Review Article

Application of Rotating Machinery Fault Diagnosis Based on Deep Learning

Wei Cui [iD],[1] Guoying Meng,[1] Aiming Wang,[1] Xinge Zhang,[1] and Jun Ding[2]

[1]*School of Mechanical Electronic and Information Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China*
[2]*School of Mechanical and Electrical Engineering, North China Institute of Science and Technology, Yanjiao 065201, China*

Correspondence should be addressed to Wei Cui; cwcwei@163.com

With the continuous progress of modern industry, rotating machinery is gradually developing toward complexity and intelligence. The fault diagnosis technology of rotating machinery is one of the key means to ensure the normal operation of equipment and safe production, which has very important significance. Deep learning is a useful tool for analyzing and processing big data, which has been widely used in various fields. After a brief review of early fault diagnosis methods, this paper focuses on the method models that are widely used in deep learning: deep belief networks (DBN), autoencoders (AE), convolutional neural networks (CNN), recurrent neural networks (RNN), generative adversarial networks (GAN), and transfer learning methods are summarized from the two aspects of principle and application in the field of fault diagnosis of rotating machinery. Then, the commonly used evaluation indicators used to evaluate the performance of rotating machinery fault diagnosis methods are summarized. Finally, according to the current research status in the field of rotating machinery fault diagnosis, the current problems and possible future development and research trends are discussed.

1. Introduction

Complex electromechanical equipment is an important basis for the development of modern industries such as coal, transportation, aviation, and construction. With the continuous progress of modern industry, rotating machinery, as an important part of complex electromechanical equipment, is developing toward complexity, large scale, and intelligence [1]. Rotating machinery usually works for long periods of time under heavy loads and high speeds. Bearing, gearbox, and other key components are likely to suffer from wear, deformation, fracture, and other faults during the inherent degradation process and under time-varying operating conditions. The failure of these components will affect the normal operation and use of the equipment. Severe cases can result in downtime or damage, and they even cause casualties and huge economic losses. Therefore, research on condition monitoring and fault diagnosis of rotating machinery is one of the key means to ensure the normal operation of equipment, reduce unplanned maintenance, avoid

catastrophic failure, and ensure the safety of industrial production, which is of great significance [2, 3]. The main tasks of fault diagnosis research on rotating machinery include determining the operating state, judging the fault type, and predicting the fault trend [4, 5]. In recent years, researchers have done a lot of research work in the field of rotating machinery fault diagnosis, which has achieved fruitful research results and has been applied in actual working conditions, as shown in Figure 1.

With the rapid development of intelligent algorithms, the current research on deep learning in the fault diagnosis of rotating machinery is increasing year by year, and it has also attracted the attention of more and more researchers. In order to better promote the research progress in this field, this article reviews the related research from the two aspects of deep learning theory and its application in fault diagnosis. This research can provide convenience and inspiration for related researchers and provide reference for understanding and promoting the development of fault diagnosis research.
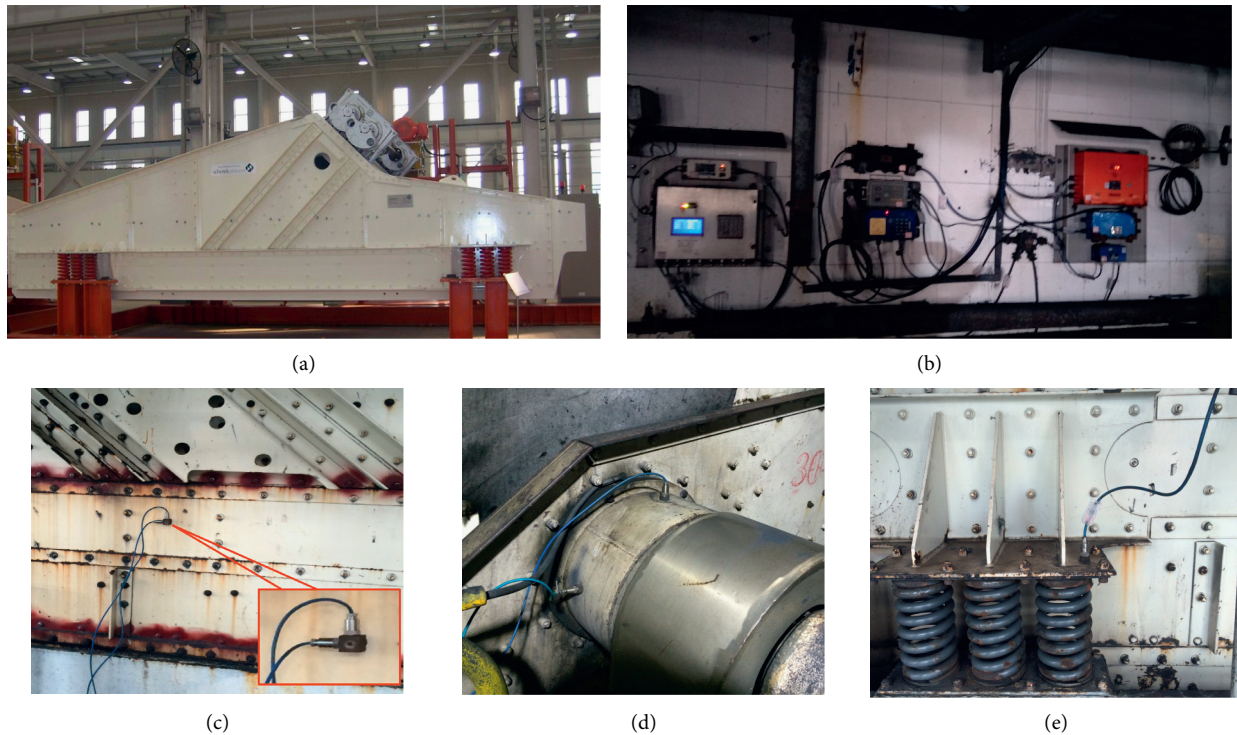
(a)

(b)

(c)

(d)

(e)

FIGURE 1: Practical application of fault diagnosis research on mine vibrating screen: (a) mining vibrating screen; (b) signal collection and transmission equipment; (c) vibration signal collection of the side plate of the vibrating screen; (d) vibration signal collection of the vibrating screen exciter; (e) vibration signal collection of the vibrating screen spring and support device.

This paper first introduces several development stages of the fault diagnosis method for rotating machinery and explains the diagnosis process, advantages, and disadvantages of each stage method. Then, several mainstream deep learning models and transfer learning methods in the field of rotating machinery fault diagnosis are reviewed from both theoretical and application aspects. Finally, the challenges faced by deep learning in the field of rotating machinery fault diagnosis and the possible research trends in the future are discussed.

2. Development Process of Rotating Machinery Fault Diagnosis

Rotating machinery will generate different physical signals when it is running. These signals can be used as characteristic signals to characterize its operating status during fault diagnosis research. According to different signal types and acquisition methods, common analysis methods include the following categories: vibration signal analysis [6], thermal imaging analysis [7], acoustic signal analysis [8], temperature signal analysis, and electrical signal analysis. As shown in Table 1, the above methods have their own advantages and disadvantages due to factors such as operating environment and sensor installation. Because vibration signal analysis can quickly respond to changes in the state of rotating machinery, the relevant signal processing methods are diverse and the diagnostic accuracy is relatively high, so vibration signal analysis is the most widely used in these methods.

The early traditional fault diagnosis of rotating machinery was realized based on simple signal processing technology [9, 10]. As shown in Figure 2, the traditional fault diagnosis method first collects the vibration, temperature, voltage, and current signals of the rotating machinery through sensors. After processing, the characteristic parameters that characterize the operating conditions of the equipment are obtained. The characteristic parameters of normal signals and fault signals are compared and analyzed, and an appropriate threshold is selected. When the characteristic parameters of the collected signals exceed the set threshold, it is determined that the device is faulty. However, in actual working conditions, rotating machinery operates in a harsh environment, and its signals are often characterized by nonstationarity and nonlinearity and contain a large amount of noise [11]. Therefore, the accuracy of the traditional fault diagnosis method which only relies on the characteristic parameters is low.

With the rise of artificial intelligence, machine learning (ML) as its core has gained widespread attention. The researchers combined the machine learning method with the feature extraction method based on signal processing to perform fault diagnosis of rotating machinery. Feature extraction methods based on signal processing are used for feature extraction and selection of equipment monitoring signals. Commonly used features include time domain features, frequency domain features, and time-frequency domain features [1, 12]. The ML method obtains the diagnostic model with high generalization through the training of input features and establishes the relationship

TABLE 1: Comparison of advantages and disadvantages of different analysis methods.

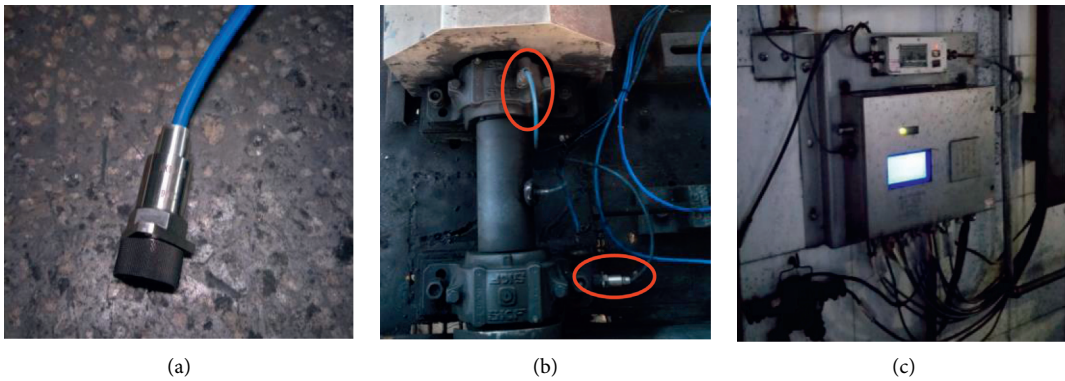| Analytical method	Advantages	Disadvantages
Vibration signal analysis	(i) Ability to quickly respond to changes in the state of rotating machinery (ii) Being easy to measure and sensitive to most mechanical failures (iii) Fast speed, high precision, low cost (iv) Various signal processing methods (v) Ability to realize online monitoring, low technical threshold (vi) Being more suitable for industrial production field equipment diagnosis	(i) It is difficult to locate equipment faults (ii) Sensor installation location is not easy to choose (iii) Noise signal interference is easy (iv) The propagation path of the vibration signal is complicated, and the vibration signal at the source of the fault is not easy to collect
Thermal imaging analysis	(i) Ability to detect electrical faults (ii) Ability to achieve nondestructive testing (iii) Strong recognition result against interference (iv) Lower hardware cost (v) Clear and intuitive test results	(i) The scope of application is limited, only suitable for mechanical equipment that will produce high temperature on the surface when a failure occurs (ii) Thermal image processing speed is slower, and software cost is higher (iii) It is not easy to choose the installation location and distance of the collection equipment (iv) It is vulnerable to the environment
Acoustic signal analysis	(i) Ability to achieve nondestructive testing (ii) Short analysis time and high recognition efficiency (iii) Low cost and fast speed (iv) Higher sampling rate of signal acquisition	(i) The quality of the acoustic signal is determined by the acquisition equipment, which has higher requirements for the equipment (ii) It is vulnerable to environmental conditions (iii) The installation distance of the collection equipment is not easy to determine (iv) It is susceptible to noise (v) It is vulnerable to the parameters of recording software
Temperature signal analysis	(i) Easiness of operation (ii) Low cost (iii) High detection accuracy	(i) It is difficult to detect early failures of equipment (ii) It is vulnerable to the environment (iii) Installation location of collection equipment is not easy to choose
Electrical signal analysis	(i) Low cost (ii) Easiness of operation (iii) Intuitive test results	The scope of application is limited to electrical faults

(a)	(b)	(c)

FIGURE 2: Equipment signal acquisition: (a) vibration sensor; (b) sensor installation location; (c)data acquisition device.

between the selected features and the health state of rotating mechanical equipment through the trained diagnostic model [13]. The commonly used ML methods include BP neural network, RBF neural network, extreme learning machine, and support vector machine (SVM) [2]. Although the fault diagnosis method of rotating machinery based on signal processing and machine learning has achieved some achievements, it still has two major defects [14, 15]. First of all, in the process of feature extraction, a large number of signal processing technologies and rich engineering experience are required to extract and select appropriate fault features. Secondly, the ML method used belongs to the shallow model [16], and the model parameters will increase with the increase of input data. It will lead to the decrease of

the training efficiency and generalization performance of the model, resulting in overfitting, and reduce the diagnostic accuracy of the diagnosis model.

With the continuous development of computer technology, sensor technology, and communication technology [17], the amount of monitoring data of key components of rotating machinery such as bearings and gearboxes increases day by day, pushing the field of fault diagnosis into the era of "big data" [5]. In the field of ML, Hinton's paper on data dimension reduction of deep neural networks [18] marked the beginning of deep learning research. Deep learning (DL) methods automatically learn features from the original input and process them by building an end-to-end diagnostic mode [19], directly establishing a connection between the growing monitoring data and the health status of the machine [1], and they solve the problem of heavy workload and high cost when selecting features from a large amount of monitoring data. Therefore, the application of deep learning methods in the fault diagnosis of rotating machinery has great significance. Figure 3 shows the implementation process of the above three-stage diagnosis method.

Rotating machinery fault diagnosis methods based on deep learning can often achieve good results in a laboratory environment, because there are enough labeled data to train the diagnosis model at this time, but this is not the case in actual working conditions [20]. In actual working conditions, rotating machinery usually undergoes a long degradation process from a healthy state to failure; when a failure occurs, it will be repaired in time, and it is time-consuming and labor-intensive to obtain failure data [21]. As a result, in the collected data, the amount of normal data is far more than the fault data, and the large difference in the number of sample types will cause serious data imbalance. Moreover, it takes a huge cost to mark a large amount of monitoring data for health categories. Therefore, most of the monitoring data collected in actual working conditions are not marked. The above reasons lead to the unsatisfactory application effect of the fault diagnosis method based on deep learning in actual working conditions.

Transfer learning (TL) is a new machine learning method that is closely related to deep learning and can use existing knowledge to solve problems in other different but related fields [22]. Therefore, DL and TL can be used as a bridge to use the rich data resources of the laboratory environment to diagnose the fault of rotating machinery under actual working conditions, and solve the problem that the diagnosis model has low accuracy in identifying the health status of the equipment due to the scarcity of fault data in practical applications.

Although the research methods used by researchers for fault diagnosis of rotating machinery are different, most of the methods follow the same process, as shown in Figure 4. In the research process, first collect the signal from the target diagnostic equipment to obtain characteristic signals that can characterize the operating status of the equipment. Then, use different research methods to establish diagnostic models to process and analyze the characteristic signals. Finally, diagnose and make decisions on the health of the target device, so as to ensure the normal operation of the target device.

3. Deep Learning-Based Fault Diagnosis Method for Rotating Machinery and Its Application

The mainstream deep learning models in the field of rotating machinery fault diagnosis include deep belief networks (DBN), autoencoders (AE) and their variants, convolutional neural networks (CNN), recurrent neural networks (RNN), and generative adversarial networks (GAN).

3.1. Deep Belief Networks and Their Application

3.1.1. Basic Theory of DBN. The deep belief network (DBN) is composed of a stack of multiple restricted Boltzmann machines (RBM) [23]. RBM is a probabilistic generative model, and its structure is shown in Figure 5.

RBM consists of a visible layer v and a hidden layer h. The visible layer represents the input sample, and the hidden layer is equivalent to the feature extractor. The visible layer and the hidden layer are connected by a weight matrix w, and there is a connection between each neuron node in the layer, but there is no connection in the layer. The energy function of RBM is as follows:

$$E(v, h) = -\sum_{i=1}^{n} a_i v_i - \sum_{j=1}^{m} b_j h_j - \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} v_i h_j. \quad (1)$$

In the formula, n and m are the numbers of visible layer and hidden layer units, respectively; a_j and b_j are the neuron bias of the visible layer and the hidden layer, respectively. w_{ij} is the connection weight between neurons. The energy function is used to define the joint probability distribution of the nodes in the explicit layer and the nodes in the hidden layer, which can be expressed as the following formula, where $Z(\theta)$ is the normalization factor.

$$P(v \mid \theta) = \frac{1}{Z(\theta)} \sum_{h} \exp[-E(v, h, \theta)],$$
$$Z(\theta) = \sum_{v,h} \exp[-E(v, h, \theta)]. \quad (2)$$

The activation conditions of the visible unit and the hidden unit are defined as follows:

$$P(v_i = 1 \mid h) = \sigma_s\left(b_i + \sum_{j=1}^{m} \omega_{i,j} \cdot h_j\right),$$
$$P(h_i = 1 \mid v) = \sigma_s\left(a_i + \sum_{j=1}^{n} \omega_{i,j} \cdot v_j\right), \quad (3)$$

where σ_s is the activation function and Sigmoid function is usually used. Then, use the maximum likelihood estimation method to get the parameters of RBM, expressed as

$$\hat{\theta} = \arg\max_{\theta} \ln[P(\theta \mid x_1, x_2, \ldots, x_k)] = \frac{1}{k} \sum_{i=1}^{k} \ln[P(x_i \mid \theta)], \quad (4)$$
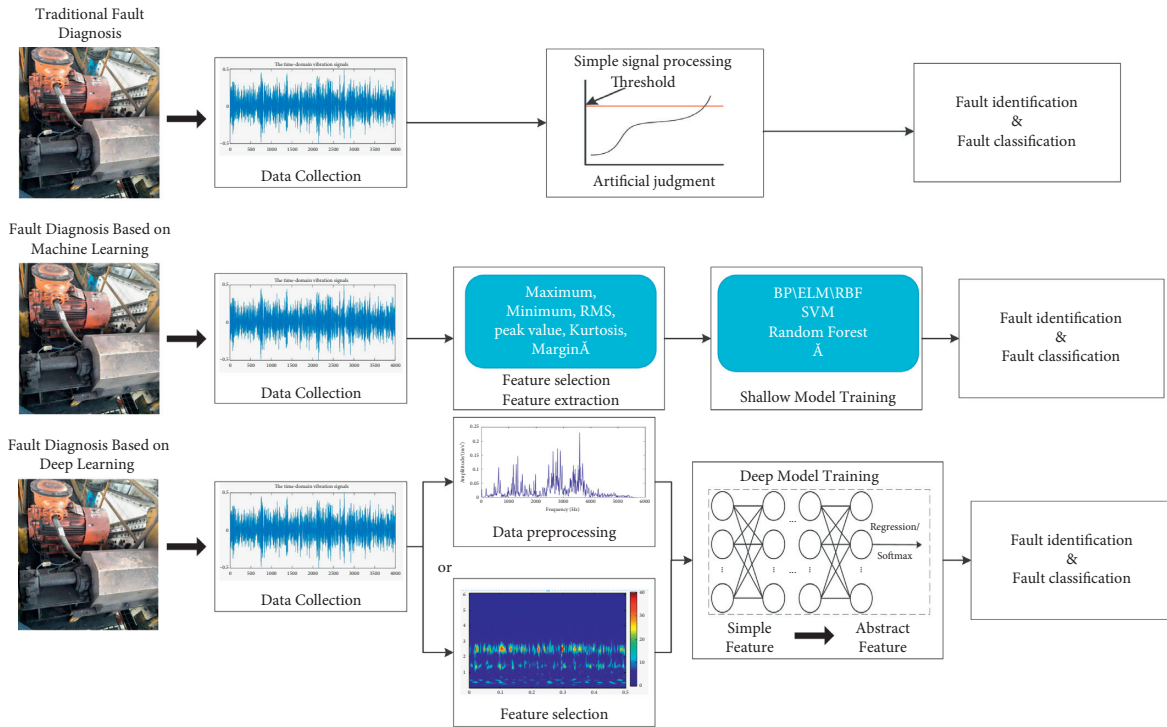
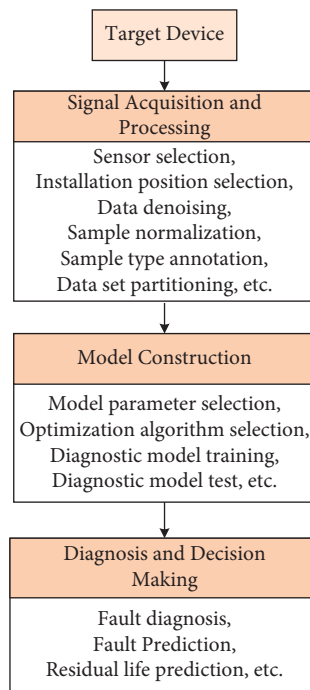Figure 3: Development process of rotating machinery fault diagnosis.

