet functions $\psi_{a,\tau}(t)$. For a function $f(t) \in L^2(R)$, the mathematical expression of CWT is defined as:

$$\begin{aligned} CWT_f(a,\tau) &= \langle f(t), \psi_{a,\tau}(t)\rangle \\ &= \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} f(t)\,\psi^*\left(\frac{t-\tau}{a}\right)dt \end{aligned} \tag{2}$$

where $\psi^*(\cdot)$ denotes the complex conjugate of $\psi(\cdot)$. Through this operation, the one-dimensional time-domain signal $f(t)$ is converted to the joint distribution on the time-frequency domain, with two parameters a, τ.

B. CONVOLUTIONAL NEURAL NETWORK

CNN is a widely used deep learning model that has powerful abilities in the field of image processing. It can achieve automatic abstracts of hierarchical features to avoid manual feature extraction operations. Relying on its characteristics of local receptive fields, weight sharing, and sparse connections, a CNN usually contains fewer parameters than fully connected network, which means it can more efficiently utilize data, dramatically reduce the training difficulty, and is not easily overfit [21]. In general, the architecture of a standard CNN commonly contains convolutional layers, pooling layers, and fully connected layers, as shown in Fig. 1.

1) CONVOLUTIONAL LAYER

The operation of feature extraction is conducted through convolutional layers, where an identical square convolutional kernel, which can be regarded as a scanner with a specified window size, is used to slide on the input feature graph and scan every pixel of the input feature graph according to a specified stride. For each step, the convolutional kernel will coincide with several pixels and corresponding elements in the overlap area will be multiplied, summed, and offset to obtain a new pixel value of the output feature image. Features are extracted through a convolution filter with multiple kernels, and the mathematical calculation of the convolution operation is given by:

$$x_j^l = f\left(\sum_{i\in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \tag{3}$$

where M_j represents the localized receptive area of feature graphs, k_{ij}^l indicates the jth weight value of the convolutional kernel in the lth layer, x_j^l is the jth pixel value of feature images om the lth layer, b_j^l is the jth bias value in the lth layer kernel, $*$ represents the convolution operation, $f(\cdot)$ is the activation function, including the sigmoid function, tanh function and rectified linear unit (ReLU), which are expressed as:

$$\text{sigmoid}(x) = \frac{1}{1+e^{-x}} \tag{4}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{5}$$

$$\text{ReLU}(x) = \max\{0, x\} \tag{6}$$

2) POOLING LAYER

Pooling layers are commonly set behind the convolutional layers, and they are combined to form convolution blocks. Multiple convolution blocks are stacked to build the deep architecture. Pooling layer perform subsampling operation to remove the redundant information of image data, reduce the amount of feature data, and improve the robustness and calculation efficiency of the algorithm. The pooling operation is mathematically defined as:

$$x_j^l = f\left(\beta_j^l \text{down}\left(x_i^{l-1}\right) + b_j^l\right) \tag{7}$$

where *down* $(\cdot)$ indicates the pooling operation, and the most commonly used pooling processes are average pooling and maximum pooling, which means the average value or the maximum value within a pooling region is selected to be propagated to the next layer, respectively.

3) FULLY CONNECTED LAYER

The fully connected layers receive the one-dimensional extracted features from previous convolution blocks and perform classification or regression. Softmax regression is commonly used at the output layer to output a probability distribution, which can represent the final predicted results of each category, and is mathematically defined as:

$$\text{Softmax}\left(X^{(i)}\right) = \begin{bmatrix} p\left(y^{(i)}=1\right) \\ p\left(y^{(i)}=2\right) \\ \vdots \\ p\left(y^{(i)}=n\right) \end{bmatrix} = \frac{1}{\sum_{i=1}^{n} e^{\theta_k^T X^{(i)}}} \begin{bmatrix} e^{\theta_1^T X^{(i)}} \\ e^{\theta_2^T X^{(i)}} \\ \vdots \\ e^{\theta_n^T X^{(i)}} \end{bmatrix} \tag{8}$$

where $X^{(i)}$ is the input value of the ith sample, $y^{(i)} \in \{1, 2, \ldots, n\}^T$ represents the corresponding predicted label of the ith sample, and $\theta \in [\theta_1, \theta_2, \ldots, \theta_n]^T$ denotes the softmax parameter. The softmax classifier makes each element of the output vector a positive value, and all elements sum to 1, which coordinates the cross-entropy loss to update the parameters in the network.

C. SUPPORT VECTOR MACHINE

Support vector machine (SVM) is a generalized linear classifier of binary data through supervised learning, whose decision boundary is the maximum-margin hyperplane of learning samples. By using the kernel method, the nonlinear classification can be addressed. With many unique advantages, SVM is mostly applied to complete classification tasks of multiclass, nonlinear, and high-dimensional samples by searching for the optimal separating hyperplane with minimum generalization error.

For a training dataset $Z = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, where x_i denotes the ith input feature vector, y_i is the corresponding category label and $i = 1, 2, \ldots, n$ represents the sample number. The maximum-margin separating hyperplane $w^T x_i + b = 0$ is built in feature space, where the weight vector w and offset term b are optimization parameters, and $1/\|w\|$ represents the margin, whose maximization is equal to minimizing $\|w\|^2$. By introducing the slack variable ξ_i and penalty factor C, the learning problem of SVM, which is equivalent to the soft interval maximization problem, can be described as follows:

$$\begin{aligned} &\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \\ &s.t.\ \ y_i\left(w^T x_i + b\right) \geq 1 - \xi_i, \quad \forall (x_i, y_i) \in Z \\ &\qquad \xi_i \geq 0, i = 1, 2, \ldots, N \end{aligned} \tag{9}$$

where w and b are adjustable parameters to minimize the objective function, the penalty parameter C controls the tradeoff between smoothing decision boundary and correcting classification training points, and the slack variable ξ_i relaxes training data to prevent overfitting. By selecting a suitable kernel function $K(x, x_i)$ and a penalty parameter C, the classification decision function can be given by:

$$f(x) = \text{sign}\left(\sum_{i=0}^{n} a_i^* y_i K(x, x_i) + b^*\right) \tag{10}$$

where a_i^* and b^* are the optimal solutions, and kernel function $K(x, x_i)$ is the critical technology of SVM, which has a higher impact on the model performance. The kernel function mainly includes the linear kernel, polynomial kernel, and radial basis function (RBF). The mathematical expression of these kernel functions is defined as:

$$\text{Linear}: K(x, x_i) = x^T x_i + c \tag{11}$$

$$\text{Polynomial}: K(x, x_i) = \left(a x^T x_i + c\right)^d \tag{12}$$

$$\text{RBF}: K(x, x_i) = \exp\left(-\gamma\|x - x_i\|^2\right) \tag{13}$$

In this study, SVM is used to classify the output features extracted by CNN, which are likely linearly separable. Therefore, we first attempt to adopt a linear kernel to construct a suitable multiclass classifier.

D. TRANSFER LEARNING OF CNN

TL is an essential machine learning method, where previously learned knowledge is applied to help solve new problems faster. Instead of training a neural network with a deep architecture from scratch by randomly initializing a large number of weights, which is a time-consuming process that occupies more computing resources, transferring the weights from a pretrained model as a starting point is a more efficient process. These pretrained models are already trained by another dataset to complete a task, and the knowledge learned has been stored in the pretrained model's weights which are transferred to the new task.

The concept domain denotes the data space and distribution, and there are two domains involved in TL: source domain D_s and target domain D_t. These domains have different distributions, but they are related to each other to a certain extent. Let $D_s = \{\chi_s, P(X_s)\}$ and $D_t = \{\chi_t, P(X_t)\}$ denote the source and target-domain datasets respectively, where χ is the feature space and $P(X)$ is the marginal probability distribution. Their corresponding tasks are expressed as $T_s = \{Y_s, f_s(\cdot)\}$ and $T_t = \{Y_t, f_t(\cdot)\}$, where Y is the label space and $f(\cdot)$ is the prediction function. Then, D_s and D_t are assumed to be sampled from different marginal distributions $P(X_s)$ and $P(X_t)$, respectively. The pretrained model is obtained in the source domain D_s, and it contains knowledge of the updated weights to fit the prediction function $f_s(\cdot)$. Through the transfer of weights, the knowledge information can be applied from D_s to D_t, making the model trained in D_s faster to convergence and fit prediction function $f_t(\cdot)$, which means it can more accurately predict the label Y_t corresponding to χ_t.

