Machinery New Emerge Fault Diagnosis Based on Deep Convolution Variational Autoencoder and Adaptive Label Propagation

BO SHE AND XUAN WANG
Department of Weaponry Engineering, Naval University of Engineering, Wuhan 430000, China
Corresponding author: Bo She (she1611@126.com)

This work was supported in part by the Natural Science Foundation of Hubei Province under Grant 2019CFB362.

ABSTRACT In the research field of mechanical equipment fault diagnosis, usually only the existing fault types are identified, and the new emerge class of the fault is ignored, however, the new emerge fault class may also occur actually. In order to solve the problem, a novel fault diagnosis model based on deep convolution variational autoencoder network and adaptive label propagation (DCVAN-ALP) is proposed. Firstly, the initial high dimensional features are constructed by using the double tree complex wavelet packet method as the input of the model. Secondly, the convolutional neural network architecture is applied to construct the variational autoencoder, and the local and non-local characteristics of samples are embedded into the loss function for training, which is considered to improve the identification of hidden layer features of the neural network. Finally, t-SNE and the improved label propagation algorithm are adopted to process the hidden features of the neural network, which can achieve the purpose of diagnosing the existing fault class and especially new emerge fault class. Experimental results show that the proposed model can effectively extract the fault characteristics of the vibration signal, and it also has a significantly higher recognition accuracy rate than other typical deep learning methods and traditional classifiers in diagnosing new emerge fault class.

INDEX TERMS Convolution, variational autoencoder, label propagation, new emerge class, fault diagnosis.

I. INTRODUCTION
In industrial process and military fields, mechanical equipment is widely used, once the equipment is damaged, the system reliability will be reduced, which may cause serious impact. In order to improve the safety of industrial process, and mitigate the risks of equipment failure, automatic health management is of great significance, and fault diagnosis plays an important role in equipment health management [1], [2].

The key component of fault diagnosis research is to extract useful information from the data collected by sensors, and then the classifier is applied to obtain the diagnosis result [3], [4]. Traditional fault diagnosis methods on the basis of signal processing and expert knowledge have been applied many years ago [5], and with the development of artificial intelligence, much more intelligent algorithms have been frequently used to build fault diagnosis models in recent years. For instance, Huang et al. [6] proposed RNN–based variational autoencoder (VAE) network to extract features from raw vibration signals for motor fault detection. Jing et al. [7] proposed a convolutional neural network (CNN) to learn features directly from frequency data of vibration signals. In addition, not only one-dimensional (1-D) data like raw vibration signal, but also two-dimensional (2-D) image data or high-dimensional data are also widely served as the input of deep neural network [8]–[10]. Some studies have shown that the preprocessing of the original time-domain signal can improve the effect of the diagnostic model. Wen et al. [11] proposed a signal-to-image conversion method, the time-domain raw signals fulfill the pixels of the image by sequence, and then the input 1-D signals were converted to 2-D images. Xu et al. [12] used the continuous wavelet transform to convert bearing vibration signals into time-frequency images, and then CNN was used to extract intrinsic fault features from the images. Wang et al. [13] converted time-domain vibration signal to 3-D images based on erosion operation, and AlexNet convolutional neural network was used to diagnose the faults of coal washing machine. Zhao et al. [14] adopted discrete wavelet packet analysis to transform the gearbox vibration signal into the 2-D...
time-frequency diagram, which was used as the input of the DRN network for fault diagnosis. Liu et al. [15] constructed high-dimensional feature set as the input of deep auto-encoder (DAE) to identify the fault severities of gearbox, including time domain, frequency domain, and energy spectrum features. Shao et al. [16] developed an adaptive deep belief network (DBN) with dual-tree complex wavelet packet (DTCWPT), and DTCWPT was adopted to refine the measured vibration signals to design a manual feature set from each frequency-band signal. Tang et al. [17] decomposed raw vibration signals into intrinsic mode functions (IMFs) by CEEMD, and converted appropriate IMFs into 2-D images as the input of CNN for low-speed structural fault diagnosis. For various deep neural networks, the neurons in DAE and DBN models generally adopt full connection, while CNN network adopts local connection and weight sharing, which can greatly reduce the training of network parameters and improve the efficiency [18]–[20], in addition, VAE inherits the advantages of auto-encoder in unsupervised learning, while the characteristics of hidden layer features can better represent the input data than auto-encoder [21]. In order to improve the safety of industrial process, some other methods such as transfer learning [22], hybrid method of fault detection and diagnosis [23] are also proposed to reduce risks.