Figure 4: General process of fault diagnosis research.

where $\{x_i\}_{i=1}^{k}$ is the input data set with k samples.

In the DBN, each RBM is connected by weights, and each layer of the RBM network is independent of the others. Pretraining is performed through forward learning; that is, the output of the previous layer of RBM is used as the input of the next layer, and the features are mapped and passed layer by layer. RBM of each layer makes the weight of the layer reach the optimal mapping of the feature vector of the layer, so that a more abstract and more representational feature representation can be
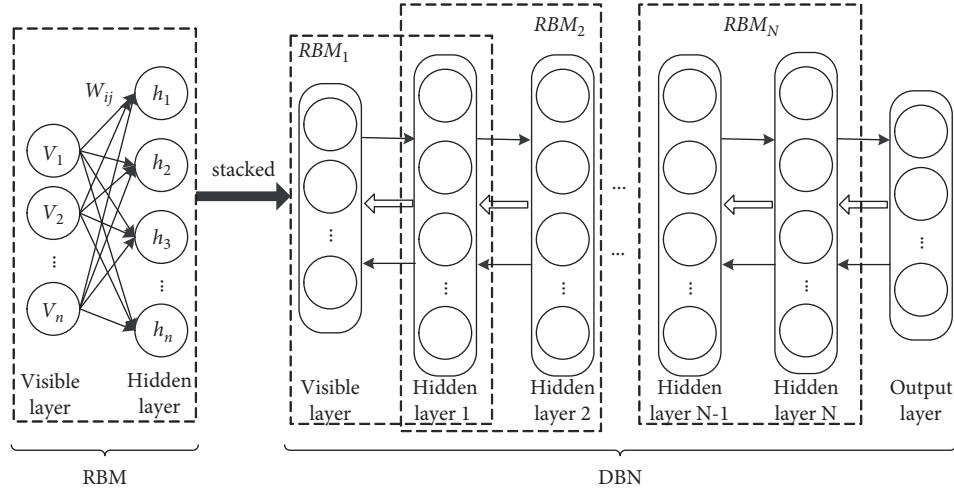
FIGURE 5: RBM and DBN structure diagram.

formed at the high level. Finally, the backpropagation algorithm is used at the high level to propagate the error from top to bottom to each RBM, so as to realize the supervised fine-tuning of the entire DBN.

The core idea of the DBN is to optimize the network parameter values using a layer-by-layer greedy algorithm and use two training methods, interlayer pretraining and reverse fine-tuning, to extract the distributed features of the input data. Pretraining adopts an unsupervised training method, which uses a large number of unlabeled samples to minimize the reconstruction error between layers, can map data from input to output, and construct complex nonlinear functions to characterize features. Reverse fine-tuning makes full use of a small number of labeled samples supervised to achieve accurate classification of DBN networks.

3.1.2. Application of DBN in Fault Diagnosis. Researchers have carried out a lot of research on the application of DBN. Jiang et al. [24] and Han et al. [25] constructed a DBN-based diagnostic model by stacking multiple RBM and achieved higher diagnostic accuracy than traditional methods. In [26, 27], the authors use the frequency domain data after fast Fourier transform (FFT) as the input of the DBN model to diagnose the induction motor fault. Tao et al. used the Teager Energy Operator (TEO) to extract the instantaneous energy in the rolling bearing vibration signal and construct the corresponding feature vector, and then they combined the DBN to diagnose the rolling bearing fault [28]. In [29], a novel intelligent ball screw degradation recognition method based on deep belief networks (DBN) and multisensor data fusion is proposed. First, the derived method calculates frequency spectrums of raw signals, and the fused frequency spectrums are calculated by the multisensor data fusion. Then, a deep learning-based recognition model that can estimate the degradation condition of ball screw automatically is established with the fused data set. The flow chart of this method is shown in Figure 6.

Zhang et al. proposed a semisupervised fault recognition model based on Laplacian feature mapping (LE) and DBN to identify the failure mode of mechanical equipment [30]. In

[31], DBN is used for automatic diagnosis of high-speed train on-board equipment, and its diagnosis performance is better than KNN and ANN. Chen et al. used the feature self-extraction capability of DBN to extract the characteristics of the vibration signal of the gear transmission system and then perform fault identification [32]. Some researchers used DBN to build diagnostic models to diagnose faults in hydraulic equipment [33], wind turbines [34], and compressors [35] and achieved higher diagnostic accuracy than that of traditional methods. Oh et al. preprocessed the vibration signal to generate a 2D image, then applied the directional gradient histogram as the input feature, and performed feature extraction and fault classification through DBN [36]. Tao et al. proposed a DBN-based multisensor information fusion bearing fault diagnosis model. The input of the model is 14 time domain statistical features of the vibration signals collected by 3 sensors [37]. The flow diagram of multisensor information fusion is shown in Figure 7.

In order to speed up the DBN model training speed and improve its generalization ability, operating efficiency, and recognition accuracy, the researchers further studied the optimization algorithm based on the DBN model, as shown in Table 2. Li et al. proposed a deep belief network (DBN) algorithm and bearing fault diagnosis model based on particle swarm optimization (PSO) [38]. In [39–41], the researchers used Nesterov momentum to adaptively optimize the training of the diagnosis model based on DBN, and the diagnosis accuracy rate was higher than that of the standard DBN. He et al. used genetic algorithm to optimize the structure of the DBN and diagnose the fault of the gear transmission chain [42]. Tao et al. proposed a rolling bearing fault diagnosis method based on bacterial foraging decision and deep confidence network, which improved the accuracy of rolling bearing fault diagnosis [43]. Shao et al. constructed an adaptive DBN, used adaptive learning rate and momentum algorithm to train the DBN model [44], and further proposed a convolutional DBN algorithm for bearing fault diagnosis, which used exponential moving average technology to improve the performance of the diagnostic model [45, 46].
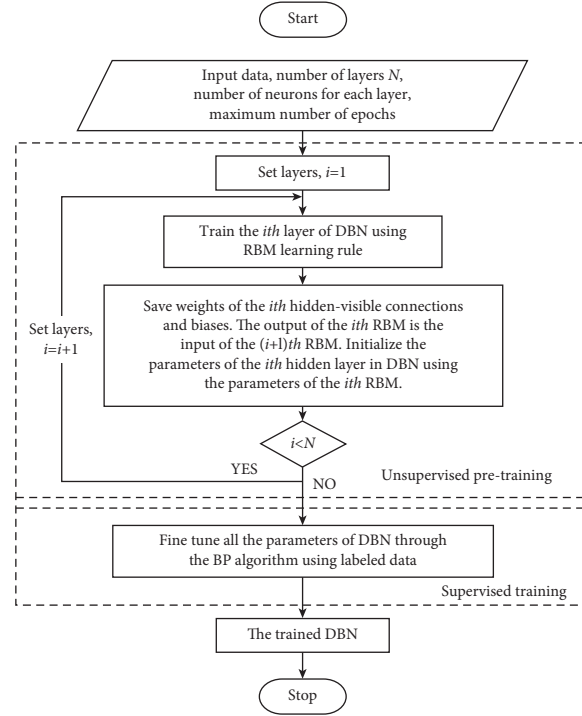
FIGURE 6: The flow chart of the diagnostic method for ball screw proposed in [29].

3.2. Autoencoder and Its Application

3.2.1. Basic Theory of AE and Its Variants

(1) Autoencoder (AE). Autoencoder (AE) is a feedforward neural network with input layer, hidden layer, and output layer. The input layer and the hidden layer form a coding network, and the hidden layer and the output layer form a decoding network. The structure is shown in Figure 8.

The basic idea of AE is to transform the input high-dimensional data into a low-dimensional encode vector through nonlinear mapping through the coding network, then reconstruct the code vector through the decoding network, and then learn new data representation [47].

Given a data set $\{x^m\}_{m=1}^M$, the encoding network uses the encoding function f_θ to transform the training sample x^m into a low-dimensional coding vector h^m.

$$h^m = f_\theta(x^m) = s_f(Wx^m + b), \qquad (5)$$

where s_f is the activation function of the coding network; θ is the parameter set of the coding network, $\theta = \{W, b\}$; W is the weight matrix from the input layer to the hidden layer; and b is the bias term coefficient. Then, the encoding vector h^m is reconstructed by the decoding function g_{θ_l} in the decoding network to obtain the reconstructed representation $\widehat{x}^m$ of the sample x^m.

$$\widehat{x}^m = g_{\theta'}(h^m) = s_g(W'h^m + d), \qquad (6)$$

where s_g is the activation function of the decoding network; θ' is the parameter set of the decoding network, $\theta' = \{W', d\}$; W' is the weight matrix from the hidden layer to the output layer; and d is the bias term coefficient. The autoencoder minimizes the reconstruction error between x^m and $\widehat{x}^m$ by optimizing the parameter set $\{\theta, \theta'\}$.

$$J_{AE}(\theta, \theta') = \frac{1}{M} \sum_{m=1}^M L(x^m, \widehat{x}^m). \qquad (7)$$

As shown in Figure 9, after stacking multiple layers of AE, a deep network structure can be formed: stacked autoencoder (SAE). This structure uses the hidden layer output of the previous layer of AE as the input layer of the next AE. The training of the model includes two stages: pretraining and fine-tuning. The unsupervised layer-by-layer pretraining stage is used to extract fault features in the signal, and the supervised global fine-tuning stage is used to optimize the model's expression of fault features and make it have diagnostic capabilities.

(2) Denoising Autoencoder (DAE). Aiming at solving the problem that the original signal of rotating machinery contains a lot of noise and nonlinear components, Vincent et al. [48] proposed adding noise with certain statistical characteristics to the sample data to train the autoencoder, so as to obtain more robust features to initialize the deep network structure, which is denoising encoder (DAE).
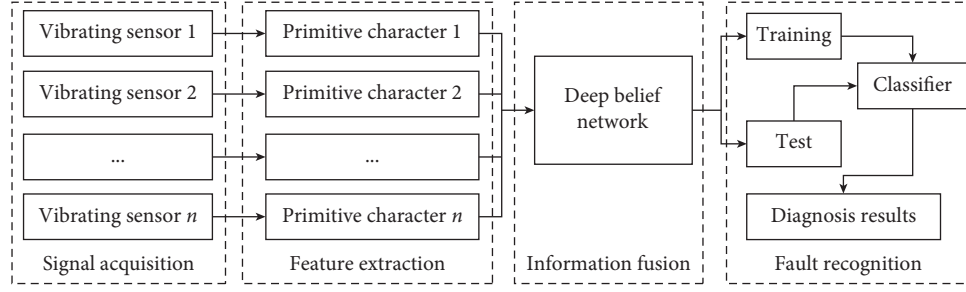
FIGURE 7: The flow diagram of multisensor information fusion proposed in [37].

TABLE 2: Summary of fault diagnosis based on optimization algorithm and DBN model.

References	Optimizing method
[38]	Particle swarm optimization (PSO)
[39–41]	Adaptive learning rate combined with Nesterov momentum
[42]	Genetic algorithm(GA)
[43]	Bacterial foraging decision
[44]	Algorithm of adaptive learning rate and momentum
[45]	The exponential moving average technique

During training, first add random noise to the sample x^m according to the q_D distribution to obtain the noisy sample $\widetilde{x}^m$:

$$\widetilde{x}^m \sim q_D\left(\widetilde{x}^m \mid x^m\right),$$
$$\arg\min_{\theta,\theta'} L\left(x^m, g_{\theta'}\left(f_\theta\left(\widetilde{x}^m\right)\right)\right). \tag{8}$$

Then, train the DAE by optimizing the objective function such as (8).

(3) Sparse Autoencoder (SAE). A sparse penalty term is added to the loss function of the autoencoder to constrain it to reduce the probability that the autoencoder network would completely copy the input information to the hidden layer during the training process. The sparse data features learned by this method can better express the input data, and better hidden layer features can be obtained when the number of model neurons is large. Usually, choose Kullback–Leibler (KL) divergence to determine the penalty term, and the penalty term PN can be expressed as follows:

$$PN = \sum_{j=1}^{S} KL\left(\rho \,\|\, \rho_j\right). \tag{9}$$

In the formula, S is the number of neurons in the hidden layer and $KL(\rho \| \rho_j)$ is KL divergence. The general cost function of neural network is expressed as

$$J(W, b) = \left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2}\|h_{W,b}\left(x(i)\right) - y(i)\|^2\right)\right]$$
$$+ \frac{\gamma}{2}\sum_{l=1}^{m_l-1}\sum_{i=1}^{S_l}\sum_{j=1}^{S_{j+1}}\left(W_{ij}(l)\right). \tag{10}$$

After adding the sparse penalty item, it can be expressed as

$$J_{\text{sparse}}(W, b) = J(W, b) + \beta \sum_{j=1}^{S} KL\left(\rho \,\|\, \rho_j\right). \tag{11}$$

In the formula, β is the weight of the sparse penalty term.

The optimized parameters W and b, which are also parameters in the sparse cost function J_{sparse}, are finally obtained in the encoding process. Therefore, the optimal parameters W and b can be obtained by minimizing the sparse cost function, and this process can be achieved by the backpropagation algorithm.

(4) Convolutional Autoencoder (CAE). Masci et al. [49] proposed to use the unsupervised learning method of the traditional autoencoder to combine the convolution and pooling operations of the convolutional neural network to achieve feature extraction, and apply the deconvolution operation to decode the features, which is convolutional autoencoder (CAE).

First, suppose there are **k** convolution kernels, and each convolution kernel is composed of parameters w^k and b^k; then, the encoding vector h^k can be expressed as

$$h^k = f\left(x * w^k + b^k\right). \tag{12}$$

By performing feature reconstruction on the obtained h^k, the following formula can be obtained:

$$y = g\left(h^k * \widehat{w}^k + d\right),$$
$$E = \frac{1}{2n}\sum\left(x_i - y_i\right)^2. \tag{13}$$

Considering (13), compare the input sample and the reconstructed result with Euclidean distance, and obtain a complete convolutional autoencoder through the optimization of backpropagation algorithm.

3.2.2. Application of Autoencoder and Its Variants in Fault Diagnosis. AE models can learn representations from machinery data in an automatic way. Jia et al. constructed a diagnostic model based on stacked autoencoders (SAE), which automatically learned fault features from frequency domain data to diagnose rolling bearings and planetary gearboxes [15]. Liu et al. used AE to construct a recurrent neural network for motor bearing fault diagnosis [50]. Sun
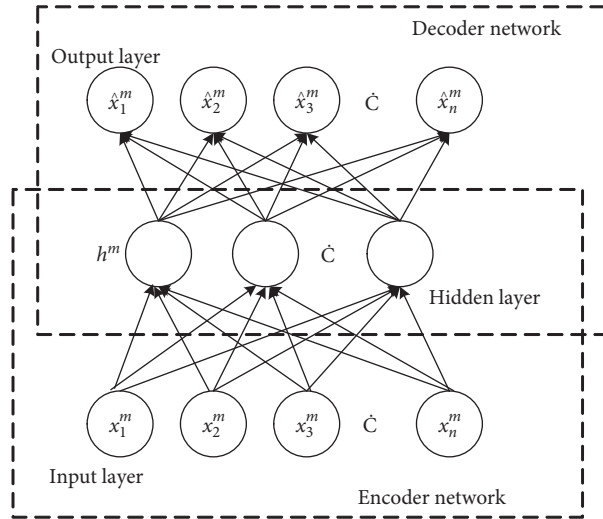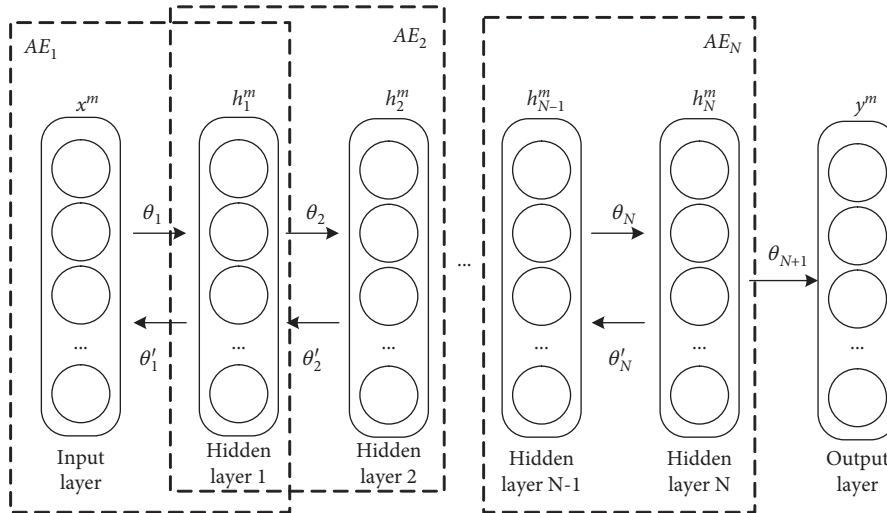
FIGURE 8: AE structure diagram.