As an algorithm suitable for learning hierarchical representations from images, CNNs usually extract common features like edges and curves from images in the lower-level layers, which are appropriate for most image classification tasks, while the high-level layers tend to learn more abstract representations that are applied to a small minority of specific situations. Therefore, the weights in the lower-level can be transferred and frozen in the newly established model, while the weights of the higher hidden layers need to be updated based on the new dataset to complete the task of the target domain; this process is called fine-tuning. TL of CNNs using pretrained model with natural images has been successful in research areas such as biomedical image recognition. In this study, a CNN model trained on ImageNet dataset will be applied to recognize time-frequency images from the mechanical fault dataset [16].

III. METHODOLOGY

We proposed a novel machinal fault diagnosis method that combines CWT, deep CNNs, TL, and SVMs to detect the running status of rolling machinery with high precision accuracy. In CWT, one-dimensional signals are converted to 2-D images. In CNNs, deep architectures are used to extract highly abstract features from input data. In TL, an optimization strategy is adopted to help improve model performance and reduce time consumption. In SVMs, classifiers are created to complete status recognition and output final results.

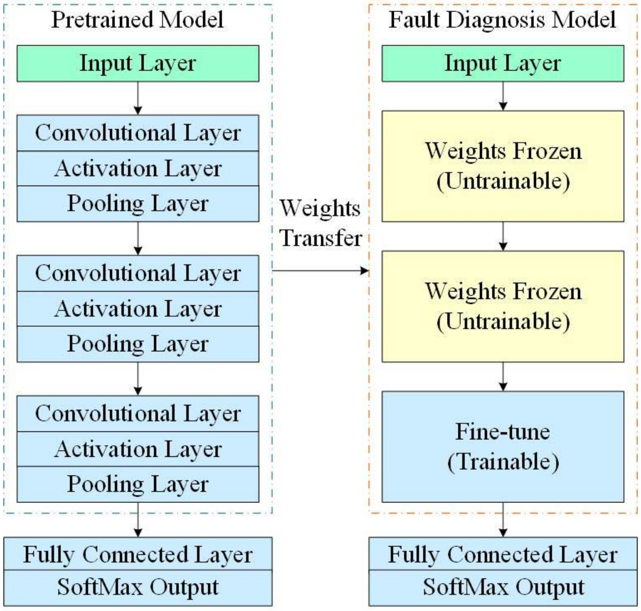


FIGURE 2. Transfer learning procedure.

The proposed method consists of the following stages: data acquisition, time-frequency imaging, data partitioning, pretrained model building, CNN model fine-tuning, SVM classifier training, and performance evaluation. The whole framework of the proposed method is shown in Fig. 3.

1) Data acquisition: The data used in this study are open-source public datasets, that can be obtained from major websites, and the CWRU bearing fault dataset will be used as a typical case. These data are one-dimensional vibration signals acquired by sensors.
2) Time-frequency imaging: As the input shape of CNN is 2-D images with 3 channels, the one-dimensional signals need to be divided into fixed length samples and then converted to time-frequency images by CWT.
3) Data partitioning: All of the images are divided into a training dataset and a testing dataset. The training data are fed into the pretrained model to update the internal parameters, and the testing data are used to verify the performance of model.
4) Pretrained model building: The pretrained models in this work are classical deep CNNs trained on the ImageNet dataset. These CNNs are loaded with weights trained on ImageNet by removing top layers and replacing them with global average pooling. Then, two fully connected layers with softmax regression followed, whose weights were randomly initialized.

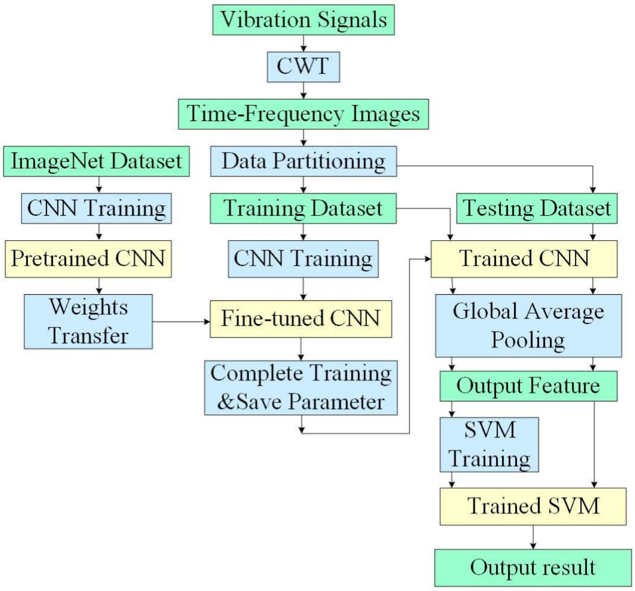


FIGURE 3. The workflow of the proposed fault diagnosis method using TL strategy.

5) Fine-tuning: The parameters in the low-level layers are set to be untrainable, while the parameters of the high-level layers are all trainable, they will be fine-tuned based on the mechanical fault datasets. After model convergence, all parameters in the deep architecture will be saved.
6) SVM training: The saved CNNs extract high-level features of images by applying GAP, and these features are used to train the SVM classifier in the form of a 2-D tensor. By adjusting hyperparameters γ and C, the decision function with the highest classification accuracy will be used to recognize health status.
7) Performance evaluation: The final fault diagnosis models are comprised of the completely trained CNNs and SVMs. The testing dataset is used to evaluate the performance of this model with indicator accuracy.

IV. EFFECTIVENESS VALIDATION

A. DATASET DESCRIPTION

The fault experimental dataset utilized in this study is provided by the Case Western Reserve University (CWRU) Data Center [22]. This database is a standard reference and is extensively used to validate the proposed approach. As shown in Fig. 4, the test rig for the CWRU dataset consists of a 2 hp Reliance Electric motor (left), a torque transducer/encoder (middle), and a dynamometer (right).

The vibration data used in this study were collected at a sampling frequency of 12 kHz from the drive end bearings of the motor under four different operational conditions with bearing loads ranging from 0-3 hp. Single-point failures of SKF deep-groove ball bearings were manufactured by using electrical discharge machining (EDM) technology, with wear diameters of 0.007 in (0.1778 mm), 0.014 in (0.3556 mm), and 0.021 in (0.5334 mm) seeded on the rolling elements, inner raceway, and outer raceway, respectively. Therefore, each operational condition includes 9 fault types corresponding to 3 different fault severities and 3 different fault locations. Adding the normal condition into the consideration, bearing fault diagnosis can be considered a classification task with 10 running statuses. Fig. 5 shows these 9 types of time-domain vibration signals, which vary in amplitude and frequency components.

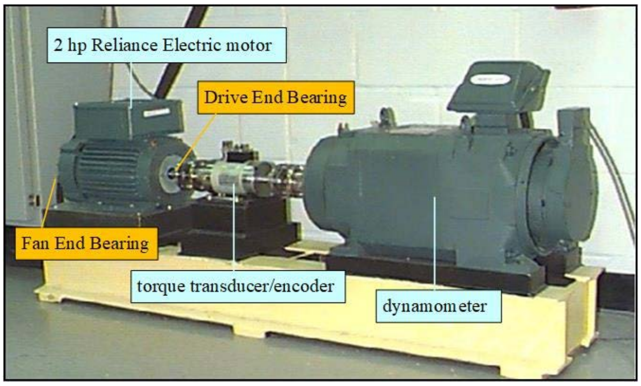


FIGURE 4. CWRU bearing fault diagnosis testbench.