From above studies, although the validity of these models has been proven, the common problem of these fault diagnosis methods is to assume that the classes of target domain samples belong to the source domain samples, in other words, these methods can just diagnose the existing faults. However, the source domain samples cannot contain all fault classes, which means new fault class may emerge with the long-term operation of the mechanical equipment. Thus, the method can not only diagnose existing fault, but also diagnose new emerge fault class, which is more in line with the actual diagnosis requirements. Wang et al. [24] proposed a novel deep metric learning model for imbalanced fault diagnosis and new emerge fault classification, but few samples of the new fault were labeled for training, although the proposed model had achieved effective results, in fact, the real label of the new fault samples cannot be determined in advance.

To overcome the problem, a novel diagnosis model is proposed based on deep convolution variational autoencoder network with adaptive label propagation algorithm. The main contributions of this work are summarized as follows.

1) The loss function of deep convolution variational autoencoder network is constructed by minimizing intra-class compactness (locality) and maximizing inter-class separability (non-locality) in feature space.
2) Label propagation algorithm is improved to adaptively acquire parameter to build neighborhood graph.
3) The proposed DCVAN-ALP model can not only diagnose the existing fault, but also diagnose new emerge fault class in domain adaptation.

The rest of the paper is organized as follows. Section II describes the VAE method. The improved label propagation algorithm is described in Section III. Section IV is dedicated to detail the framework of the proposed DCVAN-ALP model. In Section V, two case studies are given to illustrate the effectiveness of the proposed method. Finally, the conclusion is drawn in Section VI.

II. VARIATIONAL AUTOENCODER ALGORITHM

A. THE STRUCTURE OF VAE

VAE is an unsupervised learning method based on the generation of autoencoder, it inherit autoencoder’s encoder-decoder architecture that is trained to reconstruct its own input, but the difference is that the latent representation z of given data x are replaced with stochastic variables. The basic structure of VAE is shown in Figure 1. VAE forces the distribution of the latent variables z approach to a standard normal distribution, as a result, the latent variables z have stable statistical properties. The idea of VAE is to use the hidden variable z to generate data set x by optimizing the generation parameter θ. In order to make x similar to the original data x with high probability, that is, to maximize the marginal distribution pθ(x).

\[
p_\theta(x) = \int p_\theta(x|z)p_\theta(z) dz
\]

where pθ(x|z) represents the probability distribution of the original data x reconstructed from the hidden variable z, and pθ(z) represents the prior distribution of the hidden variable z. Since the real posterior distribution pθ(z|x) is difficult to be calculated, the approximate posterior qφ(z|x) which obeys Gaussian distribution can be used instead of the real posterior [25].

The objective function of VAE is the variational lower bound to the marginal likelihood of the data, it can be rewritten for individual data points x(i) as follows:

\[
L(\theta, \phi; x^{(i)}) = -D_{KL}(q_\phi(z|x^{(i)}))||p_\theta(z)) + E_{q_\phi(z|x^{(i)})}[\log p_\theta(x^{(i)}|z)]
\]

where DKL is Kullback-Leibler divergence, DKL(q_\phi(z|x^{(i)}))||p_\theta(z) represents the regularization term, and E_{q_\phi(z|x^{(i)})}[\log p_\theta (x^{(i)}|z)] denotes the reconstruction error.

As the reconstruction error term in Eq. (2) requires the Monte Carlo estimate of the expectation, which is not easily differentiable and will cause the network parameters not be updated by the gradient descent method. In order to solve the problem, the reparameterization technique [26] is proposed.
to make variables $z$ reparameterized, let $z = \mu + \epsilon \sigma$, with $\epsilon \in N(0, 1)$. With this trick, a differentiable estimator of the variational lower bound can be calculated and the gradient descent algorithm can be applied to train VAE.

**B. VAE LOSS FUNCTION WITH LOCAL AND NON-LOCAL INFORMATION**

In order to identify the features of different classes more effectively, local and non-local characteristics among samples are embed into the feature representation space, the following objective function is minimized as

$$F = \frac{F_{\text{local}}}{F_{\text{non-local}}}$$  

(3)

In other words, samples of the same class should be clustered in the feature space, and samples of different classes should be scattered in the feature space.

$$F_{\text{local}} = \frac{1}{C} \sum_{i=1}^{C} \frac{1}{n_i} \sum_{j=1}^{n_i} (f^j_i - u_i)^2$$  

(4)

$$F_{\text{non-local}} = \frac{1}{C} \sum_{i=1}^{C} (u - u)^2$$  

(5)

where $C$ denotes the number of classes, $n_i$ is the number of the $i$th class of samples, $f^j_i$ is defined as the feature matrix of the $i$th class of the $j$th sample in the latent $z$ feature space, $u_i$ represents the mean value matrix of the feature matrix $f^j_i$, $u$ denotes the mean value matrix of all classes of sample features.