FIGURE 9: Stacked autoencoder (SAE) composition structure.

et al. extracted low-dimensional features from the original time domain signal by compressive sensing and used them as the input features of the SAE diagnostic model [51]. Reference [52] used nonlinear soft threshold method and digital wavelet framework to preprocess the vibration signal, and then input the frequency coefficient into the SAE model to diagnose the bearing fault. Some researchers also input multidomain statistical features such as time domain features, frequency domain features, and time-frequency domain features into the SAE model for fault diagnosis [53, 54]. Shao et al. compared the diagnostic performance of the SAE-based diagnostic model when using different activation functions [55] and used Gaussian wavelet function as the activation function to propose a depth tracking wavelet adaptive encoder (TDWAE) to diagnose the bearing of electric locomotive [56]. The structure of the wavelet autoencoder used in the article is shown in Figure 10. Liu et al. used STFT to process the acoustic signal into a normalized spectrum and input it into a two-layer deep neural
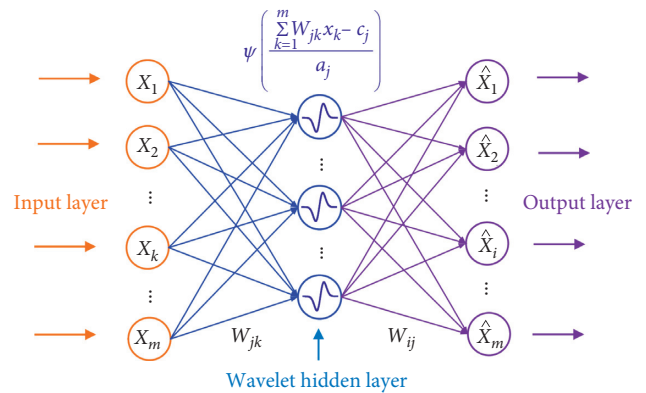

FIGURE 10: The structure of the wavelet autoencoder used in [56].

network based on SAE for rolling bearing fault diagnosis [57]. Cheng et al. [58] used SAE to extract features in the time domain, frequency domain, and time-frequency

domain. SVM is used for fault classification at the end of the network. Zhou et al. cascaded three SAE modules and used them to classify fault mode, fault type, and fault severity [59].

The original signal of rotating machinery contains a lot of noise and nonlinear components [60], AE may learn similar features during feature extraction, and the learned features have shifting mutation characteristics, which will lead to misclassification of machinery health status. Therefore, the researchers applied the AE variant structure to the fault diagnosis of rotating machinery.

Lei et al. used frequency domain signals as input and stacked multiple denoising autoencoders (DAE) to form a DNN for fault diagnosis [61]. Wang et al. [62] proposed a rolling bearing fault diagnosis method based on EMD and sparse stacked autoencoder (SSAE), as shown in Figure 11. This method uses the EMD method to obtain the IMF component of the bearing vibration signal and constructs the Hankel matrix to obtain the singular values as the input samples of SSAE.

Lu et al. used stacked denoising AE for bearing fault diagnosis, and the results showed that the diagnosis performance of the proposed method is better than traditional SVM, ANN, and other methods [63]. Sun et al. proposed to integrate denoising coding on the basis of sparse encoder to improve the robustness of feature expression [64]. Wang et al. proposed a continuous sparse autoencoder (CSAE) to identify transformer faults [65]. Shen et al. [66] and Liu et al. [67] constructed the fault diagnosis model of rotating machinery by using contraction AE and convolution AE, respectively. Zhang et al. used convolutional autoencoders (CAE) to diagnose rolling bearing faults [68]. The results showed that the method has the ability to eliminate noise, strengthen fault characteristic signals, and attenuate non-characteristic impact signals. In [69], Chen et al. proposed an improved ensemble deep autoencoder (IEDAE). Firstly, the loss function of the autoencoder is improved, and three kinds of wavelet convolution autoencoders are designed. Then, five kinds of autoencoders, such as discriminative autoencoder and wavelet convolution autoencoder, are employed to construct the corresponding deep autoencoders, and a "cross-layer" connection is designed to alleviate the gradient disappearance of the deep network. Finally, the recognition result is given using the weighted averaging method to ensure accurate and stable diagnosis result. Wu et al. modified the mean square error (MSE) commonly used in unsupervised autoencoders and proposed a semisupervised fault diagnosis method called hybrid classification autoencoder [70]. This method can use both labeled and unlabeled data to train the model. The architecture of the proposed method is shown in Figure 12.

In order to improve the performance of the fault diagnosis model, researchers have successively optimized the AE-based diagnosis model using different optimization algorithms, as shown in Table 3. In [71, 72], a batch normalization layer was added to SAE to solve the problem of internal covariate shift in multilayer network training and speed up the convergence speed. The results show that this method can achieve higher accuracy than the original SAE method. In [73], Saufi et al. proposed a method that

combines differential evolution, and a resilient back-propagation approach is proposed to improve the performance of SSAE networks in bearing fault classification. Hou et al. proposed a rolling bearing fault identification model based on particle swarm optimization of stacked noise reduction autoencoder (PSO-SDAE). The model can obtain a robust and deep-level representation of the bearing fault state characteristics [74]. Chen et al. used SAE to diagnose diesel engine faults [75]. The authors use the harmony search (HS) algorithm to optimize hyperparameters, adaptively adjust the network structure of SAE, and improve the feature extraction ability of the network. In [76], Wang et al. proposed a deep neural network based on kernel function and denoising autoencoder (DAE). Then, the chaotic firefly algorithm is used to optimize the kernel parameters and the undetermined parameters in the deep network. Based on the AE networks, [77, 78] have proposed hybrid diagnostic models for motor bearings and wind turbines, respectively. The results show that the proposed model has better generalization and convergence speed than the original AE network.

3.3. Convolutional Neural Network (CNN)

3.3.1. Basic Theory of CNN.
Convolutional Neural Network (CNN) is a feedforward neural network proposed by LeCun [79], which has been widely used in computer vision [80, 81], speech recognition [82], and other fields [83]. The typical structure of CNN is shown in Figure 13. The convolutional layer, pooling layer, and fully connected layer are all hidden layers of the convolutional neural network, and the input layer and output layer are the visible layers of the convolutional neural network.

CNN learns abstract features by alternately superimposing convolutional layers and pooling layers. The convolutional layer convolves multiple local filters with the original input data to generate translation-invariant local features. Each filter uses the same kernel to extract the local features of the input local area. The form of the convolutional layer is as follows:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j \right), \tag{14}$$

where l is the current number of layers, $f(\cdot)$ is the activation function, k_{ij} is the weight matrix of the convolution kernel, M_j represents the set of selected input features, and b_j is a bias item corresponding to each feature in the convolutional layer.

Pooling is a downsampling operation that can reduce the spatial size of features, the length of feature maps, and the number of model parameters. The pooling layer extracts fixed-length features on the sliding window according to the rules. Commonly used pooling operations include maximum pooling and average pooling [84]. The fully connected layer refers to sequentially expanding all the finally obtained feature maps to form a feature vector, and the feature vector is fully connected with the output layer. The full connection
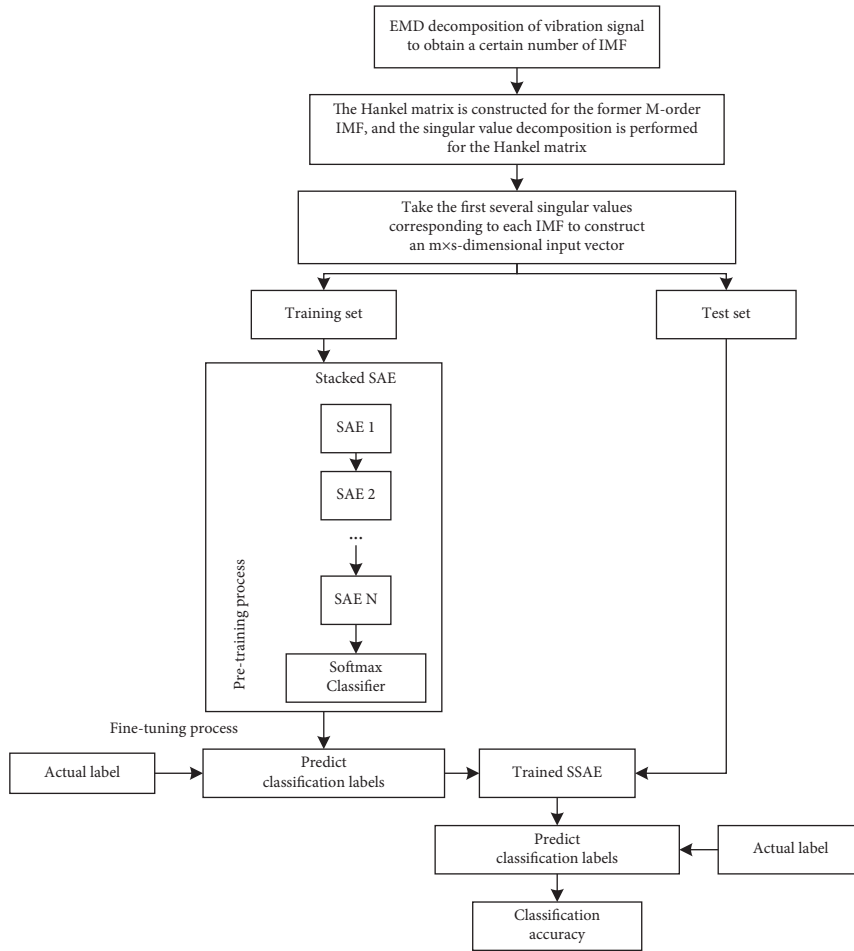
FIGURE 11: Illustrations of SSAE for rolling bearings fault diagnosis in [62].
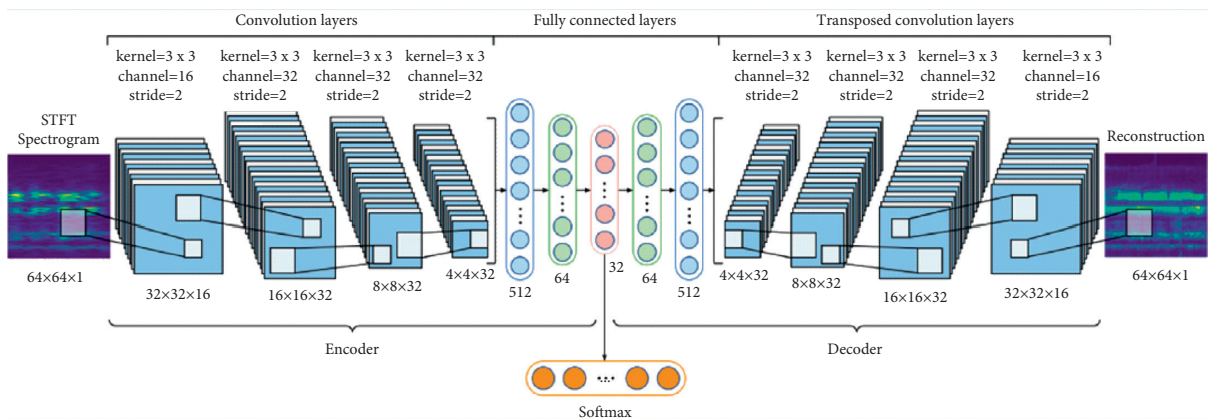

FIGURE 12: The architecture of the proposed hybrid classification autoencoder in [70].

TABLE 3: Summary of fault diagnosis based on optimization algorithm and AE model.

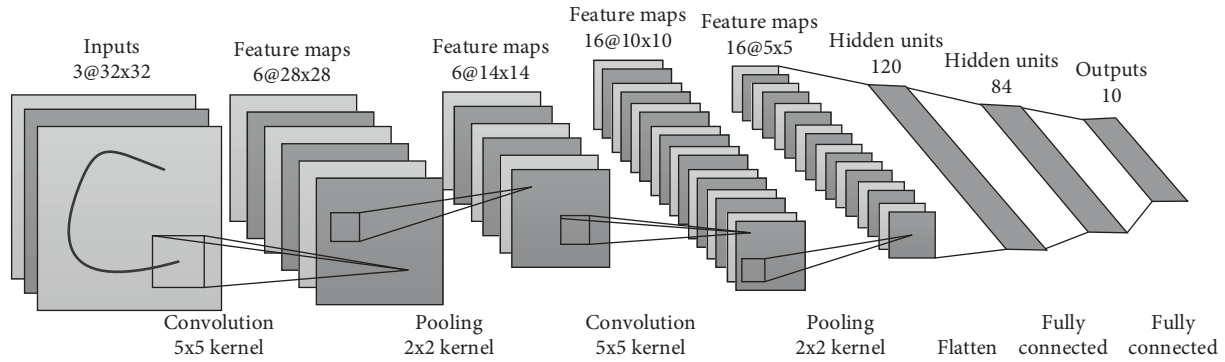| References	Optimizing method
[71, 72]	Batch normalization (BN)
[73]	Differential evolution/resilient backpropagation approach
[74]	Particle swarm optimization (PSO)
[75]	Harmony search (HS)
[76]	Chaotic firefly algorithm (CFA)
[76, 77]	Hybrid diagnostic model

FIGURE 13: CNN structure diagram.

layer is located at the end of the CNN and is used to calculate the output of the whole network.

CNN is a feature learning method with multilayer processing units, which can convert the data of the input layer into more easily recognizable features layer by layer. CNN uses local connections and weight sharing to reduce the complexity and computational complexity of the network. The local connection method effectively reduces the number of weight parameters. Weight sharing means that the weights connected by the same convolution kernel are the same, which reduces the number of training parameters, improves the convergence rate, and can effectively suppress overfitting. The training process of convolutional neural network includes two stages: forward propagation and backpropagation. Forward propagation is to input samples into the network, initialize the network parameters, and finally get the output. Backpropagation adjusts network parameters by minimizing the error cost function until the network converges or the specified iteration termination condition is reached.

3.3.2. Application of CNN in Fault Diagnosis. CNN can efficiently extract the feature information contained in massive data, which is very suitable for processing large quantities of data. Since the vibration signal is a 1D signal, researchers often use 1D CNN to extract features from vibration signals, and they conducted health detection and fault diagnosis classification for a variety of equipment, such as rolling bearings [85], motors [86], planetary gearboxes [87], and fixed gearboxes [88]. Among the above references, it is worth noting that Han et al. proposed an enhanced convolutional neural network (ECNN) that expands the receptive field [87]. This method uses a 1D convolutional layer to initially amplify the receiving field and capture the fault information in the adjacent point group in the vibration signal, then build multiple fusion expanded convolutional layers to further expand the receiving field, fully capture the long-distance dependence of the original signal, and directly input the original vibration signal into the developed fault neural network for training. The structural framework of the method is shown in Figure 14.

Reference [89] uses the motor current signal as input combined with an improved one-dimensional convolutional neural network to achieve real-time monitoring of motor faults. Janssens et al. use the original frequency domain data as the input of the 2D-CNN model. The model consists of a single-layer convolutional layer combined with a fully connected layer to complete the fault diagnosis of bearings under four types of rotating machinery conditions [90]. She et al. [91] input the multichannel signal into the multichannel 1D CNN for fault diagnosis of rolling bearing. The network structure is shown in Figure 15. Reference [92] uses Ensemble Empirical Mode Decomposition (EEMD) to decompose the original signal into intrinsic mode functions (IMF) selected based on the combined model functions (CMF) algorithm and used as the input of CNN.