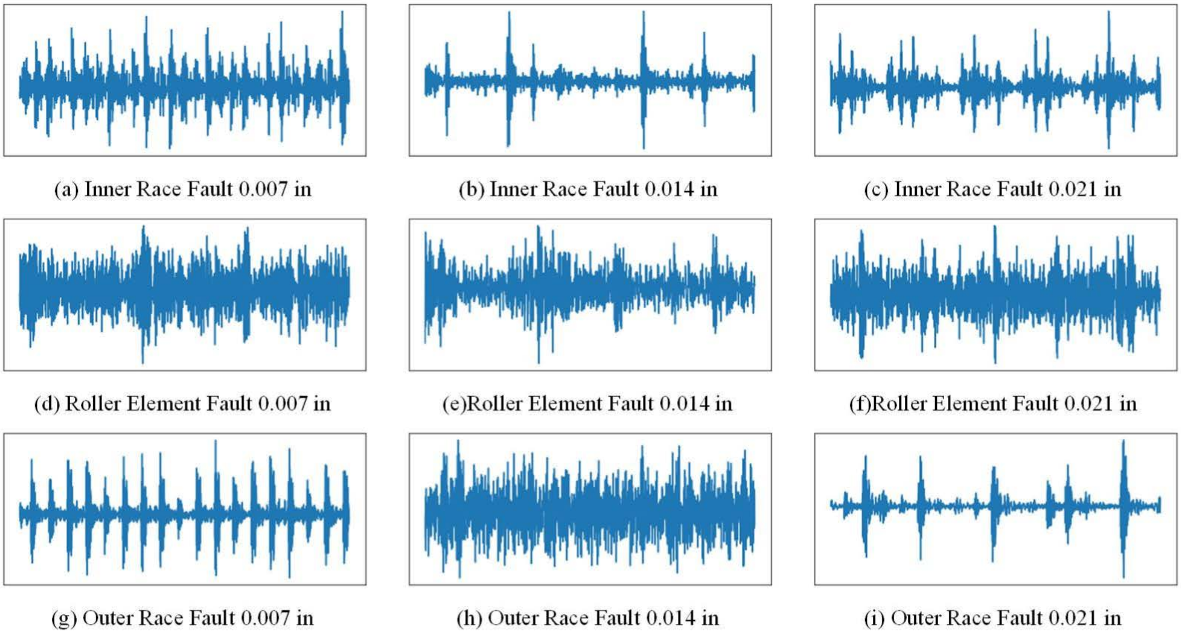
TABLE 1. Description of the rolling bearing fault dataset under each operational condition.

Class Label	Fault Type (Fault Location/Size)	Training Samples	Test Samples
0	Normal	500	100
1	0.007 in inner race (IF07)	500	100
2	0.014 in inner race (IF14)	500	100
3	0.021 in inner race (IF21)	500	100
4	0.007 in roller element (RF07)	500	100
5	0.014 in roller element (RF14)	500	100
6	0.021 in roller element (RF21)	500	100
7	0.007 in outer race (OF07)	500	100
8	0.014 in outer race (OF14)	500	100
9	0.021 in outer race (OF21)	500	100

The overlap sampling method was adopted to obtain training samples from the raw vibration signal. There is overlap between each segment of the signal and the one following it, and we set the slip length to 200 data points. The training set and test set are comprised of 500 training samples and 100 testing samples for each running status, respectively. Each sample contains 1000 data points and then will be transformed into a 64 × 64 size image with 3 channels by CWT. Data from all load conditions are included in the experimental dataset. Table 1 shows the sample distribution in the experiment.

B. HYPERPARAMETER SELECTION

The CNN model as a feature exactor is the most significant component of the fault diagnosis model. Therefore, there is a sufficient necessity to select appropriate hyperparameters to build an effective CNN. In general, the hyperparameters mainly include learning rates, optimizers, mini-batch



(a) Inner Race Fault 0.007 in

(b) Inner Race Fault 0.014 in

(c) Inner Race Fault 0.021 in

(d) Roller Element Fault 0.007 in

(e)Roller Element Fault 0.014 in

(f)Roller Element Fault 0.021 in

(g) Outer Race Fault 0.007 in

(h) Outer Race Fault 0.014 in

(i) Outer Race Fault 0.021 in

FIGURE 5. **Raw time-domain signal of vibration for each status.**

TABLE 2. **The experimental results of five optimizers.**

Optimizer	Learning Rate	Test Accuracy	Training time (s)
SGD	0.01	0.985	69.837
RMSProp	0.0001	0.992	78.544
AdaGrad	0.001	0.964	71.696
AdaDelta	0.01	0.978	72.821
Adam	0.0001	0.993	71.549

size, network structure, and dropout rates. In this section, we implement multiple trials to investigate the influence of different hyperparameter settings by using the Keras library.

1) OPTIMIZER SELECTION

The optimizer determines the training time and convergence speed of the network. The alternative optimizers include the steepest gradient descent (SGD), RMSProp, AdaGrad, AdaDelta, and Adam. We trained CNN models in 15 epochs by using these optimizers with an effective learning rate, and training time and test accuracy are listed in Table 2. Fig. 6 illustrates the training accuracy curves of the five optimizers. It can be clearly seen that RMSProp and Adam achieve the highest accuracy and the fastest convergence speed. However, RMSProp takes the longest training time. To sum up, Adam with a learning rate of 0.0001 is adopted as the optimizer in subsequent experiments.

2) MINI-BATCH SIZE SELECTION

To prevent local optima, the mini-batch gradient descent method is usually used to train network models. In this section, the appropriate value of mini-batch size is selected and training time and test accuracy are noted as evaluation indicator. Table 3 records the experimental results. When the mini-batch size is 64 or 128 does the model achieve a slightly higher accuracy than the others. However, the time consumption is the lowest when the mini-batch size is 128. Hence, we set the value of the mini-batch size as 128 in the following experiment to accelerate model training.

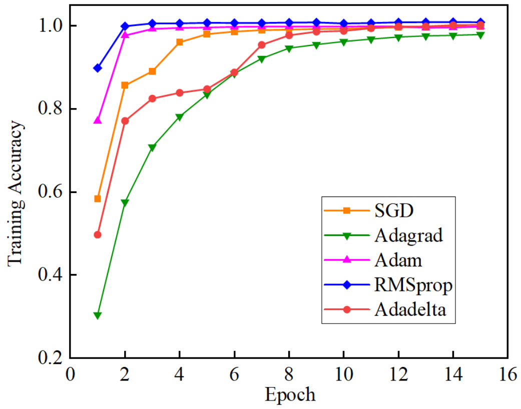


FIGURE 6. **The accuracy curves comparison of five optimizers.**

3) MODEL ARCHITECTURE SELECTION

The architecture of a neural network has a critical impact on the model performance. There are several CNNs can be obtained with pretrained weights from the application module in Keras Library, including VGGNet, ResNet, Xception, MobileNet, and DenseNet. These CNNs have been verified

TABLE 3. The experimental results of various mini-batch sizes.

Mini-batch Size	Test Accuracy	Training Time(s)
16	0.986	138.031
32	0.989	97.527
64	0.992	79.636
128	0.993	71.549
256	0.987	73.617

TABLE 4. The experimental results of five CNNs.

Model	Parameters	Test Accuracy	Training time (s)	Size (MB)
VGG16	14,982,474	0.991	84.213	162.110
ResNet50	24,641,930	0.853	136.092	237.379
Xception	21,915,698	0.984	94.802	210.358
MobileNet	3,758,794	0.985	26.968	42.042
DenseNet121	7,567,434	0.993	71.549	70.426

TABLE 5. The overview of CNNs trained on ImageNet.

Model	Size (MB)	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth
Xception	88	0.790	0.945	22,190,480	126
VGG16	528	0.713	0.901	138,357,544	23
VGG19	549	0.713	0.900	143,667,240	26
ResNet50	98	0.749	0.921	25.636.712	-
InceptionV3	92	0.779	0.937	23.851.784	159
Inception-ResNetV2	215	0.803	0.953	55,873,736	572
MobileNet	16	0.704	0.895	4,253,864	88
MobileNetV2	14	0.713	0.901	3,538,984	88
DenseNet121	33	0.750	0.923	8,062,504	121
NASNet-Mobile	23	0.744	0.919	5,326,716	-
NASNet-Large	343	0.825	0.960	88,949,818	-

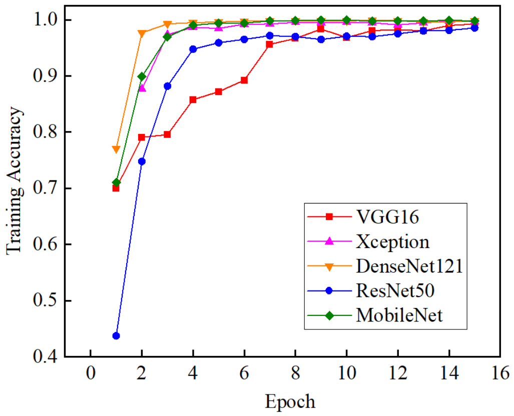


FIGURE 7. The accuracy curves comparison of five CNNs.

to be valid and are easily implemented on our dataset. Table 5 shows the top-1 and top-5 accuracies refer to the CNN's performance on the ImageNet validation dataset.

These CNNs are introduced simply as follows:

1) VGGNet: VGGNet is a deep CNN developed by Visual Geometry Group of Oxford University. It uses multiple small convolution kernels to replace a large convolution kernel and successfully increases the network depth to 16-19 layers with fewer parameters [23].
2) ResNet: ResNet was proposed by He *et al.* in 2015. By using residual structures, a deep architecture with 152 layers was successfully trained. This mechanism greatly alleviates the vanishing or exploding gradients caused by the deep architecture [24].
3) Xception: Xception is an improvement to InceptionV3 structure proposed by Google. It uses depthwise separable convolution to replace the original convolution operation and introduces residual connections to accelerate convergence [25].
4) MobileNet: MobileNet uses deep separable convolution to build a lightweight network, which is appropriate for mobile and embedded devices. Compared to other advanced models, it shows almost equally powerful performance with minimal memory [26].
5) DenseNet: DenseNet is proposed in 2017. It uses the novel dense connection mechanism to realize feature reuse, which avoids vanishing gradients and achieves better performance with fewer parameters [27].

These CNNs were trained in 15 epochs, except ResNet50 was trained longer to alleviate overfitting. The performance of different CNNs is compared through convergence speed, training time, and test accuracy, as shown in Fig. 7 and Table 4. Fig. 7 illustrates the convergence of each CNN through accuracy curves. All CNNs achieve extremely high accuracy, almost 100%, on the training set after 15 epochs, and DenseNet121 has the fastest convergence speed. Table 4 demonstrates the efficiency of five CNNs. It is obvious that VGG16 and DenseNet121 have the highest precision within a shorter training time, but VGG16 occupies more space. In addition, the accuracy achieved by MobileNet is similar to Xception, while MobileNet occupies lower space. It is noteworthy that ResNet50 has the worst performance on the test set, as overfitting has not been completely eliminated. In summary, DenseNet121 is the best choice to complete this task, and VGG16 and MobileNet are also satisfying alternatives with higher accuracy and lower memory, respectively.

C. VERIFICATION OF TL AND SVM

To verify the effectiveness of the TL strategy, comparative experiments were conducted with pretrained models and models trained from scratch. The test accuracies and training times of five CNNs under two conditions are presented in Fig. 8 and Fig. 9, respectively. It can be clearly seen that the performance of VGG16, ResNet50, and MobileNet was significantly improved by the TL strategy, where the accuracies obtained by the pretrained model are much higher than that of trained from scratch. As the original architectures

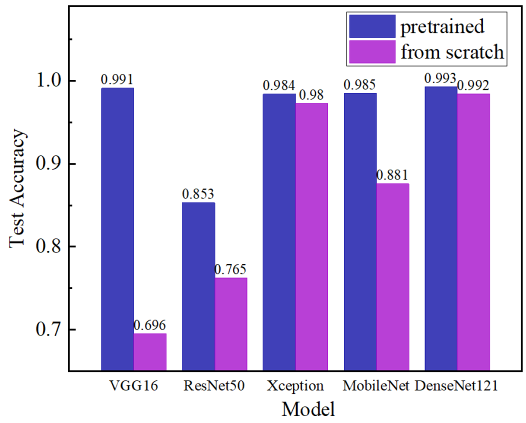


FIGURE 8. Test accuracy comparison of CNNs trained from scratch and using pretrained models.

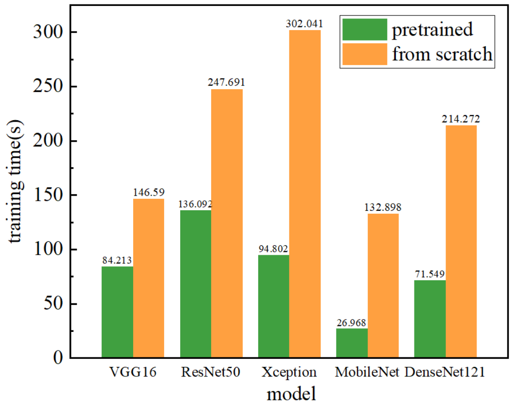


FIGURE 9. Training time comparison of CNNs trained from scratch and using pretrained models.

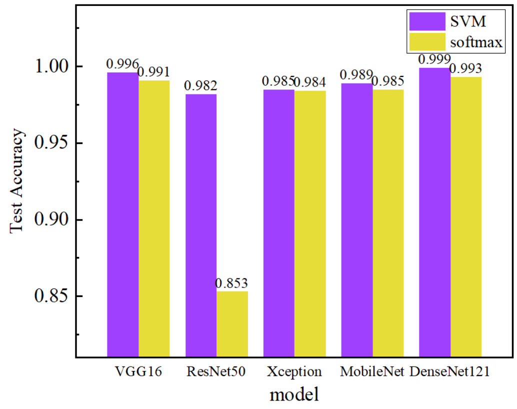


FIGURE 10. Test accuracy comparison of SVMs and fully connected layers with softmax regression.

of Xception and DenseNet121 are sufficiently powerful to realize the task of this section, there is little difference in test accuracies under two conditions. However, the training time of every CNN was drastically reduced when TL was applied. Therefore, it can be concluded that TL improves model training efficiency and accelerates the training process.

The previous experiments all utilized fully connected layers and softmax regression as classifiers to output predicted value, as they easily coordinate with categorical cross-entropy functions to update weights in networks. To further promote the CNNs accuracies on the test set, we attempted to replace fully connected layers by SVMs with linear kernels to implement classification. The results are shown in Fig. 10, where the CNNs combined with SVMs show better performance than softmax classifiers. As the softmax classifiers essentially conform exacted features to the probability distribution, it is not as powerful as SVMs in a multiclassification task. Moreover, the linear kernels used in SVMs alleviate the overfitting phenomenon existing in fully connected layers. That is probably the reason that ResNet50, whose performance is the worst among these models, has been greatly improved when utilizing the SVMs.

D. T-SNE-BASED VISUALIZATION

To verify the feature extraction ability of the proposed method, t-distributed stochastic neighbor embedding (t-SNE) is used to visualize the data distribution of the extracted features [28]. t-SNE is a technology capable of mapping high-dimensional data into a 2-dimensional space map, where the mutual distance of data points is determined by the similarity of samples, as shown in Fig. 11. Samples with different labels are presented in different colors. Fig. 11(a) demonstrates the distribution of 1000 samples of raw data in the test dataset, which has the highest degree of confusion. It is difficult for an algorithm to recognize the correct class of these data. Some data processing methods can extract valid features and filter out distractions, including CWT, FFT, EMD, and WPD, as shown in Fig. 11(b), Fig. 11(c), Fig. 11(d), Fig. 11(e). It can be seen that these methods have the ability to extract features to a certain degree, but they are not completely competent for it, and the distinction of features extracted by these methods is still rather fuzzy. Even though FFT successfully separates different samples from each other, it cannot gather the same samples together. However, the CNN model of the proposed method nearly realized complete feature extraction; for samples, each status is gathered together, and different groups are clearly separated. Therefore, the CNN adopted in this study is proven useful to assist high accuracy classification.

E. RESULTS ANALYSIS AND EVALUATION

For binary classification, 4 situations can be defined according to the combination of the actual class and predicted results, which contain true positive (TP), false positive (FP), true negative (TN) and false negative (FN), as shown in Table 6. It is obvious that the total number of samples is equal to the sum total of TP, FP, TN, and FN.