The input data of CNN is usually 2D data. Therefore, some scholars use other methods to decompose or reconstruct the input vibration signal to make it suitable for the diagnosis model based on CNN to complete the intelligent fault diagnosis of the equipment [93–96]. As shown in Figure 16, some researchers use the short-time Fourier transform (STFT), the Hilbert–Huang transform (HHT) [97], the continuous wavelet transform (CWT) [98], the synchrosqueezing transform (SST) [99], and other methods to convert 1D vibration signals into 2D time-frequency images as the input of CNN. Some other researchers used gray image [100], wavelet packet energy diagram [101], infrared thermal image [102–104], root mean square diagram [105], feature statistics diagram [106], and other images as the input of CNN for fault diagnosis of rotating machinery. Wang et al. used the short-time Fourier transform (STFT) to obtain the time-frequency map of the vibration signal and then adaptively extracted the time-frequency map features through CNN [107]. The effects of preprocessing methods and hyperparameters on the accuracy of network diagnosis are also studied. The results show that batch size is the main factor affecting training accuracy and efficiency. Xiao et al. converted the vibration signal into a 2D gray image to extract image features. Then, the feedforward denoising convolutional neural network is used for noise reduction, and the CNN gradient descent algorithm is optimized for parameter adaptive learning rate [108]. Zhu et al. used CNN to perform adaptive feature extraction on the symmetrical point map converted from the vibration signal [109]. Wen et al. proposed a method to convert time domain signals into 2D images, combined with improved
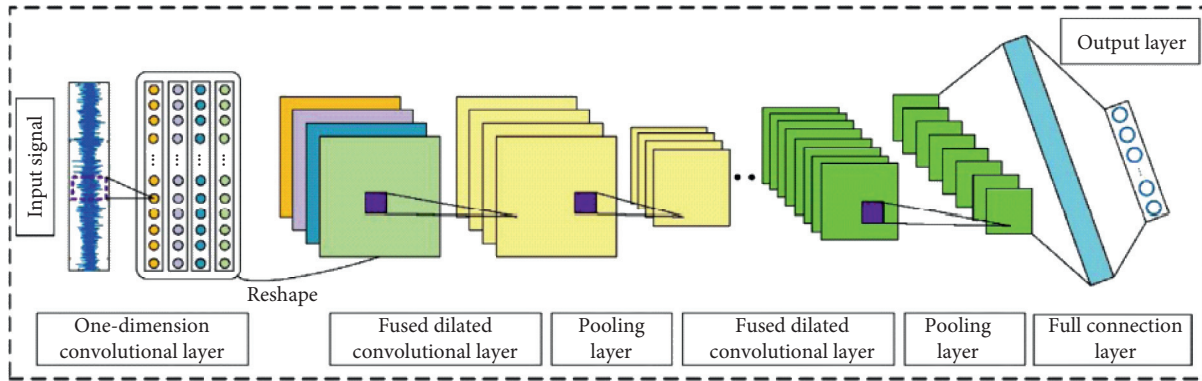
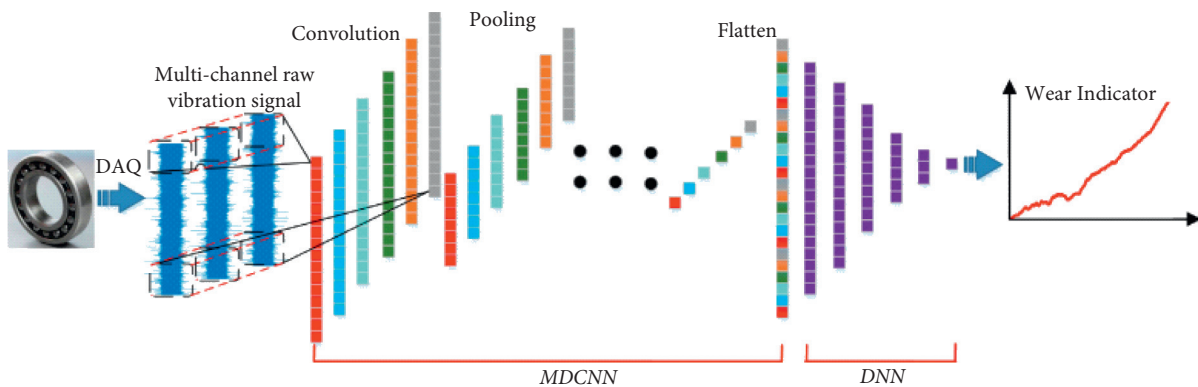FIGURE 14: The structural framework of the ECNN method proposed in [87].

FIGURE 15: The multichannel 1D CNN structure proposed in [91].

LeNet5 model to achieve fault diagnosis of bearings and centrifugal pumps [100]. Ren et al. extracted the main spectrum energy of the vibration signal, combined it into a feature map as the input of CNN, and predicted the remaining life of the bearing [110].

As shown in Table 4, some researchers have optimized and improved the diagnosis model based on CNN. Cao et al. proposed an adaptive deep convolutional neural network. The network uses particle swarm optimization (PSO) to determine the structure and parameters of the CNN model and uses principal component analysis (PCA) to visualize the diagnosis process [111]. Zhang et al. proposed a WDCNN model for bearing fault diagnosis [112]. The model has a wide first-layer kernel and a small convolution kernel. The results show that the proposed model also has high diagnostic accuracy in noisy environments. Reference [113] proposes an end-to-end convolutional neural network method. This method takes the original time signal as input and does not require any denoising or batch normalization preprocessing. It has high diagnostic accuracy under noisy environments or when the workload changes. The model structure is shown in Figure 17. Pan et al. proposed a fault classification model based on CNN and second-generation wavelet transform (SGWT): LiftingNet. They used a large-size kernel function and a nonlinear function to filter out noise [114]. Dong et al. proposed a bearing fault diagnosis method based on multilayer noise reduction technology and improved

convolutional neural network (ICNN) [115]. This method introduces an attention mechanism in the feature extraction layer of CNN, which improves the ability to extract nonsensitive features. Ye et al. proposed a multichannel weighted convolutional neural network (MCW-CNN) for feature learning and fault diagnosis of gearbox vibration signals [116]. Guo et al. used a hierarchical deep convolutional neural network to extract fault features and perform fault diagnosis on rolling bearings. The network contains two CNN modules to identify faults and determine the severity of faults [117]. Gong et al. [118] proposed an improved CNN-SVM method for motor bearing fault diagnosis. This method uses 1*1 transitional convolutional layer and global mean pooling layer to replace the fully connected layer structure of traditional CNN to reduce training parameters. As in [119], SVM is used instead of Softmax classifier to classify fault features to further improve the accuracy of diagnosis. In order to prevent part of the effective information from being filtered out by the pooling operation, Li et al. proposed a deep CNN that ignores the pooling layer to predict the remaining life of multivariable equipment [120].

As the number of CNN layers increases, it will cause gradients to disappear or explode during training. Therefore, the researchers added a residual unit on the basis of the CNN architecture and proposed the ResNet structure to solve the problem that parameters such as weights and deviations in the deep-level CNN architecture are usually not easy to
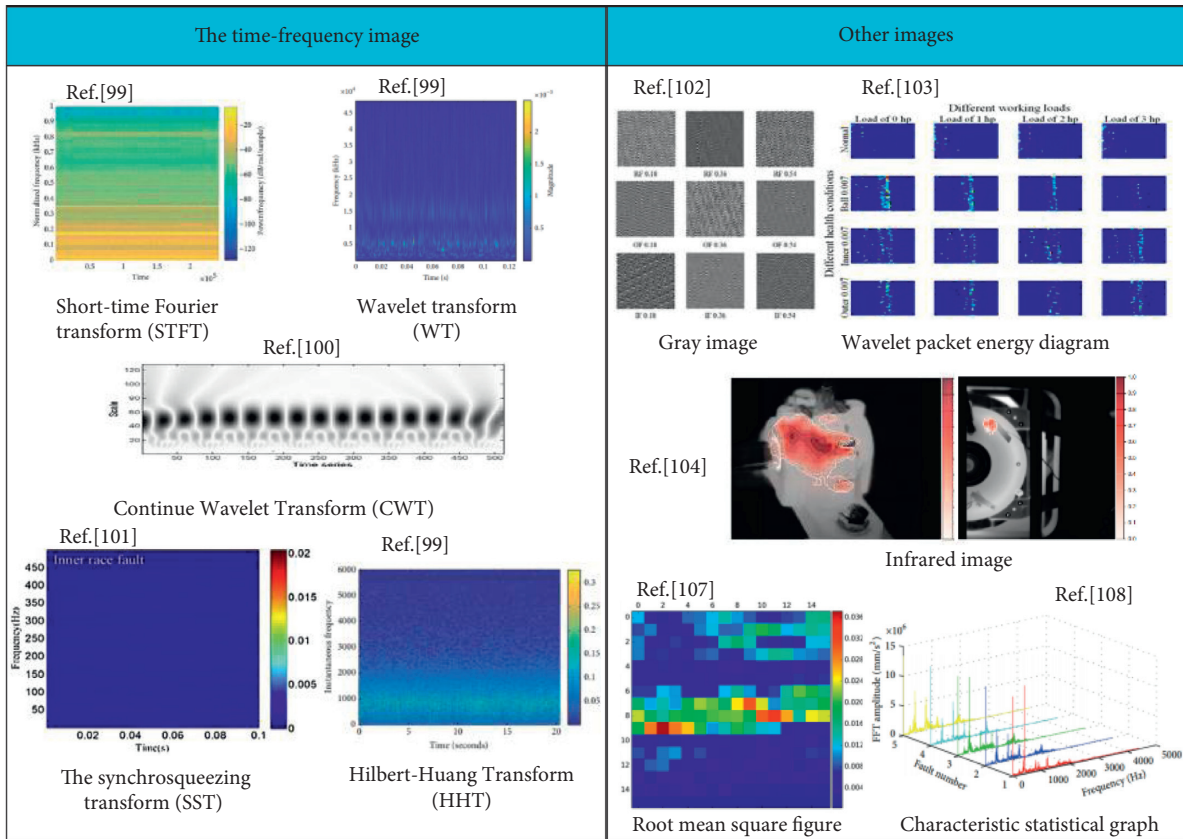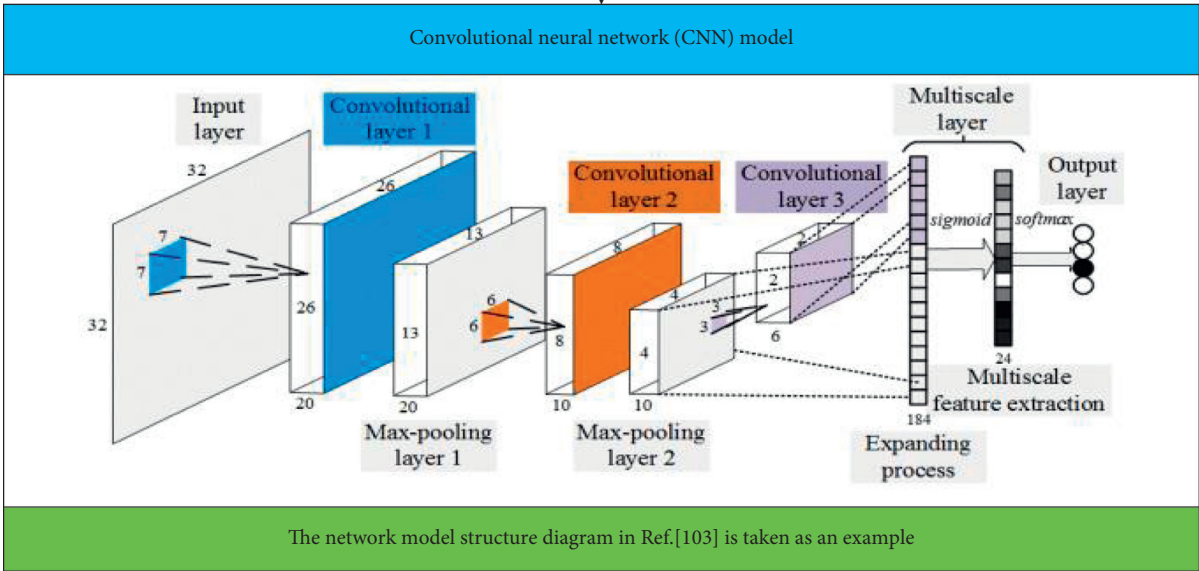
FIGURE 16: The process of diagnosis method based on CNN model with 2D image as input.

optimize [121, 122]. Under complex conditions such as variable speed conditions or variable load conditions, the fault diagnosis model of rotating machinery based on the ResNet structure can achieve higher diagnostic accuracy and generalization performance [123–126]. Reference [125] proposed a method based on time-frequency analysis and deep residual network to diagnose planetary gearbox faults. Zhao et al. used a dynamic weighted wavelet coefficient and multiwavelet coefficient fusion method to improve the performance of a ResNet-based diagnostic model. Then, the

TABLE 4: Summary of fault diagnosis based on optimization algorithm and CNN mode.

| References	Optimization and improvement
[111]	Particle swarm optimization (PSO)/principal component analysis (PCA)
[112]	Wide first-layer kernel/small convolutional kernels/AdaBN algorithm
[113]	Training intervention/end-to-end CNN
[114]	Large-size kernel function and nonlinear function to filter noise
[115]	Multilayer noise reduction/attention mechanism
[116]	Multichannel weighted convolutional neural networks
[117]	Layered deep convolutional neural networks
[118]	1*1 transitional convolution layer/global mean pooling layer
[118, 119]	Using SVM instead of Softmax classifier
[120]	Depth CNN of the pool layer is ignored
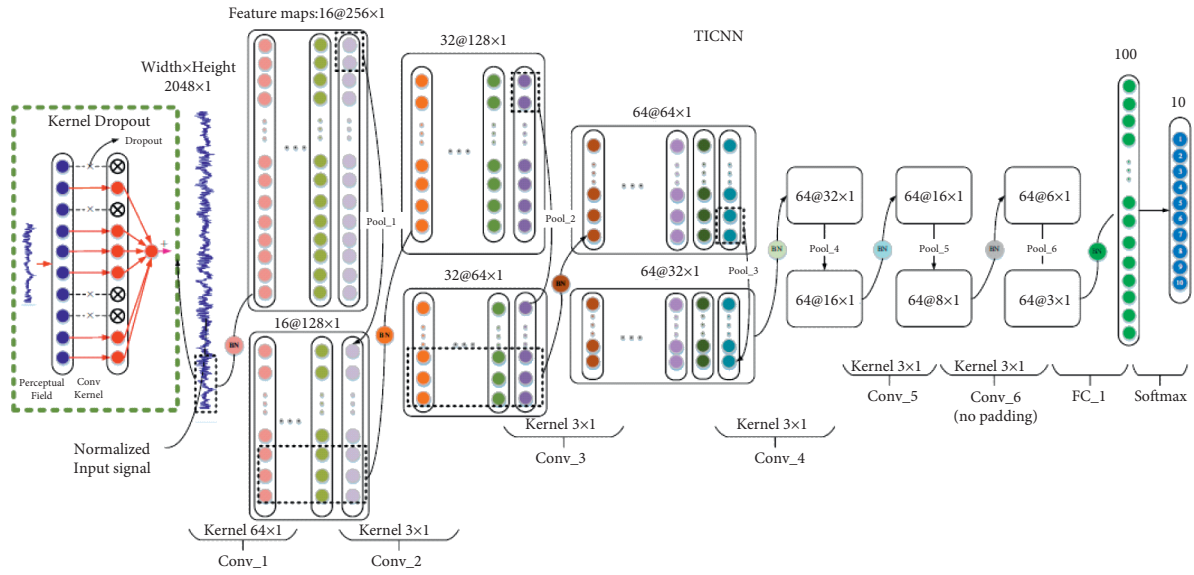
FIGURE 17: The model structure of the method proposed in [113].

fault diagnosis of planetary gearbox under severe noise environment is carried out [127, 128].

3.4. Recurrent Neural Networks (RNN)

3.4.1. Basic Theory of RNN. Recurrent neural network (RNN) is a neural network that includes feedforward connections and internal feedback connections. RNN is often used to deal with sequence problems. RNN adds self-connected neurons in the hidden layer to form internal memory and uses backpropagation to realize the memory of historical information and network state feedback. The special network structure of RNN can retain the state information of the hidden layer at a moment, and it has a powerful advantage in the field of complex dynamic system modeling. The model structure is shown in Figure 18.

In the figure, x_t is the input unit at time t, h_t is the hidden state at time **t**, and $\mathbf{o_t}$ is the output of the network at time t. U, V, W represent the connection weight between layers. After the repetitive structure in the network is expanded, network parameters such as the weight matrix and bias items can be shared. The calculation formula at time t is as follows:
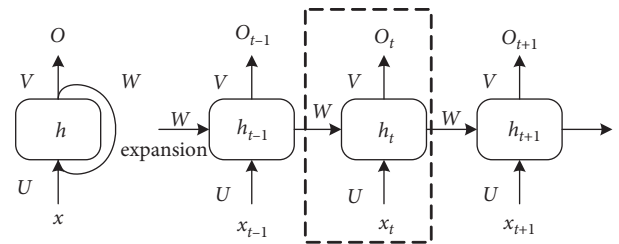
FIGURE 18: RNN model structure.

$$h_t = f\left(Wh_{t-1} + Ux_t + b_h\right). \tag{15}$$

In the formula, b_h is the bias vector at time t and $f(\cdot)$ is the activation function. The value of each hidden layer of the RNN is determined by the input at the current moment and the value of the hidden layer at the previous moment. RNN can be divided into two different networks of Jordan type and Elman type according to different feedback paths.

The traditional RNN is equivalent to a multilayer feedforward neural network. As the length of the time series increases, the number of network layers will increase and the amount of calculation will increase significantly. When

dealing with long-term monitoring sequences, large pre-
diction deviations will occur, and gradients may disappear or
explode [129]. Researchers have developed improved models
such as long short-term memory models (LSTM) and gated
recurrent units (GRU) based on standard RNN to solve the
shortcomings of RNN.

Long short-term memory (LSTM) is the most repre-
sentative variant of RNN. LSTM uses a memory unit con-
taining a gate structure to replace neurons in the hidden
layer, which can add information to or forget the cell state
and allowing information to pass through selectively.

As shown in Figure 19, the LSTM memory cell structure
includes input gates, forget gates, output gates, and input
modulation gates. The input gate, forget gate, and output
gate use the Sigmoid function to control the switching state
of the gate, and the input modulation gate uses the Tanh
function to control the switching state of the gate [130]. The
output of the forget gate is the product of the input in-
formation and the cell state at the last moment. The forget
gate controls the forgetting of information. The output of the
input gate is the product of the input information and the
output of the modulation gate. The function of the input gate
is to control the input of information. The sum of the above
two is used as the cell state of the LSTM memory unit at the
current moment, and the product of the output gate and the
cell state processed by the Tanh layer is used as the input
information of the next unit.

3.4.2. Application of RNN in Fault Diagnosis.

RNN can
retain the state information of the hidden layer at a moment,
which overcomes the limitations of simple neural networks.
It is mainly used to process sequential data or degraded data.
Jiang et al. proposed an adaptive RNN for intelligent fault
diagnosis of bearings and used an adaptive learning algo-
rithm to further improve the performance of the diagnostic
model [131]. Reference [132] uses RNN to model the op-
erating behavior of wind power generation systems. The
fault identification is achieved by comparing the residuals
between the real system output and the model output.
Reference [133] uses RNN to solve the robustness problem of
actuator fault diagnosis in dynamic nonlinear systems.
Helmes trains the RNN-based model through the BPTT
algorithm and the extended Kalman filter method to achieve
the prediction of the remaining life [134]. Lin et al. used the
recursive fuzzy neural network model to perform fault-
tolerant control of the permanent magnet synchronous
motor position servo drive [135].