To analyze the generalization ability of the trained model, several evaluation criteria are essential for performance measure [29], including accuracy rate (Acc), precision (P),

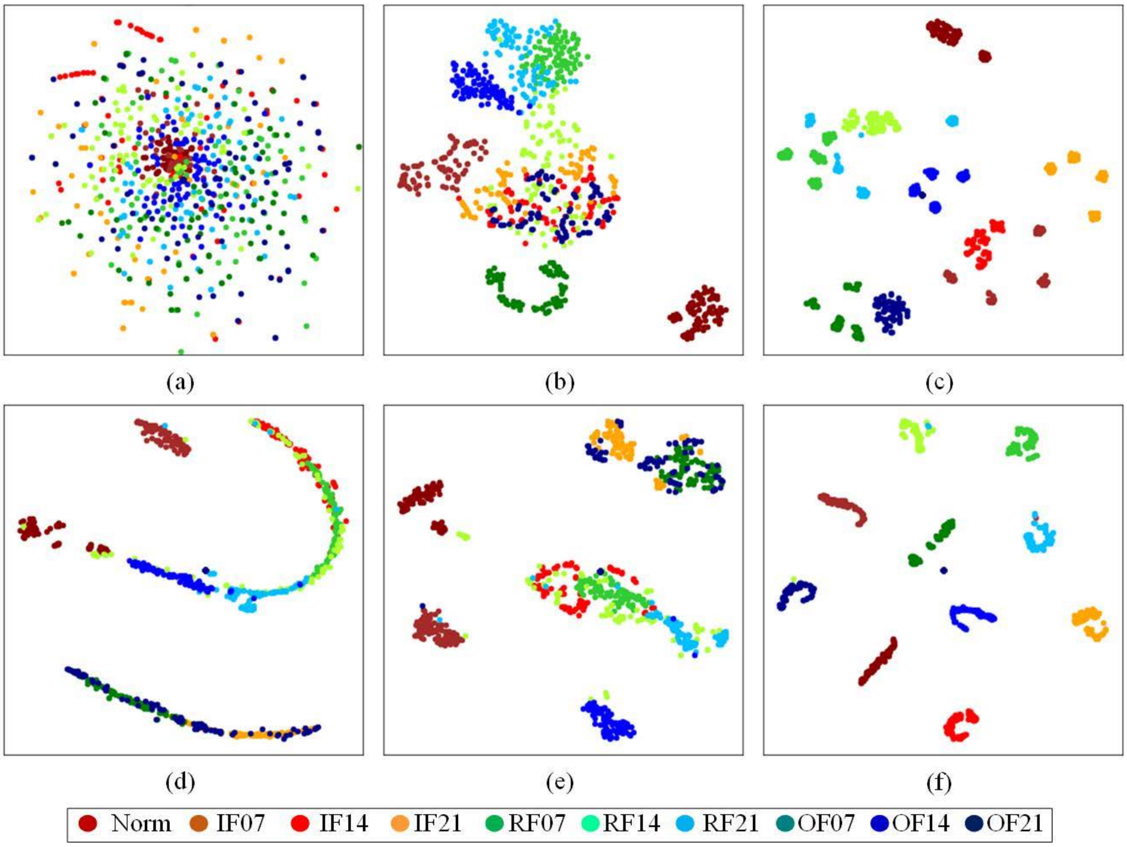

(a) (b) (c) (d) (e) (f)

FIGURE 11. Visualization of data processing based on t-SNE: (a) Raw Signal; (b) CWT; (c) FFT; (d) EMD; (e) WPD; and (f) CNN.

TABLE 6. Confusion matrix for binary classification.

Actual Results	Predicted Results	
	Positive	Negative
Positive	TP (Positive samples are judged as positive)	FN (Positive samples are judged as negative)
Negative	FP (Negative samples are judged as positive)	TN (Negative samples are judged as negative)

recall (R), F1-score (F1) and confusion matrix, which are calculated as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

$$P = \frac{TP}{TP + FP} \tag{15}$$

$$R = \frac{TP}{TP + FN} \tag{16}$$

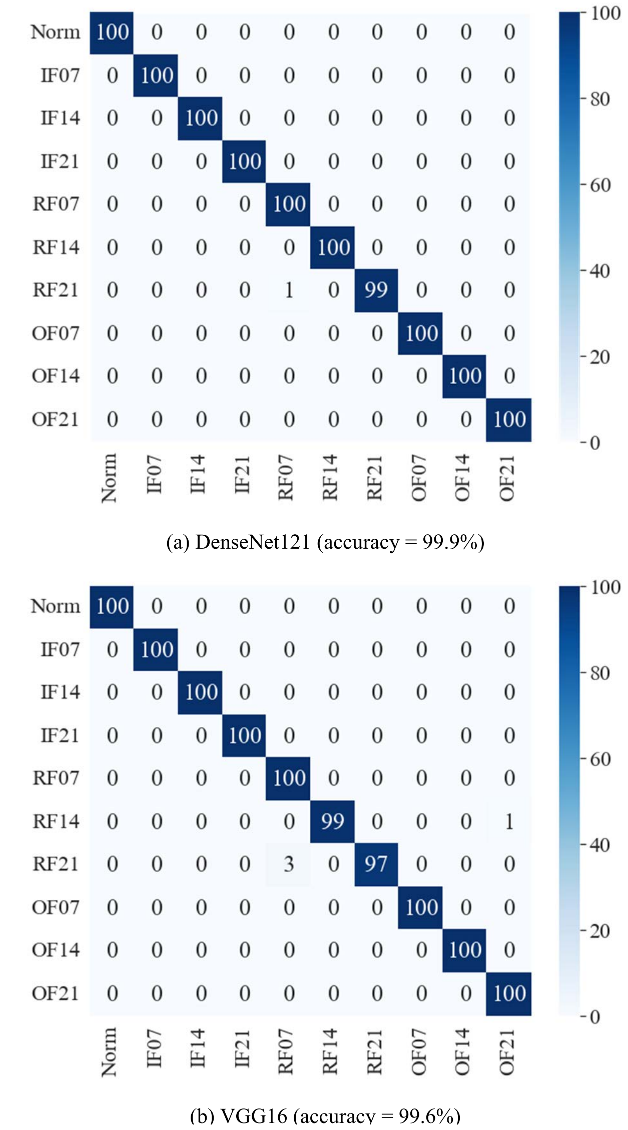
$$F1 = \frac{2 \times P \times R}{P + R} \tag{17}$$

where accuracy represents the proportion of correctly classified samples, precision represents the proportion of truly predicted positive samples in all predicted positive results, recall is the ratio of truly predicted positive samples to all actual positive samples, and F1 indicates the comprehensive consideration of precision and recall. Table 7 lists the P, R, and F1 of our proposed method by applying DenseNet121 based on TL and SVM.

TABLE 7. The evaluation results of the proposed method.

Health Status	Precision	Recall	F1-score	Samples
Normal	1.000	1.000	1.000	100
IF07	1.000	1.000	1.000	100
IF14	1.000	1.000	1.000	100
IF21	1.000	1.000	1.000	100
RF07	0.990	1.000	0.995	100
RF14	1.000	1.000	1.000	100
RF21	1.000	0.990	0.995	100
OF07	1.000	1.000	1.000	100
OF14	1.000	1.000	1.000	100
OF21	1.000	1.000	1.000	100
Average/total	0.999	0.999	0.999	1000

The confusion matrix is useful to intuitively explain the details of fault misjudgment and location for each category in the multiclassification task, where the vertical axis indicates the values of the practical label and the horizontal axis shows the predicted results. The elements on the main diagonal are the number of true judgment samples, while other points present misjudgment numbers. Fig. 12 shows the multiclass confusion matrix of CNN-SVM models by applying DenseNet121 and VGG16 as feature exactors.



	Norm	IF07	IF14	IF21	RF07	RF14	RF21	OF07	OF14	OF21
Norm	100	0	0	0	0	0	0	0	0	0
IF07	0	100	0	0	0	0	0	0	0	0
IF14	0	0	100	0	0	0	0	0	0	0
IF21	0	0	0	100	0	0	0	0	0	0
RF07	0	0	0	0	100	0	0	0	0	0
RF14	0	0	0	0	0	100	0	0	0	0
RF21	0	0	0	0	1	0	99	0	0	0
OF07	0	0	0	0	0	0	0	100	0	0
OF14	0	0	0	0	0	0	0	0	100	0
OF21	0	0	0	0	0	0	0	0	0	100

(a) DenseNet121 (accuracy = 99.9%)

	Norm	IF07	IF14	IF21	RF07	RF14	RF21	OF07	OF14	OF21
Norm	100	0	0	0	0	0	0	0	0	0
IF07	0	100	0	0	0	0	0	0	0	0
IF14	0	0	100	0	0	0	0	0	0	0
IF21	0	0	0	100	0	0	0	0	0	0
RF07	0	0	0	0	100	0	0	0	0	0
RF14	0	0	0	0	0	99	0	0	0	1
RF21	0	0	0	0	3	0	97	0	0	0
OF07	0	0	0	0	0	0	0	100	0	0
OF14	0	0	0	0	0	0	0	0	100	0
OF21	0	0	0	0	0	0	0	0	0	100

(b) VGG16 (accuracy = 99.6%)

FIGURE 12. Confusion matrixes of CNN-SVM models.

Through the confusion matrix, it can be seen that only 1 sample belonging to 'RF21' is misjudged by DenseNet121 as 'RF07', while VGG16 misjudges 3 samples in the same way and 1 'RF14' sample is wrongly judged as 'OF21'. Therefore, conclusions can be drawn that misjudgment easily occurs among different fault severities of rolling element fault. Moreover, DenseNet121 and VGG16 both have excellent generalization performance without misjudgment among normal samples and inner race faults.

In conclusion, our proposed method achieves desired results on CWRU bearing dataset, and the maximum accuracy reached 99.9% when DenseNet121 is used as a feature extractor. In addition, we simply compare our method with other existing algorithms, which are all validated on CWRU dataset, and take accuracy as a performance indicator [30], as shown in Table 8. It can be concluded that our proposed method has achieved current state-of-the-art results.

TABLE 8. Comparison of approaches on CWRU bearing dataset.

Feature extractor	Classifier	Accuracy	Ref.
TSFFCNN	PSO-SVM	97.70%	[28]
MCNN, MSCF	Softmax	94.91%	[31]
1D CNN	Softmax	99.49%	[32]
LMD	SNN	98.95%	[33]
TDFDA	DMRVFLNN	99.42%	[34]
CNN, GN	CSCoh	99.39%	[35]
CWT, CNN	GcForest	99.20%	[36]
AMRXU-VSPMI	SVM	98.65%	[37]
LSTM-GAN-AE	Softmax	99.74%	[38]
CWT, CNN, TL	SVM	99.90%	This study

MCNN = multi-channel CNN, MSCF = multiscale clipping fusion, TSFFCNN = two-stream feature fusion CNN, PSO-SVM = particle smarm optimized-SVM, LMD = local mean decomposition, SNN = spiking neural network, TDFDA = time-domain and frequency-domain analysis, DMRVFLNN = discriminative manifold random vector functional link neural network, GN = group normalization, CSCoh = cyclic spectral coherence, GcForest = deep forest, AMRXU = autoregressive with external uncertainty, VSPMI = proportional multi-integral combined with variable structure-Lyapunov, LSTM = long short-term memory, GAN = generative adversarial network.

V. COMPARATIVE EXPERIMENTS

To further demonstrate the validity and superiority of the proposed method, we conducted three groups of experiments to test the model's performance in the form of accuracy, and compare it with four other mainstream intelligent methods on the same fault dataset. These experiments separately explored the application of three different aspects, including the cross-domain diagnosis of a single point fault, performance tests under simulated composite faults, and experiments on multiple different datasets. The comparative algorithm includes two traditional machine learning methods, pure support vector machine (SVM) and k-nearest neighbor (KNN), and two deep learning algorithms, deep neural network (DNN) and one-dimensional convolutional neural network (1D CNN). Several methods of signal processing are used in combination with these algorithms, including fast Fourier transform (FFT), empirical mode decomposition (EMD), and wavelet packet decomposition (WPT). These methods are introduced simply as follows:

1) EMD-SVM: The vibration signal is decomposed into a series of intrinsic mode functions (IMFs) through EMD, and the first six IMFs are selected to perform data dimension reduction by principal component analysis (PCA). Then, the obtained low dimensional feature vectors are delivered into the SVM classifier as the input data space to search for the optimal separating hyperplane, where the kernel function adopts the radial basis function.
2) WPT-KNN: Wavelet packets decompose the original signal stepwise by using multiple iterations, and the decomposition level of the wavelet packet tree is 3 layers. As a result, 8 terminal nodes of subbands with different frequencies are obtained. Then, we calculate the wavelet packet energy entropy of each subband and

TABLE 9. Working environments of each subdatasets.

Subdataset	A	B	C	D	E	F	G	H
Working load of training data (hp)	0	0	1	1	2	2	3	0-3
Working load of testing data (hp)	0	1	1	2	2	3	3	0-3

integrate them into 8-dimensional vectors as the input data space, which is used to build the KNN classifier. KNN determines the category of the samples according to the category of the nearest samples.

3) FFT-DNN: FFT is implemented to convert the time-domain signal to the frequency-domain first, and then the obtained data are delivered into DNN to update the weights and basis. DNN adopts a classical five-layer structure, in which the unit number of the input layer is equal to the shape of the input vector, and the unit number of the output layer is determined by the number of recognition categories. The unit numbers of the three hidden layers are set liberally in descending order. Finally, the output layer uses softmax regression to output the predicted probability distribution of each category [39].

4) 1D CNN: For this method, no transformation is implemented, and the raw temporal vibration signals are directly subjected to all operations. The architecture of the adopted CNN model is composed of two convolutional layers and two pooling layers. With the increasing of layers, the depth of the feature vector increases while the width decreases [40]. The following two fully connected layers accomplish the classification task and output final results.

A. CROSS-DOMAIN DIAGNOSIS OF SINGLE POINT FAULT

To analyze the model's performance with other intelligent algorithms, the accuracy results, one of the most critical indicators of diagnosis methods, are compared in this section. All fault statuses are single point failures, which means that one element will not appear in more than two kinds of faults. Furthermore, to explore the generalization and stability of the proposed method, we divided the dataset into several subdatasets under various working environments. These subdatasets are expressed as A-H, and the difference among them is the working load of training data and testing data, as shown in Table 9. Especially, for subdatasets B, D, and F, the working loads of the training data and testing data are different and they belong to cross-domain fault diagnosis. The accuracy results of cross-domain diagnosis will more fully explain whether the model adapts to complicated and volatile working situations.

Fig. 13 shows the experimental results of the proposed methods and four other intelligent algorithms. It can be observed that the proposed method generally achieves the best performance. Moreover, CNN and DNN are both deep learning models, while SVM and KNN are shallow architectures, so the former performs better than the latter. It is worth noting that FFT-DNN performs excellently on A, C, E, G, and H; however, it achieves the worst results on cross-domain diagnosis tasks for subdatasets B, D, and F. It can be inferred that FFT makes the samples under different working loads too diverse. In contrast, the algorithms that adopt the CNN architecture can always remain stable on the cross-domain diagnosis. Through the testing accuracy of two traditional algorithms, it can be seen that WPT shows better feature extraction ability than EMD. In other words, deep architectures more easily achieve high accuracy than conventional methods. CNN can better adapt to cross-domain diagnosis, and wavelet analysis performs well in the field of feature extraction. Therefore, the proposed method combined with the above advantages has the best performance.

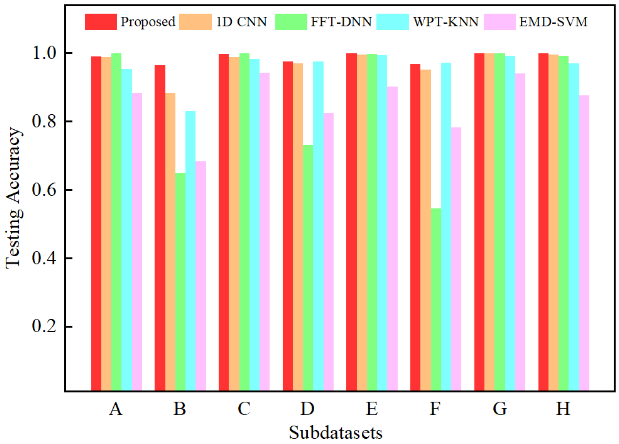


FIGURE 13. The comparison of testing accuracies on several subdatasets.

B. PERFORMANCE TEST FOR SIMULATED COMPOSITE FAULTS

In practical applications, composite faults are more common than single faults. Therefore, it is necessary to evaluate the algorithm's ability to detect compound faults. However, the CWRU bearing dataset contains no type of composite faults, so we construct simulated multiple fault signals through mathematical calculations. For example, samples of inner race fault and outer race fault with the same length are selected, we add them point by point and then determine the arithmetic average, so the samples with inner race fault and outer race fault are obtained. Using this process, we constructed 6 kinds of simulated compound faults, as shown in Table 10. Note that "OF 07 × 2" means that 2 points of failure exist on the bearing's outer race, whose samples are calculated from six o'clock OF07 and three o 'clock OF07.

Fig. 14 indicates the performance of the five algorithms on compound fault recognition based on the F1-score. It can be observed that the proposed method has superior F1-scores compared with the other four algorithms, as the F1-scores of

TABLE 10. The description of simulated composite faults.