It is difficult for traditional RNN methods to analyze and
process multidimensional data, and there are large prediction
errors when processing long-term sequences. Therefore, re-
searchers have improved the RNN method. Wu et al. used
LSTM to predict the remaining life of engineering equipment
and used dropout technology to improve the generalization
ability of LSTM [136]. The results showed that the prediction
effect of the proposed method is better than that of the
traditional RNN model. Reference [137] uses the three models
of traditional RNN, LSTM, and GRU to diagnose and predict
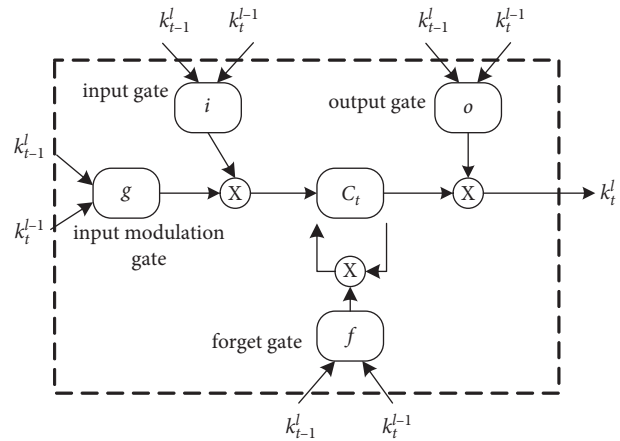aircraft engine faults. The results show that the performance

FIGURE 19: LSTM memory cell structure.

of LSTM and GRU is better than traditional RNN. Zhang et al.
proposed a residual life prediction method based on bidi-
rectional LSTM, an architecture that is specialized in dis-
covering the underlying patterns embedded in time series, to
track the system degradation and consequently to predict the
RUL [138]. The model structure is shown in Figure 20.

Zhao et al. proposed a convolutional bidirectional long
short-term memory network (CBLSTM) to predict tool wear
failure [139]. This method uses CNN to extract local features
and uses bidirectional LSTM to encode the time information
output by CNN. Finally, the fully connected layer and the
linear regression layer are superimposed to predict the target
value. Liu et al. established an intelligent prediction model
for rolling mill flutter energy value based on long short-term
memory (LSTM) cyclic neural network and analyzed the
influence of different time steps on the prediction effect to
obtain the optimal prediction step [140]. Reference [141]
uses the weighted feature averaging method and bidirec-
tional GRU to build an enhanced two-way GRU network for
machine health monitoring.

3.5. Generative Adversarial Networks (GAN)

3.5.1. Basic Theory of GAN.

The monitoring data of rotating
machinery under actual working conditions has the problem
of data imbalance; that is, the amount of normal data is far
more than the fault data, and the number of different types of
fault data is not balanced. When using imbalanced data for
fault diagnosis, the classification boundary of the classifier
will be biased toward most classes. As a result, it is difficult to
identify minority samples, and the performance of the
classifier will be seriously affected.

Generative adversarial network (GAN) [142] is a feature
learning algorithm based on game scenarios. Feature
learning is performed through adversarial learning, which
can be used to solve the above-mentioned data imbalance
problem. The structure of GAN is shown in Figure 21. The
generative confrontation network consists of two parts:
generator and discriminator. The input of the generator is
random noise z obeying a certain distribution, the output is
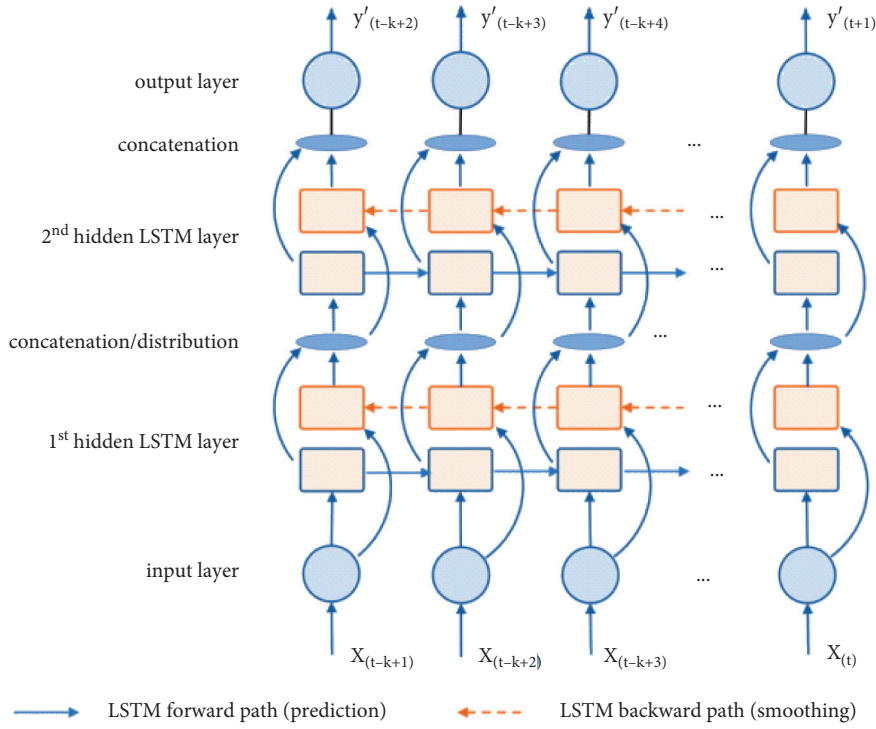a generated sample $G(z)$ similar to the real sample x. The

FIGURE 20: The structure of residual life prediction method based on bidirectional LSTM proposed in [138].
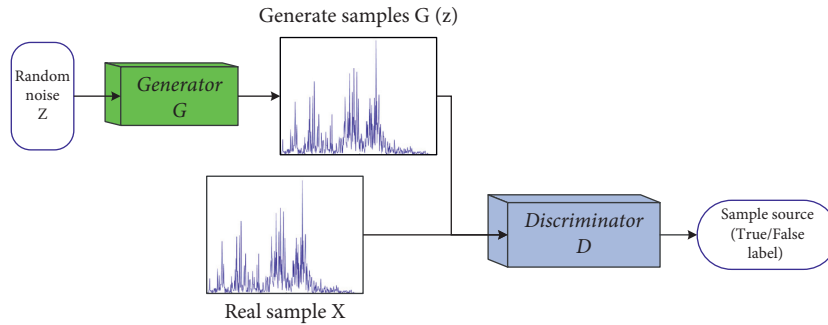

FIGURE 21: The structure of GAN.

input of the discriminator is the real sample and the generated sample. The function of the discriminator is to distinguish the source of the input sample, that is, the probability that $G(z)$ comes from x. When the discriminator inputs x, $D(x)$ is close to 1; when $G(z)$ is input, $D(G(z))$ is close to 0.

In the training process, the generator optimizes the discriminant results of the discriminator to improve the generation ability, making $G(z)$ as similar as possible to the real sample x, so that the discriminator cannot distinguish the source of samples. The discriminator optimizes itself through the probability of misjudgment so as to improve the discriminant ability of generated samples. The whole network is optimized through mutual antagonism; that is, the final output of the generator is close to the generated samples of the real sample distribution, and the discriminator cannot distinguish the generated samples. The objective function of GAN training can be expressed as follows:

$$\min_{G} \max_{D} V(D, G) = E_{x \sim p_{data}(x)}[\log D(x)] \\ + E_{z \sim p_z(z)}[\log(1 - D(G(z)))]. \quad (16)$$

The discriminator in GAN can be regarded as a kind of classifier to distinguish the authenticity of samples, and cross entropy is often used to distinguish the similarity of sample distribution. The formula is as follows:

$$H(p, q) = -\sum_i p_i \log q_i. \quad (17)$$

3.5.2. Application of GAN in Fault Diagnosis. Based on GAN, data similar to minority samples can be generated to solve the problem of data imbalance. Wang et al. used GAN to generate new samples with similar distributions to the original samples to expand the failure sample set, combined with the stacked noise reduction autoencoder for the fault

diagnosis of the planetary gearbox [143]. Liu et al. [144] and Mao et al. [145] used GAN and stacked denoising autoencoder to solve the problem of data imbalance in bearing fault diagnosis. The results show that the fault samples generated by GAN can improve the fault diagnosis accuracy in the case of data imbalance. Sun et al. proposed an adversarial generative oversampling model based on a generative adversarial network (GAN) to produce valuable artificial samples for minority class to balance the data distribution and used it for tool breakage detection [146]. Different from the previous research using GAN, it uses the discriminator to filter the samples generated by the generator to achieve effective oversampling. The framework structure and network model structure of the proposed method are shown in Figures 22 and 23.

Xie et al. proposed using periodic consistency GAN to generate false samples for auxiliary diagnosis to diagnose bearing faults under different operating conditions [147]. Li et al. used a GAN-based generative model to generate false fault samples for fault diagnosis of bearings under different operating conditions [148]. Reference [149] proposed a method based on deep convolutional generative adversarial networks (DCGAN) to accurately detect rolling bearing faults under data imbalance and complex dynamic conditions. Dai et al. combined generative adversarial networks and autoencoders to construct an encoding-decoding-recoding network model to detect anomalies in mechanical systems [150].

4. Transfer Learning (TL)

4.1. Basic Theory of TL. The current machine learning methods used for fault diagnosis of rotating machinery are mostly based on the assumption that training data and test data are in the same feature space and have the same distribution. However, there is a data imbalance between the health data and fault data of rotating machinery in actual working conditions, and the cost of labeling large-scale data is extremely high or even impossible. Therefore, it is difficult to construct a large-scale and well-labeled data set. At the same time, the above assumptions are often not true in actual working conditions. The diagnostic model trained with the training set will have poor performance on the test set.

Transfer learning (TL) relaxes the assumption that training data and test data must be independent and identically distributed. TL can apply knowledge or patterns learned in a certain field or task to different but related fields and has been widely used in many fields [151, 152].

Assume that the sample space of a machine learning task T is $X \times Y$, where X is the input space, Y is the output space, and its probability density function is $p(x, y)$. Suppose X is a subset of the d-dimensional real number space and Y is a discrete set. A sample space and its distribution can be called a domain: $D = (X, Y, p(x, y))$. Given two domains, they are considered different if at least one of their input spaces, output spaces, or probability distributions is different, then the two domains are considered to be different. Transfer learning refers to the process of knowledge transfer between two different domains. Features or knowledge structures are

transfered from the source domain to help with learning tasks in the target domain to complete or improve learning in the target domain, where labeled data in the target domain are missing or not available. The number of training samples in the source domain is generally much larger than that in the target domain. Different from traditional machine learning methods, as shown in Figure 24, transfer learning frameworks focus on using transferable characteristics or knowledge of the source domain to improve model performance, which can reduce the number of samples required in the target domain.

Deep learning methods can learn the deep representation of data and provide cross-domain invariant features for transfer learning. Transfer learning based on cross-domain invariant features can effectively reduce the difference between source domain and target domain, so transfer learning and deep learning methods are usually closely combined. Transfer learning is generally divided into instance-based transfer learning, feature-representation transfer learning, parameter transfer learning, and relational knowledge transfer learning [153].

4.2. Application of TL in Fault Diagnosis of Rotating Machinery. Using the transfer learning method to carry out fault diagnosis research on rotating machinery can solve the problem of low accuracy of the diagnostic model caused by insufficient labeling data under actual working conditions through sufficient labeling data in the laboratory environment. Zheng et al. proposed a bearing fault diagnosis model to fuse the diagnosis knowledge of multiple working conditions and complete the diagnosis tasks under other working conditions [154]. Tong et al. carried out fault diagnosis research on motor bearing [155] and conveyor roller bearing [156] based on the transfer learning method. Chen et al. proposed an improved LSSVM transfer learning method based on auxiliary data to solve the problem of insufficient bearing data available under different working conditions [157]. Lei et al. proposed a deep transfer diagnosis method for the transfer diagnosis between laboratory bearings and electric locomotive bearings [22]. This method extracts transfer fault characteristics from the monitoring data of different devices by constructing a domain-sharing deep residual network and then imposes domain adaptation regular term constraints during the training process to form a deep transfer diagnosis model. Lu et al. used stacked autoencoders (SAE) to extract fault features with similar distributions to perform transfer learning fault diagnosis on motor bearings and gears [158]. For motor bearings under different operating conditions, Zhang et al. [159] and Hasan et al. [160] used target operating conditions samples to fine-tune the pretrained diagnostic model. Compared with the diagnostic model that only trains a small number of target domain samples, the fine-tuned diagnostic model has faster convergence speed and higher diagnostic accuracy. The schematic diagram of the method is shown in Figure 25. Shao et al. proposed a method based on the enhanced depth transfer autoencoder for the diagnosis of bearing fault transfer between different mechanical equipment [161]. This method
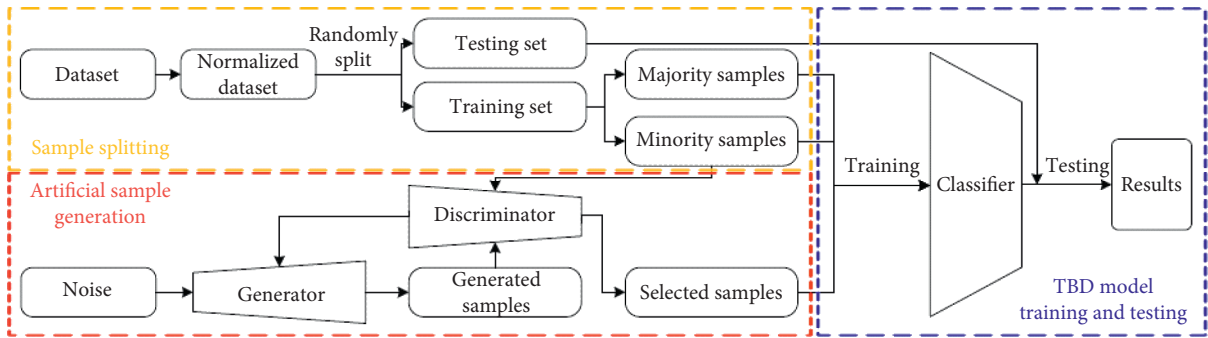
FIGURE 22: The structure of method for detecting tool breakage proposed in [146].
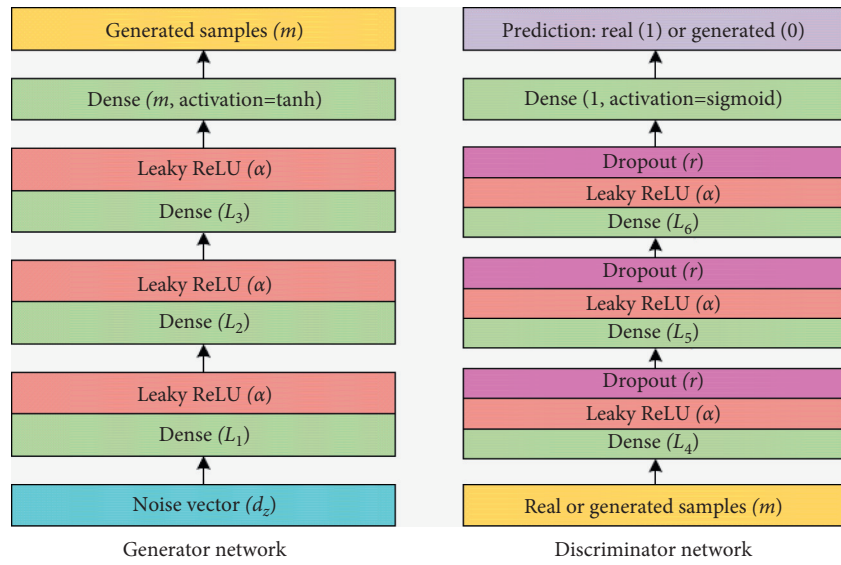

FIGURE 23: The network structure of the adversarial generative oversampling model proposed in [146].
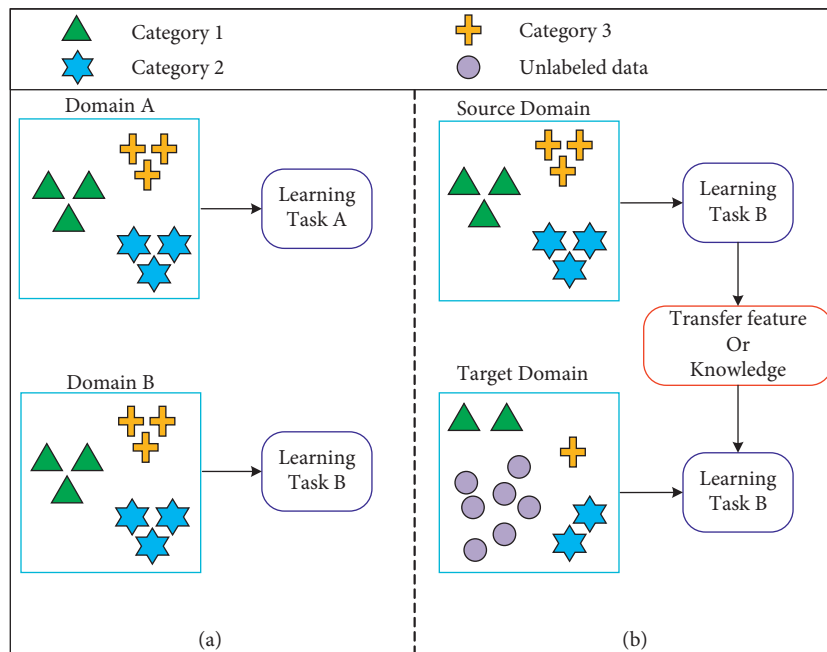

FIGURE 24: Schematic diagram of traditional method and transfer learning method: (a) traditional method; (b) transfer learning method.
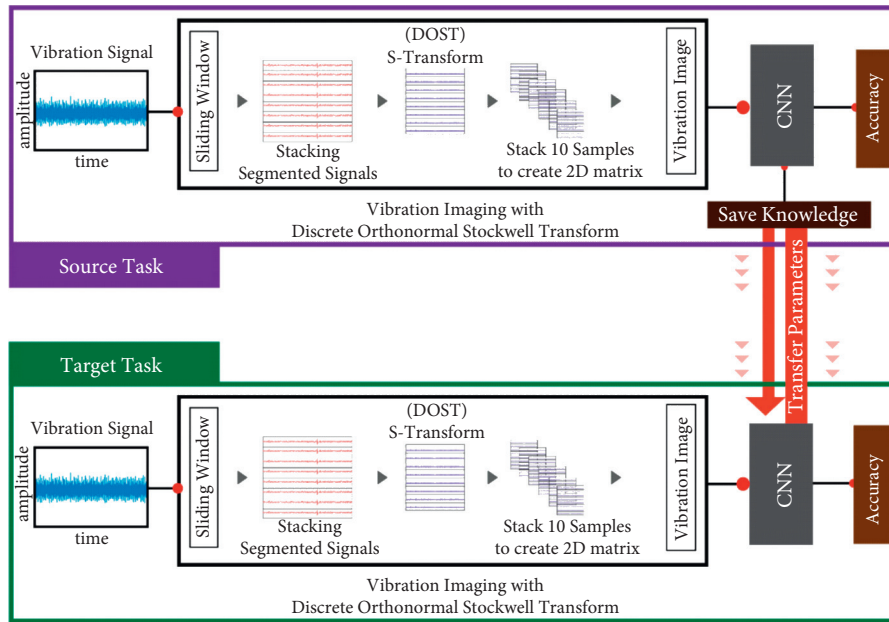
FIGURE 25: Illustrations of the fine-tuned diagnostic model based on the TL method proposed in [160].

first obtains the initialization parameters of the target model by pretraining with sufficient source domain data and then uses a sample of the target domain to fine-tune the target model to adapt to the remaining samples. Cao et al. [162] and Shao et al. [163] converted vibration data into 2D time-frequency images and then performed parameter transfer on the pretrained image recognition model to obtain the pretrained network. Finally, the fine-tuned network is used for feature extraction and fault classification of time-frequency images. Reference [164] used the deep network model trained in the source domain to complete information transfer in the target domain by fine-tuning parameters, and it constructed a RUL prediction model with good feature representation.