Class Label	Running Status	Training Samples	Test Samples
0	Normal	500	100
1	Single point failure	500	100
2	IF07+OF07	500	100
3	IF07+RF07	500	100
4	IF07+RF14	500	100
5	IF07+RF07+OF07	500	100
6	IF07+RF14+OF21	500	100
7	OF07×2	500	100

each running condition of the proposed method reach over 95%, and the average value is nearly 98%. FFT-DBN also has excellent performance based on the F1-score, followed by 1D CNN, WPT-KNN, and EMD-SVM. It is obvious that composite fault diagnosis is more difficult than single point failure, especially for classes 2, 3, 4, 5, 6, the F1-scores of all five algorithms are visibly lower than other conditions. This illustrates that samples under these conditions are less discriminative. However, even with this disadvantage, the proposed method and the FFT-DNN method can still maintain high accuracy, which shows favorable stability and capability.

C. EXPERIMENTS ON MULTIPLE DATASETS

To further explain the good performance of the proposed method in various cases, this section conducts diagnosis experiments on several datasets. These datasets are all vibration signals of roller machinery faults, including bearings, gearboxes, and rotors, and furthermore involve the most common faults in rotating machinery. For CWRU Bearing dataset, we implemented two experiments corresponding to sampling frequencies 12k and 48k. Other datasets, such as the MFPT Fault Datasets, IMS Bearing Datasets, the UPB Datasets, Gear Fault Dataset, MaFaulDa, and the Rotor Fault Dataset, are also used to verify the effectiveness of the algorithm. These datasets are introduced simply as follows:

1) MFPT Fault Datasets: Data were assembled and prepared on behalf of Machinery Failure Prevention Technology (MFPT) by Dr. Eric Bechhoefer, Chief Engineer, NRG Systems. These datasets are comprised of data from a bearing test rig (nominal bearing data, an outer race fault at various loads, and inner race fault and various loads) and three real-world faults [41].
2) IMS Bearing Datasets: These datasets were provided by the Center for Intelligent Maintenance Systems (IMS), University of Cincinnati. Three datasets are included in the data packet. Each dataset describes a test-to-failure experiment and consists of individual files that are 1-second vibration signal snapshots recorded at specific intervals. Each file consists of 20,480 points with the sampling rate set at 20 kHz [42].
3) UPB Datasets: The test platform was developed at the University of Paderborn in Germany. In total, experiments with 32 different bearing damages in ball bearings of type 6203 were performed, including undamaged (healthy) bearings (6x), artificially damaged bearings (12x), and bearings with real damages caused by accelerated lifetime tests (14x) [43].
4) Gear Fault Dataset: Gearbox failure data were shared by Professor Jiong Tang and his team from the University of Connecticut. This dataset is comprised of time-domain gear fault vibration data and gear fault data after angle-frequency domain synchronous analysis. Gear fault types include healthy, missing, crack, spall, and 5 kinds of chips [44].
5) MaFaulDa: The Machinery fault database (MaFaulDa) is composed of 1951 multivariate time-series acquired by sensors on a SpectraQuest Machinery Fault Simulator (MFS) Alignment-Balance-Vibration (ABVT). The dataset is comprised of six different simulated states: normal function, imbalance fault, horizontal misalignment faults, vertical misalignment faults, inner bearing faults, and outer bearing faults [45].
6) Rotor Fault Dataset: These data are denoised signals processed by wavelet thresholding-based denoising. They are represented by a 2-dimensional matrix. The vibration signals belong to the normal rotor, contact-rubbing, unbalance, and misalignment. Each column represents the length of data, 2048, or time, 1 s [46].

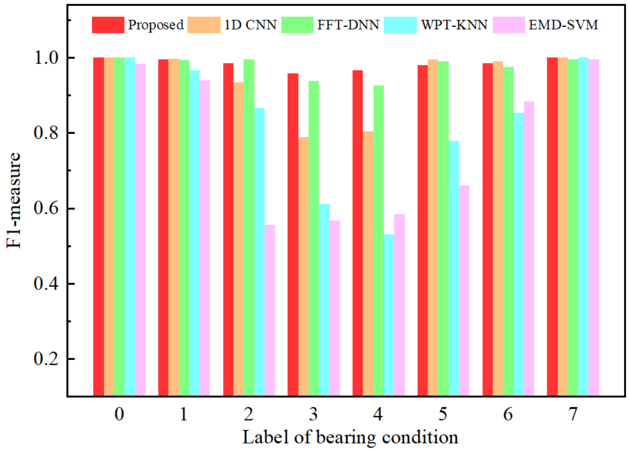


FIGURE 14. The F1-scores of composite faults of different methods.

The results are shown in Fig. 15. It can be observed that three deep learning algorithms perform obviously better than the other two. These algorithms achieved nearly 100% accuracy on several datasets, such as the IMS, MaFaulDa, and Rotor Fault Dataset. For the other datasets, the 1D CNN is slightly inferior but still maintains an accuracy over 90%. Therefore, time-domain analysis is insufficient compared with frequency-domain analysis and time-frequency-domain analysis. The difference in the results on various datasets mainly depends on the performance of the algorithm and the data distribution. For the same algorithm, high accuracy is easily achieved when the distinction among

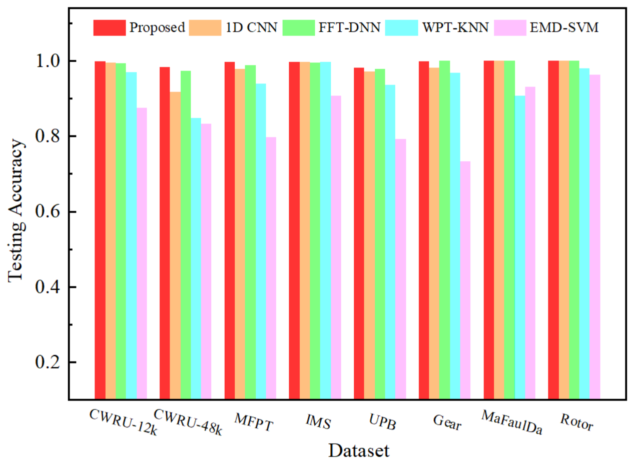


FIGURE 15. The testing accuracies on several datasets of five algorithms.

data is higher, as CWRU-12k is easier to recognize than CWRU-48k. For the same dataset, the algorithm with better performance achieves better results. In conclusion, the proposed method and FFT-DNN are the best performers in this experiment, followed by 1D CNN, WPT-KNN, and EMD-SVM.

VI. CONCLUSION AND FUTURE WORK

In conclusion, we proposed a novel intelligent fault diagnosis method, where raw signals are processed by CWT, a diagnosis model composed of CNN and SVM is used as feature extractor and classifier, respectively, and TL is used as the optimal strategy. From comparative experiments, the validity of TL and SVM has been verified. In addition to testing accuracy, t-SNE visualization, evaluation reports, and confusion matrixes are used to present details of model performance. Furthermore, we made comparisons between the proposed method and other algorithms, implemented them on seven mechanical datasets, and comprehensively analyzed the effectiveness, generalization, and stability of our approach, which increased diagnosis precision at a lower calculation cost. In future work, we will consider expanding the application field of diagnosis algorithms, focusing on more practical cases and more complicated reality, and sequentially improving algorithm performance by incorporating other state-of-the-art research.

REFERENCES

[1] J. Gu, Y. Peng, H. Lu, B. Cao, and G. Chen, "Compound fault diagnosis and identification of hoist spindle device based on Hilbert Huang and energy entropy," *J. Mech. Sci. Technol.*, vol. 35, no. 10, pp. 4281–4290, Oct. 2021.

[2] H. Wang and L. Chan, "Wavelet packet transform-assisted least squares support vector machine for gear wear degree diagnosis," *Math. Problems Eng.*, vol. 2021, pp. 1–9, Sep. 2021.

[3] L. Yang, G. Jia, F. Wei, W. Chang, C. Li, and S. Zhou, "The CIPCA-BPNN failure prediction method based on interval data compression and dimension reduction," *Appl. Sci.*, vol. 11, no. 8, p. 3448, Apr. 2021.

[4] X. Li, H. Shao, S. Lu, J. Xiang, and B. Cai, "Highly efficient fault diagnosis of rotating machinery under time-varying speeds using LSISMM and small infrared thermal images," *IEEE Trans. Syst., Man, Cybern. Syst.*, early access, Mar. 14, 2022, doi: 10.1109/TSMC.2022.3151185.

[5] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikainen, "Deep learning for generic object detection: A survey," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 261–318, Jan. 2020.

[6] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding, "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation," *Med. Image Anal.*, vol. 63, Jul. 2020, Art. no. 101693.

[7] D. W. Otter, J. R. Medina, and J. K. Kalita, "A survey of the usages of deep learning for natural language processing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 604–624, Feb. 2021.

[8] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106587.