Another method based on transfer learning is to realize fault diagnosis by reducing the difference in the distribution of sample characteristics between the source domain and the target domain. Wen et al. [165] proposed a SAE-based domain adaptive method to diagnose bearings under different working conditions and used the maximum mean difference (MMD) term to measure the difference between domains. Li et al. [166] and Zhang et al. [167] used MMD as a loss function to measure the difference in the distribution of data in the two domains and constructed a CNN-based transfer diagnosis model. As shown in Figure 26, Yang et al. [20] used a CNN-based transfer diagnosis model to transfer the diagnosis knowledge of laboratory motor bearings to the diagnosis of locomotive bearings. Multilayer MMD and its improvement method [168,169] were added to the diagnosis model to improve the transmission performance and robustness of the diagnosis model.

Qian et al. [170] and Zhang et al. [112] used adaptive batch normalization (AdaBN) operation in the CNN-based bearing transfer diagnosis model to improve the model diagnosis performance under different working conditions. Chen et al. used TCA to reduce the distribution difference of rolling bearing monitoring data under different operating conditions and extract transfer characteristics [171]. Han et al. [172] proposed a deep transfer intelligent fault diagnosis framework that extends marginal distributed adaptation (MDA) to joint distributed adaptation (JDA). As shown in Figure 27, the framework uses a discriminant structure related to the source domain labeled data to adapt to the conditional distribution of unlabeled target data, thereby ensuring more accurate distribution matching. Mao et al. [173] used TCA and SVM classifiers to study gearbox transfer fault diagnosis methods and RUL predictions.

5. Performance Evaluation Index of Rotating Machinery Fault Diagnosis Method Based on Deep Learning

After a researcher proposes a specific research method, it is usually necessary to conduct a rigorous evaluation, that is, to evaluate the performance of the method through a suitable evaluation standard, judging the validity and practicability of the method model. However, in the evaluation process, factors such as the selection of evaluation indicators, data types, and model application scenarios will affect the performance of the method model. Therefore, the selection criteria of evaluation indicators are different in different application fields. Some scholars analyzed various performance indicators and classification methods. Some commonly used performance indicators and error measures are systematically organized, and their basic principles, application suggestions, and limitations are summarized [174,175]. This section provides an overview of the performance evaluation indicators that are widely used in the field of rotating machinery fault diagnosis. However, in specific research, researchers still need to make targeted selection of the evaluation indicators used according to the
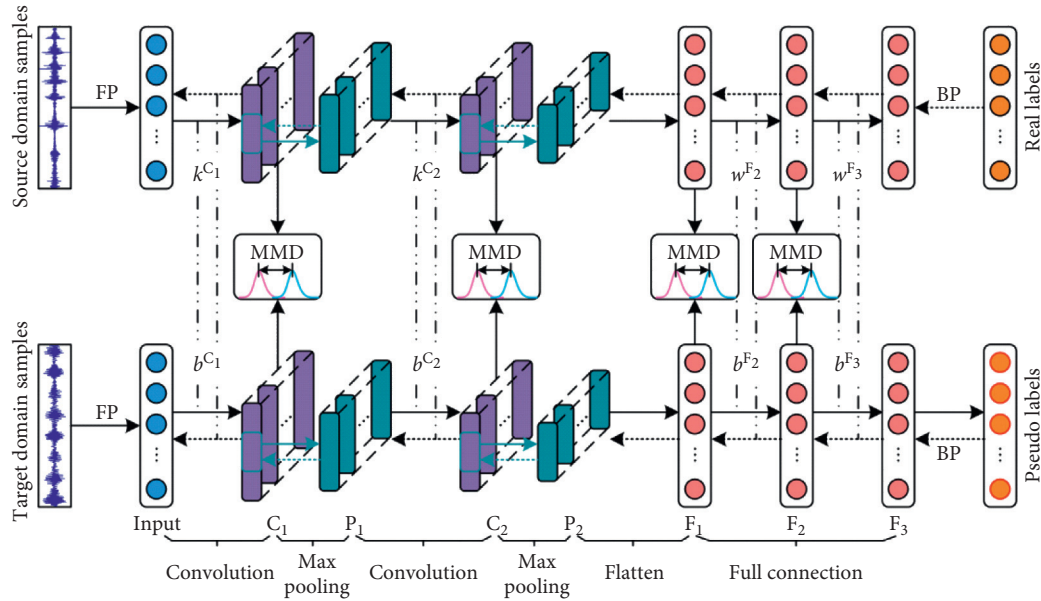
FIGURE 26: Illustrations of the proposed CNN-based transfer diagnosis model in [20].
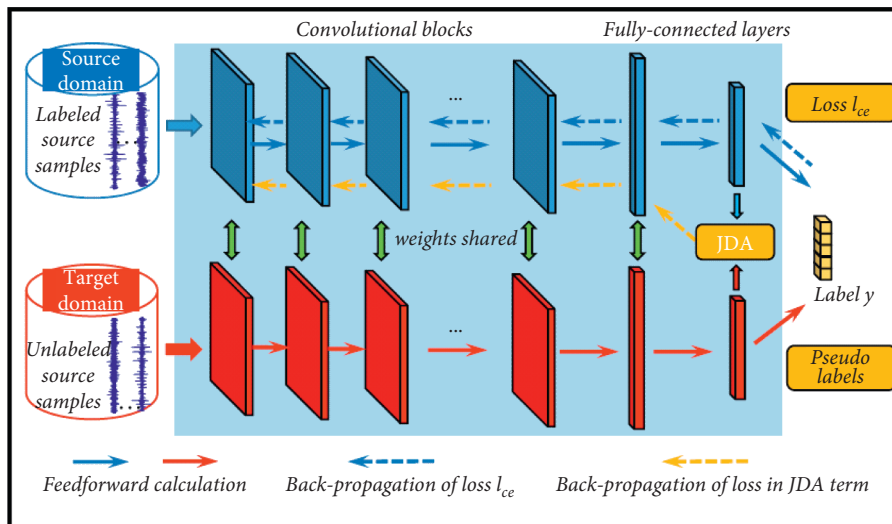
FIGURE 27: Illustrations of the proposed intelligent fault diagnosis framework in [172].

actual situation in order to correctly evaluate the performance of the method.

5.1. Classification Task Performance Evaluation Index.
Related research in the field of fault diagnosis of rotating machinery mainly focuses on the motors, bearings, gearboxes, and other components in mechanical equipment. Most of these studies are classified tasks. Take the fault diagnosis research of rolling bearing as an example; the diagnosis target is the fault type of the bearing (normal bearing, bearing inner ring fault, bearing outer ring fault, bearing rolling element fault). Performance evaluation indicators widely used in these classification tasks include the following:

(1) Accuracy (ACC)

ACC is defined as evaluating the ratio of the number of correct predictions to the total number of samples. Its expression is

$$ACC = \frac{TP + TN}{P + N} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (18)$$

where P denotes the number of real positives, N denotes the number of real negatives, TP denotes true positives, TN denotes true negatives, FP denotes false positives, and FN denotes false negatives.

(2) True Positive Rate (TPR)/Recall

TPR is defined as measuring the proportion of actual positives that are correctly identified as positives. It can be used to describe the proportion

of actual positives correctly identified. Its expression is

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN}. \tag{19}$$

(3) Positive Predictive Value (PPV)/Precision

PPV is defined as the proportions of positive observations that are true positives. Its expression is

$$PPV = \frac{TP}{TP + FP}. \tag{20}$$

(4) False Positive Rate (FPR)

FPR is defined as measuring the proportion of positive cases that are correctly identified as positives. It can be described proportion of negative cases incorrectly identified as positive cases. Its expression is

$$FPR = \frac{FP}{N} = \frac{FP}{FP + TN}. \tag{21}$$

(5) False Discovery Rate (FDR)

FDR is defined as expected proportion of false observations. It can be described as the proportion of the individuals with a positive test result for which the true condition is negative. Its expression is

$$FDR = \frac{FP}{FP + TP}. \tag{22}$$

(6) F1-Score

F1-score is defined as harmonic mean of the precision and recall. Its expression is

$$F_1 = \frac{2TP}{2TP + FP + FN}. \tag{23}$$

5.2. Predictive Task Performance Evaluation Index. In addition to the classification of fault types, the research field of fault diagnosis of rotating machinery also includes another important research direction, that is, the prediction of the remaining life of the equipment. The remaining life prediction research of rotating machinery is a kind of prediction task. Given sample set $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, where y_i is the true label of sample x_i, to evaluate the performance of method model f, it is necessary to compare the model prediction result $f(x)$ with the real label y. The most commonly used performance evaluation indicators for this type of task include the following:

(1) Error (E)

E is defined as the amount by which an observation differs from its actual value. Its expression is

$$E = A - P, \tag{24}$$

where A is actual measurements and P is predictions.

(2) Mean Error (ME)

ME is defined as the average of all errors in a set. Its expression is

$$ME = \frac{1}{n} \sum_{i=1}^{n} E_i. \tag{25}$$

(3) Mean Absolute Error (MAE)

MAE is defined as measuring the difference between two continuous variables. Its expression is

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |E_i|. \tag{26}$$

(4) Mean Squared Error (MSE)

MSE is defined as measuring the average of the squares of the errors. Its expression is

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (E_i)^2. \tag{27}$$

(5) Root Mean Squared Error (RMSE)

RMSE is defined as root square of average squared error. Its expression is

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (E_i)^2}. \tag{28}$$

(6) Mean Absolute Percentage Error (MAPE)

MAPE is defined as measuring the extent of error in percentage terms. Its expression is

$$MAPE = \frac{\sum_{i=1}^{n} |E_i| / |A_i|}{n/100}. \tag{29}$$

(7) Correlation Coefficient (R)

R is defined as measuring the strength of association between variables. Its expression is

$$R = \frac{\sum_{i=1}^{n} (A_i - \overline{A_i})(P_i - \overline{p_i})}{\sum_{i=1}^{n} (A_i - \overline{A_i})^2 \sum_{i=1}^{n} (P_i - \overline{p_i})}. \tag{30}$$

5.3. Other Performance Evaluation Methods. In addition to the above-mentioned performance evaluation indicators, there are also some methods that can intuitively evaluate the performance of the method model and are widely used. These methods include the following.

5.3.1. Confusion Matrix. The content of the confusion matrix is the statistical data of actual classification and predicted classification. Each column in the matrix represents a predicted instance, and each row represents an actual instance. Confusion matrix is of great significance for understanding the accuracy measurement of model classification effect [174]. As shown in Figure 28, the actual classification results and misjudgments of each type of sample can be found through the confusion matrix [176].
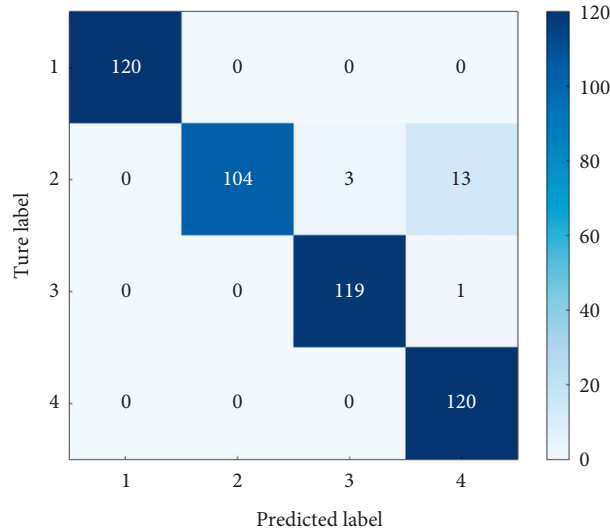
FIGURE 28: Schematic diagram of confusion matrix of classification results.

5.3.2. t-SNE Visualization. t-SNE is a nonlinear dimensionality reduction algorithm. When researching the classification of high-dimensional data, t-SNE can project the data into a 2D or 3D space for observation and judge the separability of the data (small intervals between similar types and large intervals between heterogeneous types). It is also possible to visually analyze the features extracted from each layer of the deep learning model through t-SNE. As shown in Figure 29, the classification performance of the model can be effectively evaluated through the observed feature distribution and the degree of clustering.

In addition to the above methods, the performance of the method model can be also evaluated to a certain extent through methods such as the training/validation accuracy curve of the method model, the training/validation loss curve, and the algorithm running time. However, the hardware equipment and operating environment used by different research methods are different, so these methods have limitations. Depending on the application field of the research method and the type of data, the selection criteria of the evaluation indicators will also change. When selecting performance indicators, it is necessary to select multiple indicators or a combination of multiple indicators according to different research methods to properly evaluate the performance of the method.

6. Existing Problems and Future Development Trends

Rotating machinery fault diagnosis method based on deep learning can not only quickly and effectively extract characteristic signals reflecting equipment operating conditions, but also establish a nonlinear relationship between equipment operating monitoring data and equipment operating conditions and then accurately identify the type of equipment failure and the degree of the failure. However, there are still certain problems and challenges that need to be further studied.

Regarding research on fault diagnosis when multiple fault forms exist on the same component at the same time, existing studies have mostly neglected the simultaneous occurrence of multiple failure modes in engineering applications. For example, bearing failure may be caused by various failure forms such as corrosion of the inner ring and cracks on the surface of the outer ring. Therefore, the research on the fault diagnosis method when there are multiple fault forms of the same component at the same time is worthy of further study.

In respect of research on fault diagnosis when multiple component faults are coupled with each other, rotating machinery in engineering applications runs for a long time and under heavy load, and its multiple components often have different failures at the same time. These components usually affect each other and jointly determine the operating conditions of the equipment. Most of the existing research focuses on a certain part of the equipment, such as a bearing or gearbox and conducts its fault diagnosis research separately. Therefore, the method of equipment fault diagnosis after multiple components have failed and are coupled to each other needs to be studied in depth.

Concerning research on methods to improve the quality of operating data for rotating machinery, rotating machinery data in engineering applications has the characteristics of large data volume, multiple signal sources, different sampling forms, and easiness of interference by random factors. It is difficult to establish appropriate evaluation criteria to quantitatively explain the completeness, accuracy, and timeliness of the data. Therefore, it is necessary to study intelligent data cleaning algorithms or other methods to improve data quality and increase data availability.

With regard to research on the extended application field of fault diagnosis based on deep learning, current fault diagnosis mostly focuses on key components such as bearings or gearboxes of rotating machinery. However, there are a large number of other types of mechanical equipment in engineering applications, such as vibrating machinery or unfixed machinery. The
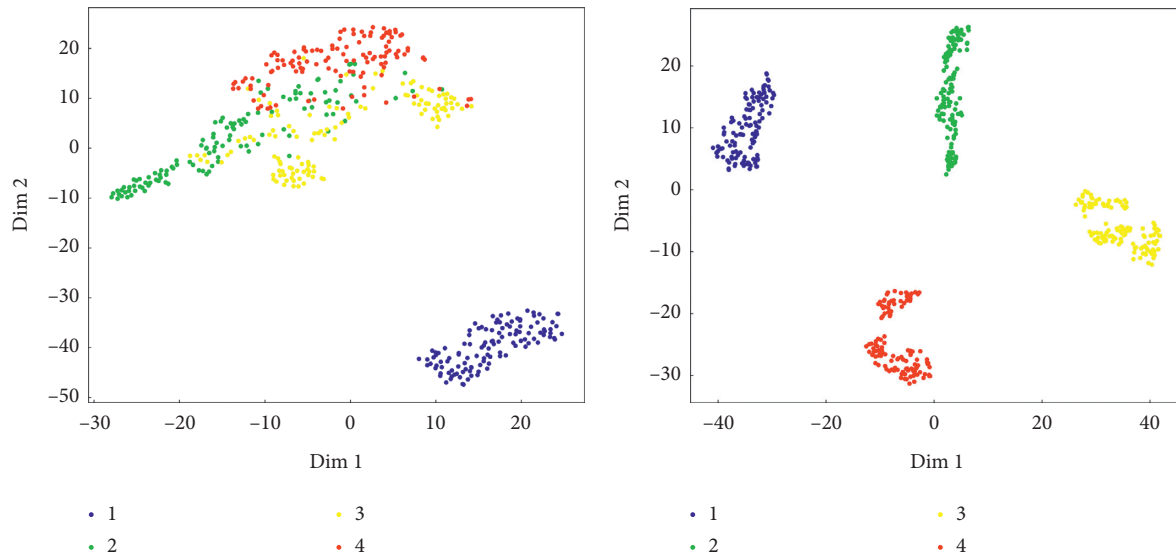
FIGURE 29: t-SNE visualization diagram.

characteristic signals of this type of mechanical equipment are different from those of rotating machinery. Therefore, it is necessary to study the application of deep learning-based fault diagnosis methods to other types of mechanical equipment such as vibrating machinery and unfixed machinery.

As for research on robustness and real-time performance of fault diagnosis methods based on deep learning, the actual working conditions of the diagnosis model are not necessarily the same. For example, a model that performs well in the diagnosis of fan bearings is not suitable for the diagnosis of coal mine belt conveyors. Therefore, it is necessary to study a more robust diagnosis method. At the same time, the diagnosis model and related algorithms need to be updated according to in-service performance to deal with new situations.

7. Conclusion

The fault diagnosis research of rotating machinery is a key link to ensure the normal operation of rotating machinery and equipment, reduce unplanned maintenance, and ensure the safety of industrial production, which is of great significance. With the rapid development of intelligent algorithms and hardware equipment, the research on fault diagnosis of rotating machinery based on deep learning has received attention. This article reviews the research on fault diagnosis of rotating machinery based on deep learning, and the conclusions and contributions obtained are as follows:

(1) The development process of fault diagnosis research in the field of rotating machinery is summarized, the advantages and disadvantages of the methods at each stage are explained, and the fault diagnosis research based on deep learning methods is pointed out as one of the future research trends.

(2) This article discusses deep belief network (DBN), autoencoder (AE) and its variants, convolutional neural network (CNN), recurrent neural network

(RNN), generative adversarial network (GAN), and transfer learning (TL) from both aspects of basic theory and specific applications. It provides a certain reference and convenience for researchers in this field to carry out follow-up research work.

(3) This article provides an overview of some commonly used evaluation indicators used to evaluate the performance of diagnostic methods. From the perspective of engineering application, the problems existing in the study of fault diagnosis of mechanical equipment with deep learning methods are analyzed, and the future research and development trends are prospected. This work provides some inspiration for researchers in this field and helps to promote the development of this field.

In future work, researchers should combine theory and practice to increase the size of the available data set as the data basis for future research, improve the accuracy and speed of the diagnosis algorithm as much as possible, and combine this with the equipment in actual working conditions to improve the diagnosis and the usability of the algorithm in the industrial field [177].

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 115, pp. 213–237, 2019.

[2] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: a review," *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, 2018.

[3] J. Jiao, M. Zhao, J. Lin, and K. Liang, "Hierarchical discriminating sparse coding for weak fault feature extraction of rolling bearings," *Reliability Engineering & System Safety*, vol. 184, pp. 41–54, 2019.

[4] K. Liang, M. Zhao, J. Lin, and J. Jiao, "An information-based K-singular-value decomposition method for rolling element bearing diagnosis," *ISA Transactions*, vol. 96, pp. 444–456, 2020.

[5] Y. Lei, F. Jia, D. Kong, J. Lin, and S. Xing, "Opportunities and challenges of machinery intelligent fault diagnosis in big data era," *Journal of Mechanical Engineering*, vol. 54, no. 5, pp. 94–104, 2018.

[6] A. Kumar, G. Vashishtha, C. P. Gandhi, Y. Zhou, A. Glowacz, and J. Xiang, "Novel convolutional neural network (NCNN) for the diagnosis of bearing defects in rotary machinery," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.

[7] A. Glowacz, "Ventilation diagnosis of angle grinder using thermal imaging," *Sensors*, vol. 21, no. 8, p. 2853, 2021.

[8] A. Glowacz, R. Tadeusiewicz, S. Legutko et al., "Fault diagnosis of angle grinders and electric impact drills using acoustic signals," *Applied Acoustics*, vol. 179, Article ID 108070, 2021.

[9] P. Gangsar and R. Tiwari, "Signal based condition monitoring techniques for fault detection and diagnosis of induction motors: a state-of-the-art review," *Mechanical Systems and Signal Processing*, vol. 144, Article ID 106908, 2020.

[10] Y. Liu, H. Gao, L. Guo, A. Qin, C. Cai, and Z. You, "A data-flow oriented deep ensemble learning method for real-time surface defect inspection," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 7, pp. 4681–4691, 2019.

[11] Y. Du, Y. Chen, G. Meng, J. Ding, and Y. Xiao, "fault severity monitoring of rolling bearings based on texture feature extraction of sparse time-frequency images," *Applied Sciences*, vol. 8, no. 9, p. 1538, 2018.

[12] B. Liang, S. Iwnicki, and Y. Zhao, "Application of power spectrum, cepstrum, higher order spectrum and neural network analyses for induction motor fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 39, no. 1-2, pp. 342–360, 2013.

[13] 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," *Mechanical Systems and Signal Processing*, vol. 138, Article ID 106587, 2020.

[14] K. Worden, W. J. Staszewski, and J. J. Hensman, "Natural computing for mechanical systems research: a tutorial overview," *Mechanical Systems and Signal Processing*, vol. 25, no. 1, pp. 4–111, 2011.

[15] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mechanical Systems and Signal Processing*, vol. 72-73, pp. 303–315, 2016.

[16] H. Pei, "Review of machine learning based remaining useful life prediction methods for equipment," *Chinese Mechanical Engineering Society*, vol. 55, no. 8, pp. 1–13, 2020.

[17] H. Jang, H. Shao, and X. Li, "Deep learning theory with application in intelligent fault diagnosis of aircraft," *Chinese Mechanical Engineering Society*, vol. 55, no. 7, pp. 27–34, 2020.

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

[19] L. Duan, M. Xie, J. Wang, and T. Bai, "Deep learning enabled intelligent fault diagnosis: overview and applications," *Journal of Intelligent and Fuzzy Systems*, vol. 35, no. 5, pp. 5771–5784, 2018.

[20] B. Yang, Y. Lei, F. Jia, and S. Xing, "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings," *Mechanical Systems and Signal Processing*, vol. 122, pp. 692–706, 2019.

[21] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: a systematic review from data acquisition to RUL prediction," *Mechanical Systems and Signal Processing*, vol. 104, pp. 799–834, 2018.

[22] Y. Lei, B. Yang, Z. Yao, and N. lv, "Deep transfer diagnosis method for machinery in big data era," *Chinese Mechanical Engineering Society*, vol. 55, no. 7, pp. 1–8, 2020.

[23] A.-r. Mohamed, G. E. Dahl, and G. Hinton, "Acoustic modeling using deep belief networks," *IEEE Transactions on Audio, Speech, Language Processing*, vol. 20, no. 1, pp. 14–22, 2011.

[24] H. Jiang, H. Shao, X. Chen, and J. Huang, "A feature fusion deep belief network method for intelligent fault diagnosis of rotating machinery," *Journal of Intelligent and Fuzzy Systems*, vol. 34, no. 6, pp. 3513–3521, 2018.

[25] D. Han, N. Zhao, and P. Shi, "A new fault diagnosis method based on deep belief network and support vector machine with Teager–Kaiser energy operator for bearings," *Advances in Mechanical Engineering*, vol. 9, no. 12, Article ID 1687814017743113, 2017.

[26] S.-Y. Shao, W.-J. Sun, R.-Q. Yan, P. Wang, and R. X. Gao, "A deep learning approach for fault diagnosis of induction motors in manufacturing," *Chinese Journal of Mechanical Engineering*, vol. 30, no. 6, pp. 1347–1356, 2017.

[27] S. Shao, W. Sun, P. Wang, R. X. Gao, and R. Yan, "Learning features from vibration signals for induction motor fault diagnosis," in *Proceedings of the 2016 International Symposium on Flexible Automation (ISFA)*, pp. 71–76, IEEE, Cleveland, OH, USA, August 2016.

[28] J. Tao, Y. L. Liu, Z. Fu, D. L. Yang, and F. Tang, "Fault damage degrees diagnosis for rolling bearing based on Teager energy operator and deep belief network," *Journal of Central South University: Science and Technology*, vol. 48, no. 1, pp. 61–68, 2017.

[29] L. Zhang, H. Gao, J. Wen, S. Li, and Q. Liu, "A deep learning-based recognition method for degradation monitoring of ball screw with multi-sensor data fusion," *Microelectronics Reliability*, vol. 75, pp. 215–222, 2017.

[30] X. Zhang, S. Guo, Y. Li, and L. Jiang, "Semi-supervised fault identification based on laplacian eigenmap and deep belief networks," *Journal of Mechanical Engineering*, vol. 56, no. 1, pp. 69–81, 2019.

[31] J. Yin and W. Zhao, "Fault diagnosis network design for vehicle on-board equipments of high-speed railway: a deep

learning approach," *Engineering Applications of Artificial Intelligence*, vol. 56, pp. 250–259, 2016.

[32] B. Chen, H. Liu, C. Xu, F. Chen, W. Xiao, and C. Zhao, "Gear fault diagnosis based on DBNS," *China Mechanical Engineering*, vol. 30, no. 02, p. 205, 2019.

[33] X. Wang, J. Huang, G. Ren, and D. Wang, "A hydraulic fault diagnosis method based on sliding-window spectrum feature and deep belief network," *Journal of Vibroengineering*, vol. 19, no. 6, pp. 4272–4284, 2017.

[34] D. Yu, Z. M. Chen, K. S. Xiahou, M. S. Li, T. Y. Ji, and Q. H. Wu, "A radically data-driven method for fault detection and diagnosis in wind turbines," *International Journal of Electrical Power & Energy Systems*, vol. 99, pp. 577–584, 2018.

[35] F. AlThobiani and A. Ball, "An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks," *Expert Systems with Applications*, vol. 41, no. 9, pp. 4113–4122, 2014.

[36] H. Oh, J. H. Jung, B. C. Jeon, and B. D. Youn, "Scalable and unsupervised feature engineering using vibration-imaging and deep learning for rotor system diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 4, pp. 3539–3549, 2017.

[37] J. Tao, Y. Liu, and D. Yang, "Bearing fault diagnosis based on deep belief network and multisensor information fusion," *Shock and Vibration*, vol. 2016, 2016.

[38] Y. Li, L. Wang, and L. Jiang, "Rolling bearing fault diagnosis based on DBN algorithm improved with PSO," *Journal of Vibration and Shock*, vol. 39, pp. 89–96, 2020.

[39] J. He, S. Yang, and C. Gan, "Unsupervised fault diagnosis of a gear transmission chain using a deep belief network," *Sensors*, vol. 17, no. 7, p. 1564, 2017.

[40] J. Tao, Y. Liu, D. Yang, and G. Bin, "Rolling bearing fault diagnosis based on bacterial foraging algorithm and deep belief network," *Journal of Vibration and Shock*, vol. 36, no. 23, pp. 68–74, 2017.

[41] 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 Transactions*, vol. 69, pp. 187–201, 2017.

[42] S. Tang, C. Shen, D. Wang, S. Li, W. Huang, and Z. Zhu, "Adaptive deep feature learning network with Nesterov momentum and its application to rotating machinery fault diagnosis," *Neurocomputing*, vol. 305, pp. 1–14, 2018.

[43] J. Xie, G. Du, C. Shen, N. Chen, L. Chen, and Z. Zhu, "An end-to-end model based on improved adaptive deep belief network and its application to bearing fault diagnosis," *IEEE Access*, vol. 6, pp. 63584–63596, 2018.

[44] C. Shen et al., "bearings fault diagnosis based on improved deep belief network by self-individual adaptive learning rate," *Journal of Mechanical Engineering*, vol. 55, no. 7, pp. 81–88, 2020.

[45] H. Shao, H. Jiang, H. Zhang, W. Duan, T. Liang, and S. Wu, "Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing," *Mechanical Systems and Signal Processing*, vol. 100, pp. 743–765, 2018.

[46] H. Shao, H. Jiang, H. Zhang, and T. Liang, "Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 3, pp. 2727–2736, 2017.

[47] D. Erhan, A. Courville, Y. Bengio, and P. Vincent, "Why does unsupervised pre-training help deep learning?" *Journal of Machine Learning Research*, vol. 11, pp. 625–660, 2010.

[48] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th international conference on Machine learning*, pp. 1096–1103, New York, NY, United States, July 2008.

[49] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Proceedings of the International conference on artificial neural networks*, pp. 52–59, Springer, Espoo, Finland, June 2011.

[50] H. Liu, J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang, "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders," *ISA Transactions*, vol. 77, pp. 167–178, 2018.

[51] J. Sun, C. Yan, and J. Wen, "Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 1, pp. 185–195, 2017.

[52] T. Junbo, L. Weining, A. Juneng, and W. Xueqian, "Fault diagnosis method study in roller bearing based on wavelet transform and stacked auto-encoder," in *Proceedings of the The 27th Chinese Control and Decision Conference (2015 CCDC)*, pp. 4608–4613, IEEE, Qingdao, China, May 2015.

[53] L. Guo, H. Gao, H. Huang, X. He, and S. Li, "Multifeatures fusion and nonlinear dimension reduction for intelligent bearing condition monitoring," *Shock and Vibration*, vol. 2016, 2016.

[54] 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, 2017.

[55] H. Shao, H. Jiang, Y. Lin, and X. Li, "A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders," *Mechanical Systems and Signal Processing*, vol. 102, pp. 278–297, 2018.

[56] S. Haidong, J. Hongkai, Z. Ke, W. Dongdong, and L. Xingqiu, "A novel tracking deep wavelet auto-encoder method for intelligent fault diagnosis of electric locomotive bearings," *Mechanical Systems and Signal Processing*, vol. 110, pp. 193–209, 2018.

[57] H. Liu, L. Li, and J. Ma, "Rolling bearing fault diagnosis based on STFT-deep learning and sound signals," *Shock and Vibration*, vol. 2016, 2016.

[58] F. Cheng, J. Wang, L. Qu, and W. Qiao, "Rotor-current-based fault diagnosis for DFIG wind turbine drivetrain gearboxes using frequency analysis and a deep classifier," *IEEE Transactions on Industry Applications*, vol. 54, no. 2, pp. 1062–1071, 2017.

[59] F. Zhou, Y. Gao, and C. Wen, "A novel multimode fault classification method based on deep learning," *Journal of Control Science and Engineering*, vol. 2017, 2017.

[60] F. Jia, Y. Lei, L. Guo, J. Lin, and S. Xing, "A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines," *Neurocomputing*, vol. 272, pp. 619–628, 2018.

[61] Y. Lei, F. Jia, X. Zhou, and J. Lin, "A deep learning-based method for machinery health monitoring with big data," *Journal of Mechanical Engineering*, vol. 51, no. 21, pp. 49–56, 2015.

[62] W. Fengtao, D. Gang, and W. Hongtao, "Rolling bearing fault diagnosis method based on EMD and SSAE," *Journal of Vibration Engineering*, vol. 32, no. 02, pp. 368–376, 2019.

[63] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising

autoencoder-based health state identification," *Signal Processing*, vol. 130, pp. 377–388, 2017.

[64] W. Sun, S. Shao, and R. Yan, "Induction motor fault diagnosis based on deep neural network of sparse auto-encoder," *Journal of Mechanical Engineering*, vol. 52, no. 9, pp. 65–71, 2016.

[65] L. Wang, X. Zhao, J. Pei, and G. Tang, "Transformer fault diagnosis using continuous sparse autoencoder," *SpringerPlus*, vol. 5, no. 1, pp. 448–513, 2016.

[66] C. Shen, Y. Qi, J. Wang, G. Cai, and Z. Zhu, "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 170–184, 2018.

[67] X. Liu, Q. Zhou, J. Zhao, H. Shen, and X. Xiong, "Fault diagnosis of rotating machinery under noisy environment conditions based on a 1-D convolutional autoencoder and 1-D convolutional neural network," *Sensors*, vol. 19, no. 4, p. 972, 2019.

[68] X. Zhang, X. Zhou, and C. Tang, "A deep convolutional autoencoding neural network and its application in bearing fault diagnosis," *Journal of Xi'an Jiaotong University*, vol. 52, no. 7, pp. 1–8, 2018.

[69] Z. Chen, X. Du, Y. Wang, and N. Zhang, "Application of improved ensemble deep auto-encoder in bearing fault diagnosis," *Control and Decision making*, vol. 36, no. 1, pp. 135–142, 2021.

[70] X. Wu, Y. Zhang, C. Cheng, and Z. Peng, "A hybrid classification autoencoder for semi-supervised fault diagnosis in rotating machinery," *Mechanical Systems and Signal Processing*, vol. 149, Article ID 107327, 2021.

[71] S. R. Saufi, Z. A. bin Ahmad, M. S. Leong, and M. H. Lim, "Differential evolution optimization for resilient stacked sparse autoencoder and its applications on bearing fault diagnosis," *Measurement Science and Technology*, vol. 29, no. 12, Article ID 125002, 2018.

[72] H. Wenqing, Y. Ming, and L. Weihua, "Rolling ele‐ ment bearing fault classification using improved stacked denoising autoencoders," *Journal of Mechanical Engineering*, vol. 54, no. 7, p. 87, 2018.

[73] K. Chen, Z. Mao, J. Zhangand, and Z. Jiang, "Diesel engine fault diagnosis based on stacked autoencoder optimized by harmony search," *Journal of Mechanical Engineering*, vol. 56, no. 11, pp. 132–140, 2020.

[74] F. Wang, X. Liu, B. Guo, G. Deng, Q. Han, and H. Li, "Application of kernel auto-encoder based on firefly optimization in intershaft bearing fault diagnosis," *Journal of Mechanical Engineering*, vol. 55, no. 7, pp. 58–64, 2019.

[75] J. Wang, S. Li, B. Han et al., "Construction of a batch-normalized autoencoder network and its application in mechanical intelligent fault diagnosis," *Measurement Science and Technology*, vol. 30, no. 1, Article ID 015106, 2018.

[76] J. Wang, S. Li, Z. An, X. Jiang, W. Qian, and S. Ji, "Batch-normalized deep neural networks for achieving fast intelligent fault diagnosis of machines," *Neurocomputing*, vol. 329, pp. 53–65, 2019.

[77] W. Mao, W. Feng, and X. Liang, "A novel deep output kernel learning method for bearing fault structural diagnosis," *Mechanical Systems and Signal Processing*, vol. 117, pp. 293–318, 2019.

[78] Z.-X. Yang, X.-B. Wang, and J.-H. Zhong, "Representational learning for fault diagnosis of wind turbine equipment: a multi-layered extreme learning machines approach," *Energies*, vol. 9, no. 6, p. 379, 2016.

[79] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[80] M. Aladem and S. A. Rawashdeh, "A single-stream segmentation and depth prediction CNN for autonomous driving," *IEEE Intelligent Systems*, vol. 36, no. 4, pp. 79–85, 2021.

[81] R. Yang, S. K. Singh, M. Tavakkoli et al., "CNN-LSTM deep learning architecture for computer vision-based modal frequency detection," *Mechanical Systems and Signal Processing*, vol. 144, Article ID 106885, 2020.