[9] P. Li, D. Wang, L. Wang, and H. Lu, "Deep visual tracking: Review and experimental comparison," *Pattern Recognit.*, vol. 76, pp. 323–338, Apr. 2018.

[10] Z. Huang, Y. Shen, J. Li, M. Fey, and C. Brecher, "A survey on AI-driven digital twins in industry 4.0: Smart manufacturing and advanced robotics," *Sensors*, vol. 21, no. 19, p. 6340, Sep. 2021.

[11] J. Ma, C. Li, and G. Zhang, "Rolling bearing fault diagnosis based on deep learning and autoencoder information fusion," *Symmetry*, vol. 14, no. 1, p. 13, Dec. 2021.

[12] T. Han, C. Liu, L. Wu, S. Sarkar, and D. Jiang, "An adaptive spatiotemporal feature learning approach for fault diagnosis in complex systems," *Mech. Syst. Signal Process.*, vol. 117, pp. 170–187, Feb. 2019.

[13] G.-B. Jang and S.-B. Cho, "Feature space transformation for fault diagnosis of rotating machinery under different working conditions," *Sensors*, vol. 21, no. 4, p. 1417, Feb. 2021.

[14] B. Zhao, Z. Niu, Q. Liang, Y. Xin, T. Qian, W. Tang, and Q. Wu, "Signal-to-signal translation for fault diagnosis of bearings and gears with few fault samples," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.

[15] H. Cao, H. Shao, X. Zhong, Q. Deng, X. Yang, and J. Xuan, "Unsupervised domain-share CNN for machine fault transfer diagnosis from steady speeds to time-varying speeds," *J. Manuf. Syst.*, vol. 62, pp. 186–198, Jan. 2022.

[16] 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.

[17] D. Zhang and T. Zhou, "Deep convolutional neural network using transfer learning for fault diagnosis," *IEEE Access*, vol. 9, pp. 43889–43897, 2021.

[18] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 1, pp. 136–144, Jan. 2019.

[19] Z. He, H. Shao, X. Zhong, and X. Zhao, "Ensemble transfer CNNs driven by multi-channel signals for fault diagnosis of rotating machinery cross working conditions," *Knowl.-Based Syst.*, vol. 207, Nov. 2020, Art. no. 106396.

[20] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: A review with applications," *Signal Process.*, vol. 96, pp. 1–15, Mar. 2014.

[21] Y. Tian, "Artificial intelligence image recognition method based on convolutional neural network algorithm," *IEEE Access*, vol. 8, pp. 125731–125744, 2020.

[22] *Case Western Reserve University Bearing Data Center.* Accessed: Oct. 25, 2021. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/home

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

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[25] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.

[26] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[27] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Apr. 2017, pp. 4700–4708.

[28] W. F. Gong, H. Chen, Z. H. Zhang, M. L. Zhang, R. H. Wang, C. Guan, and Q. Wang, "A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-SVM and multichannel data fusion," *Sensors*, vol. 19, no. 7, p. 1693, Apr. 2019.

[29] X. Wei and D. Söffker, "Comparison of CWRU dataset-based diagnosis approaches: Review of best approaches and results," in *Proc. Eur. Workshop Struct. Health Monitor. (EWSHM)*, vol. 127, P. Rizzo and A. Milazzo, Eds. Cham, Switzerland: Springer, 2020, pp. 525–532, doi: 10.1007/978-3-030-64594-6.

[30] R. Bai, Q. Xu, Z. Meng, L. Cao, K. Xing, and F. Fan, "Rolling bearing fault diagnosis based on multi-channel convolution neural network and multi-scale clipping fusion data augmentation," *Measurement*, vol. 184, Nov. 2021, Art. no. 109885.

[31] F. Xue, W. Zhang, F. Xue, D. Li, S. Xie, and J. Fleischer, "A novel intelligent fault diagnosis method of rolling bearing based on two-stream feature fusion convolutional neural network," *Measurement*, vol. 176, May 2021, Art. no. 109226.

[32] D. Neupane, Y. Kim, J. Seok, and J. Hong, "CNN-based fault detection for smart manufacturing," *Appl. Sci.*, vol. 11, no. 24, p. 11732, Dec. 2021.

[33] L. Zuo, L. Zhang, Z.-H. Zhang, X.-L. Luo, and Y. Liu, "A spiking neural network-based approach to bearing fault diagnosis," *J. Manuf. Syst.*, vol. 61, pp. 714–724, Oct. 2021.

[34] X. Li, Y. Yang, N. Hu, Z. Cheng, and J. Cheng, "Discriminative manifold random vector functional link neural network for rolling bearing fault diagnosis," *Knowl.-Based Syst.*, vol. 211, Jan. 2021, Art. no. 106507.

[35] Z. Chen, A. Mauricio, W. Li, and K. Gryllias, "A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks," *Mech. Syst. Signal Process.*, vol. 140, Jun. 2020, Art. no. 106683.

[36] Y. Xu, Z. Li, S. Wang, W. Li, T. Sarkodie-Gyan, and S. Feng, "A hybrid deep-learning model for fault diagnosis of rolling bearings," *Measurement*, vol. 169, Feb. 2021, Art. no. 108502.

[37] S. TayebiHaghighi and I. Koo, "SVM-based bearing anomaly identification with self-tuning network-fuzzy robust proportional multi integral and smart autoregressive model," *Appl. Sci.*, vol. 11, no. 6, p. 2784, Mar. 2021.

[38] H. Liu, H. Zhao, J. Wang, S. Yuan, and W. Feng, "LSTM-GAN-AE: A promising approach for fault diagnosis in machine health monitoring," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–13, 2022.

[39] F. Jia, Y. G. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 303–315, May 2016.

[40] W. Zhang, G. L. Peng, and C. H. Li, "Rolling element bearings fault intelligent diagnosis based on convolutional neural networks using raw sensing signal," *Smart Innov. Syst. Tec.*, vol. 64, pp. 77–84, Nov. 2017.

[41] D. Lee, V. Siu, R. Cruz, and C. Yetman, "Convolutional neural net and bearing fault analysis," in *Proc. Int. Conf. Data Mining (DMIN)*, 2016, pp. 194–200.

[42] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *J. Sound Vibrat.*, vol. 289, nos. 4–5, pp. 1066–1090, 2006.

[43] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, "Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification," in *Proc. Eur. Conf. Prognostics Health Manage. Soc. (PHM)*, 2016, pp. 1–17.

[44] P. Cao, S. Zhang, and J. Tang, "Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning," *IEEE Access*, vol. 6, pp. 26241–26253, 2018.

[45] M. A. Marins, F. M. L. Ribeiro, S. L. Netto, and E. A. B. da Silva, "Improved similarity-based modeling for the classification of rotating-machine failures," *J. Franklin Inst.*, vol. 355, no. 4, pp. 1913–1930, Mar. 2018.

[46] D. Liu, Z. Xiao, X. Hu, C. Zhang, and O. P. Malik, "Feature extraction of rotor fault based on EEMD and curve code," *Measurement*, vol. 135, pp. 424–712, Mar. 2019.

WENTAO ZHANG was born in 1997. He received the B.S. degree from Tongji University, in 2019. He is currently pursuing the M.S. degree with the Shanghai University of Engineering Science. His research interests include machine learning, fault diagnosis, mechanical design, and vibration control.

TING ZHANG received the Ph.D. degree in mechanical engineering from Shanghai Jiao Tong University, in 2014. She is currently an Associate Professor of mechanical and automotive engineering with the Shanghai University of Engineering Science. Her research interests include mechanical vibration control and vibration signal sensing.

GUOHUA CUI was born in 1975. He received the B.S. degree in lifting transportation and engineering machinery from the North China University of Water Resources and Electric Power, in 1997, and the M.S. degree in mechanical engineering and the Ph.D. degree in mechanical design and theory from Jilin University, in 2005 and 2009, respectively. He is currently a Professor, a Doctoral Advisor, and the Assistant Dean of the School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science. His research interests include robot mechanics, multirobot cooperative control, and industrial robot fault diagnosis and health assessment.

YING PAN received the B.S. degree in structural engineering and the M.S. degree in thermal power engineering from Northeast Electric Power University, in 1998 and 2001, respectively, and the Ph.D. degree in mechanical professional from Xi'an Jiaotong University, in 2004. She is currently an Associate Professor of mechanical and automotive engineering with the Shanghai University of Engineering Science. Her research interests include mechanical structure vibration analysis and vibration control.