[82] V. Passricha and R. K. Aggarwal, "A hybrid of deep CNN and bidirectional LSTM for automatic speech recognition," *Journal of Intelligent Systems*, vol. 29, no. 1, pp. 1261–1274, 2020.

[83] X. Zhu, J. Zhao, D. Hou, and Z. Han, "An SDP characteristic information fusion-based CNN vibration fault diagnosis method," *Shock and Vibration*, vol. 2019, 2019.

[84] M. Sun, Z. Song, X. Jiang, J. Pan, and Y. Pang, "Learning pooling for convolutional neural network," *Neurocomputing*, vol. 224, pp. 96–104, 2017.

[85] L. Eren, T. Ince, and S. Kiranyaz, "A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier," *Journal of Signal Processing Systems*, vol. 91, no. 2, pp. 179–189, 2019.

[86] Y.-M. Hsueh, V. R. Ittangihal, W.-B. Wu, H.-C. Chang, and C.-C. Kuo, "Fault diagnosis system for induction motors by CNN using empirical wavelet transform," *Symmetry*, vol. 11, no. 10, p. 1212, 2019.

[87] Y. Han, B. Tang, and L. Deng, "An enhanced convolutional neural network with enlarged receptive fields for fault diagnosis of planetary gearboxes," *Computers in Industry*, vol. 107, pp. 50–58, 2019.

[88] Y. Yao, H. Wang, S. Li et al., "End-to-end convolutional neural network model for gear fault diagnosis based on sound signals," *Applied Sciences*, vol. 8, no. 9, p. 1584, 2018.

[89] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 11, pp. 7067–7075, 2016.

[90] O. Janssens, V. Slavkovikj, B. Vervisch et al., "Convolutional neural network based fault detection for rotating machinery," *Journal of Sound and Vibration*, vol. 377, pp. 331–345, 2016.

[91] D. She and M. Jia, "Wear indicator construction of rolling bearings based on multi-channel deep convolutional neural network with exponentially decaying learning rate," *Measurement*, vol. 135, pp. 368–375, 2019.

[92] S. K. Singh, S. Kumar, and J. P. Dwivedi, "Compound fault prediction of rolling bearing using multimedia data," *Multimedia Tools and Applications*, vol. 76, no. 18, pp. 18771–18788, 2017.

[93] C. Lu, Z. Wang, and B. Zhou, "Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification," *Advanced Engineering Informatics*, vol. 32, pp. 139–151, 2017.

[94] M. Xia, T. Li, L. Xu, L. Liu, and C. W. De Silva, "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE/ASME transactions on mechatronics*, vol. 23, no. 1, pp. 101–110, 2017.

[95] W. Gong, H. Chen, Z. Zhang et al., "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, 2019.

[96] D.-T. Hoang and H.-J. Kang, "Rolling element bearing fault diagnosis using convolutional neural network and vibration image," *Cognitive Systems Research*, vol. 53, pp. 42–50, 2019.

[97] 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 and Vibration*, vol. 2017, 2017.

[98] S. Guo, T. Yang, W. Gao, and C. Zhang, "A novel fault diagnosis method for rotating machinery based on a convolutional neural network," *Sensors*, vol. 18, no. 5, p. 1429, 2018.

[99] D. Zhao, T. Wang, and F. Chu, "Deep convolutional neural network based planet bearing fault classification," *Computers in Industry*, vol. 107, pp. 59–66, 2019.

[100] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, 2017.

[101] X. Ding and Q. He, "Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 8, pp. 1926–1935, 2017.

[102] O. Janssens, R. Van de Walle, M. Loccufier, and S. Van Hoecke, "Deep learning for infrared thermal image based machine health monitoring," *IEEE*, vol. 23, no. 1, pp. 151–159, 2017.

[103] F. Jeffali, A. Ouariach, B. El Kihel, and A. Nougaoui, "Diagnosis of three-phase induction motor and the impact on the kinematic chain using non-destructive technique of infrared thermography," *Infrared Physics & Technology*, vol. 102, Article ID 102970, 2019.

[104] A. K. Al-Musawi, F. Anayi, and M. Packianather, "Three-phase induction motor fault detection based on thermal image segmentation," *Infrared Physics & Technology*, vol. 104, Article ID 103140, 2020.

[105] S. Li, G. Liu, X. Tang, J. Lu, and J. Hu, "An ensemble deep convolutional neural network model with improved D-S evidence fusion for bearing fault diagnosis," *Sensors*, vol. 17, no. 8, p. 1729, 2017.

[106] Z. Chen, C. Li, and R.-V. Sanchez, "Gearbox fault identification and classification with convolutional neural networks," *Shock and Vibration*, vol. 2015, 2015.

[107] L.-H. Wang, X.-P. Zhao, J.-X. Wu, Y.-Y. Xie, and Y.-H. Zhang, "Motor fault diagnosis based on short-time Fourier transform and convolutional neural network," *Chinese Journal of Mechanical Engineering*, vol. 30, no. 6, pp. 1357–1368, 2017.

[108] X. Xiao, J. Wang, Y. Zhang, Q. Guo, and S. Suo, "A two-dimensional convolutional neural network optimization method for bearing fault diagnosis," *Proceedings of the CSEE*, vol. 39, no. 15, pp. 4558–4567, 2019.

[109] X. Zhu, D. Hou, P. Zhou et al., "Rotor fault diagnosis using a convolutional neural network with symmetrized dot pattern images," *Measurement*, vol. 138, pp. 526–535, 2019.

[110] L. Ren, Y. Sun, H. Wang, and L. Zhang, "Prediction of bearing remaining useful life with deep convolution neural network," *IEEE Access*, vol. 6, pp. 13041–13049, 2018.

[111] J. Cao, S. Wang, X. Yue, and N. Lei, "Rolling bearing fault diagnosis of launch vehicle based on adaptive deep CNN," *Journal of Vibration and Shock*, vol. 2020, 2020.

[112] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, 2017.

[113] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mechanical Systems and Signal Processing*, vol. 100, pp. 439–453, 2018.

[114] J. Pan, Y. Zi, J. Chen, Z. Zhou, and B. Wang, "LiftingNet: a novel deep learning network with layerwise feature learning from noisy mechanical data for fault classification," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 6, pp. 4973–4982, 2017.

[115] S. Dong, X. Pei, W. Wu, B. Tang, and X. Zhao, "Rolling bearing fault diagnosis method based on multilayer noise reduction technology and improved convolutional neural network," *Journal of Mechanical Engineering*, vol. 57, no. 1, pp. 148–156, 2021.

[116] Z. Ye and J. Yu, "Feature extraction of gearbox vibration signals based on multi-channels weighted convolutional neural network," *Journal of Mechanical Engineering*, vol. 57, no. 1, pp. 110–120, 2021.

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

[118] W. Gong, H. Chen, M. Zhang, and Z. Zhang, "Intelligent diagnosis method for incipient fault of motor bearing based on deep learning," *Chinese Journal of Scientific Instrument*, vol. 41, no. 1, pp. 195–205, 2020.

[119] W. You, C. Shen, X. Guo, X. Jiang, J. Shi, and Z. Zhu, "A hybrid technique based on convolutional neural network and support vector regression for intelligent diagnosis of rotating machinery," *Advances in Mechanical Engineering*, vol. 9, no. 6, pp. 168–176, 2017.

[120] X. Li, Q. Ding, and J.-Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural networks," *Reliability Engineering & System Safety*, vol. 172, pp. 1–11, 2018.

[121] X. Li, W. Zhang, Q. Ding, and J.-Q. Sun, "Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation," *Journal of Intelligent Manufacturing*, vol. 31, no. 2, pp. 433–452, 2020.

[122] R. Liu, F. Wang, B. Yang, and S. J. Qin, "Multiscale kernel based residual convolutional neural network for motor fault diagnosis under nonstationary conditions," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 3797–3806, 2019.

[123] Z. Zhuang, H. Lv, J. Xu, Z. Huang, and W. Qin, "A deep learning method for bearing fault diagnosis through stacked residual dilated convolutions," *Applied Sciences*, vol. 9, no. 9, p. 1823, 2019.

[124] L. Su, L. Ma, N. Qin, D. Huang, and A. H. Kemp, "Fault diagnosis of high-speed train bogie by residual-squeeze net," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 3856–3863, 2019.

[125] S. Ma, F. Chu, and Q. Han, "Deep residual learning with demodulated time-frequency features for fault diagnosis of planetary gearbox under nonstationary running conditions," *Mechanical Systems and Signal Processing*, vol. 127, pp. 190–201, 2019.

[126] W. Zhang, X. Li, and Q. Ding, "Deep residual learning-based fault diagnosis method for rotating machinery," *ISA Transactions*, vol. 95, pp. 295–305, 2019.

[127] M. Zhao, M. Kang, B. Tang, and M. Pecht, "Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4290–4300, 2017.

[128] M. Zhao, M. Kang, B. Tang, and M. Pecht, "Multiple wavelet coefficients fusion in deep residual networks for fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 6, pp. 4696–4706, 2018.

[129] A. Malhi, R. Yan, and R. X. Gao, "Prognosis of defect propagation based on recurrent neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 3, pp. 703–711, 2011.

[130] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.

[131] H. Jiang, X. Li, H. Shao, and K. Zhao, "Intelligent fault diagnosis of rolling bearings using an improved deep recurrent neural network," *Measurement Science and Technology*, vol. 29, no. 6, Article ID 065107, 2018.

[132] N. Talebi, M. A. Sadrnia, and A. Darabi, "Robust fault detection of wind energy conversion systems based on dynamic neural networks," *Computational Intelligence and Neuroscience*, vol. 2014, p. 580972, 2014.

[133] M. Mrugalski, M. Luzar, M. Pazera, M. Witczak, and C. Aubrun, "Neural network-based robust actuator fault diagnosis for a non-linear multi-tank system," *ISA Transactions*, vol. 61, pp. 318–328, 2016.

[134] F. O. Heimes, "Recurrent neural networks for remaining useful life estimation," in *Proceedings of the 2008 international conference on prognostics and health management*, pp. 1–6, IEEE, Denver, CO, USA, October 2008.

[135] F.-J. Lin, I.-F. Sun, K.-J. Yang, and J.-K. Chang, "Recurrent fuzzy neural cerebellar model articulation network fault-tolerant control of six-phase permanent magnet synchronous motor position servo drive," *IEEE Transactions on Fuzzy Systems*, vol. 24, no. 1, pp. 153–167, 2015.

[136] Y. Wu, M. Yuan, S. Dong, L. Lin, and Y. Liu, "Remaining useful life estimation of engineered systems using vanilla LSTM neural networks," *Neurocomputing*, vol. 275, pp. 167–179, 2018.

[137] M. Yuan, Y. Wu, and L. Lin, "Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network," in *Proceedings of the 2016 IEEE international conference on aircraft utility systems (AUS)*, pp. 135–140, IEEE, Beijing, China, October 2016.

[138] J. Zhang, P. Wang, R. Yan, and R. X. Gao, "Long short-term memory for machine remaining life prediction," *Journal of Manufacturing Systems*, vol. 48, pp. 78–86, 2018.

[139] R. Zhao, R. Yan, J. Wang, and K. Mao, "Learning to monitor machine health with convolutional bi-directional LSTM networks," *Sensors*, vol. 17, no. 2, p. 273, 2017.

[140] L. Yang, G. Zhiying, Z. Xiaomin, and Z. Qingdong, "Industrial data-driven intelligent forecast for chatter of cold rolling of thin strip with LSTM recurrent neural network," *Journal of Mechanical Engineering*, vol. 56, no. 11, pp. 121–131, 2020.

[141] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1539–1548, 2017.

[142] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[143] Z. Wang, J. Wang, and Y. Wang, "An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition," *Neurocomputing*, vol. 310, pp. 213–222, 2018.

[144] H. Liu, J. Zhou, Y. Xu, Y. Zheng, X. Peng, and W. Jiang, "Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks," *Neurocomputing*, vol. 315, pp. 412–424, 2018.

[145] W. Mao, Y. Liu, L. Ding, and Y. Li, "Imbalanced fault diagnosis of rolling bearing based on generative adversarial network: a comparative study," *IEEE Access*, vol. 7, pp. 9515–9530, 2019.

[146] S. Sun, X. Hu, and Y. Liu, "An imbalanced data learning method for tool breakage detection based on generative adversarial networks," *Journal of Intelligent Manufacturing*, pp. 1–15, 2021.

[147] Y. Xie and T. Zhang, "A transfer learning strategy for rotation machinery fault diagnosis based on cycle-consistent generative adversarial networks," in *Proceedings of the 2018 Chinese Automation Congress (CAC)*, pp. 1309–1313, IEEE, Xi'an, China, December 2018.

[148] X. Li, W. Zhang, and Q. Ding, "Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 7, pp. 5525–5534, 2018.

[149] J. Viola, Y. Chen, and J. Wang, "FaultFace: deep convolutional generative adversarial network (DCGAN) based ball-bearing failure detection method," *Information Sciences*, vol. 542, pp. 195–211, 2021.

[150] J. Dai, J. Wang, Z. Zhu, C. Shen, and W. Huang, "Anomaly detection of mechanical systems based on generative adversarial network and auto-encoder," *Chinese Journal of Scientific Instrument*, vol. 40, no. 9, pp. 16–26, 2019.

[151] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, "Using deep transfer learning for image-based plant disease identification," *Computers and Electronics in Agriculture*, vol. 173, Article ID 105393, 2020.

[152] V. Chouhan, S. K. Singh, A. Khamparia et al., "A novel transfer learning based approach for pneumonia detection in chest X-ray images," *Applied Sciences*, vol. 10, no. 2, p. 559, 2020.

[153] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[154] H. Zheng, R. Wang, Y. Yang, Y. Li, and M. Xu, "Intelligent fault identification based on multisource domain generalization towards actual diagnosis scenario," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 2, pp. 1293–1304, 2019.

[155] Z. Tong, W. Li, B. Zhang, F. Jiang, and G. Zhou, "Bearing fault diagnosis under variable working conditions based on domain adaptation using feature transfer learning," *IEEE Access*, vol. 6, pp. 76187–76197, 2018.

[156] Z. Tong, W. Li, B. Zhang, and M. Zhang, "Bearing fault diagnosis based on domain adaptation using transferable features under different working conditions," *Shock and Vibration*, vol. 2018, 2018.

[157] C. Chen, F. Shen, and R. Q. Yan, "Enhanced least squares support vector machine-based transfer learning strategy for bearing fault diagnosis," *Chinese Journal of Scientific Instrument*, vol. 38, no. 1, pp. 33–40, 2017.

[158] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, "Deep model based domain adaptation for fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 3, pp. 2296–2305, 2016.

[159] 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.

[160] M. J. Hasan and J.-M. Kim, "Bearing fault diagnosis under variable rotational speeds using stockwell transform-based vibration imaging and transfer learning," *Applied Sciences*, vol. 8, no. 12, p. 2357, 2018.

[161] H. Shao, X. Zhang, J. Cheng, and Y. Yang, "Intelligent fault diagnosis of bearing using enhanced deep transfer auto-encoder," *Journal of Mechanical Engineering*, vol. 56, no. 9, pp. 84–91, 2020.

[162] 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.

[163] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2446–2455, 2018.

[164] A. Zhang, H. Wang, S. Li et al., "Transfer learning with deep recurrent neural networks for remaining useful life estimation," *Applied Sciences*, vol. 8, no. 12, p. 2416, 2018.

[165] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Transactions on systems, man, and cybernetics: Systems*, vol. 49, no. 1, pp. 136–144, 2017.

[166] X. Li, W. Zhang, and Q. Ding, "A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning," *Neurocomputing*, vol. 310, pp. 77–95, 2018.

[167] B. Zhang, W. Li, X.-L. Li, and S.-K. Ng, "Intelligent fault diagnosis under varying working conditions based on domain adaptive convolutional neural networks," *IEEE Access*, vol. 6, pp. 66367–66384, 2018.

[168] B. Yang, Y. Lei, F. Jia, and S. Xing, "A transfer learning method for intelligent fault diagnosis from laboratory machines to real-case machines," in *Proceedings of the 2018 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*, pp. 35–40, IEEE, Xi'an, China, August 2018.

[169] B. Yang, Y. Lei, F. Jia, N. Li, and Z. Du, "A polynomial kernel induced distance metric to improve deep transfer learning for fault diagnosis of machines," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 11, pp. 9747–9757, 2019.

[170] W. Qian, S. Li, J. Wang, Y. Xin, and H. Ma, "A new deep transfer learning network for fault diagnosis of rotating machine under variable working conditions," in *Proceedings of the 2018 Prognostics and System Health Management Conference (PHM-Chongqing)*, pp. 1010–1016, IEEE, Chongqing, China, October 2018.

[171] C. Chen, Z. Li, J. Yang, and B. Liang, "A cross domain feature extraction method based on transfer component analysis for rolling bearing fault diagnosis," in *Proceedings of the 2017 29th Chinese Control And Decision Conference (CCDC)*, pp. 5622–5626, IEEE, Chongqing, China, May 2017.

[172] T. Han, C. Liu, W. Yang, and D. Jiang, "Deep transfer network with joint distribution adaptation: a new intelligent fault diagnosis framework for industry application," *ISA Transactions*, vol. 97, pp. 269–281, 2020.

[173] W. Mao, J. He, and M. J. Zuo, "Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 4, pp. 1594–1608, 2019.

[174] M. Naser and A. Alavi, "Insights into performance fitness and error metrics for machine learning," 2020, https://arxiv.org/abs/2006.00887.

[175] A. Botchkarev, "A new typology design of performance metrics to measure errors in machine learning regression algorithms," *Interdisciplinary Journal of Information, Knowledge, and Management*, vol. 14, pp. 045–076, 2019.

[176] Y. Du, A. Wang, S. Wang, B. He, and G. Meng, "Fault diagnosis under variable working conditions based on STFT and transfer deep residual network," *Shock and Vibration*, vol. 2020, 2020.

[177] H. L. Zhen, W. Rixin, Y. Yuantao, Y. Jiancheng, and X. Minqiang, "An empirical analysis about the generalization performance of data-driven fault diagnosis methods," *Journal of Mechanical Engineering*, vol. 56, no. 9, pp. 102–117, 2020.