The loss function of VAE with local and non-local information can be rewritten as

$$J_L = \min(-D_{KL}(q_\theta(z|x)||p_\theta(z)) + E_{q_\theta(z|x)}[\log p_\theta(x|z)] + F)$$  

(6)

**III. THE IMPROVED LABEL PROPAGATION ALGORITHM**

**A. LABEL PROPAGATION ALGORITHM**

In this study, label propagation algorithm is used as a classifier to diagnose faults, including new emerge fault class. The algorithm assumes that if the features of two samples are more similar, the greater the probability that they belong to the same class. Moreover, graph-based label propagation is to transfer the label information of labeled samples to the neighboring nodes through the weight of the graph edge of any node. The label information of unlabeled samples can be derived after reaching the global stable state through several cyclic iterations [27], [28]. Generally, $k$-nearest neighbor method is used to construct the node neighborhood weight graph, the edge with a larger weight is easier to transfer information.

Define the initial sample matrix $X = \{x_i\}_{i=1}^{n}$, the data set $X$ is divided into two nonoverlapped parts, the first $l$ samples $x_i (i \leq l)$ in sample matrix $X$ are taken as labeled samples, and the remaining $n - l$ samples $x_i (l + 1 \leq i \leq n)$ are unlabeled, which means the labels of samples $x_i (i \leq l)$ are known in advance, and the labels of samples $x_i (l + 1 \leq i \leq n)$ are unknown in advance. The purpose of label propagation algorithm is to predict the labels $y_i (l + 1 \leq i \leq n)$ by using samples $x_i (i \leq l)$ and these labels $y_i (i \leq l)$. $C$ is the number of classes, the key part of the algorithm is to assume that the classes of target domain is one more than that of source domain, that is, the number of classes is set to $C + 1$ in target domain. Denote initial label matrix of sample $X$ is $y = [y_1, \ldots, y_l, \ldots, y_{C+1}]^T \in \mathbb{R}^{n \times (C+1)}$, $y_i = [y_{i1}, y_{i2}, \ldots, y_{i,C+1}] \in \mathbb{R}^{C+1}$. For the labeled samples, $y_i^l = 1$ if $x_i (i \leq l)$ is labeled as $j (1 \leq j \leq C)$ and $y_i^l = 0$ otherwise. For the unlabeled samples $x_i (l + 1 \leq i \leq n)$, $y_i^l = 1$ if $j = C + 1$ and $y_i^l = 0$ otherwise. According to the $k$-nearest neighbor method, if $x_i$ and $x_j$ are neighbors, then $x_i$ and $x_j$ are linked by a weight computed by

$$w_{ij} = \begin{cases} 
  e^{-\|x_i - x_j\|^2/\sigma^2} & x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i) \\
  0 & \text{others}
\end{cases}$$  

(7)

Here, $N_k(x_j)$ denotes $k$ neighborhood samples of sample $x_j$, $N_k(x_i)$ denotes $k$ neighborhood samples of sample $x_i$, and $\sigma$ is the variance, in this paper, $\sigma$ is set to be 0.1.

Let $F = [F_1, \ldots, F_n]^T \in \mathbb{R}^{n \times (C+1)}$ be the predicted label matrix of sample $X$, where $F_i \in \mathbb{R}^{C+1} (1 \leq i \leq n)$ and $0 \leq F_{ij} \leq 1$. The label propagation diagram is shown in Figure 2. The blue circle data denote the neighbors of the orange square data $x_i$, and the green triangle denotes the initial label $y_i$. The label information of the data at time $t + 1$ is propagated based on the following equation:

$$F(t+1) = I_o PF(t) + I_o y$$  

(8)

where $I_o$ is an $n \times n$ diagonal matrix with the $i$th entry being $\alpha_i$, and $\alpha_i$ is a parameter for data $x_i$ to balance the initial label information of $x_i$ and the label information received from its neighbors during the iteration. For labeled samples $x_i (i \leq l)$, $\alpha_i$ is desirable to a value close to 0, such as $\alpha_i = 0.1$, while for unlabeled samples $x_i (l + 1 \leq i \leq n)$, $\alpha_i$ is desirable to a value

![FIGURE 2. Label propagation diagram.](image-url)
close to 1, such as $\alpha_i = 0.9$. In addition, $I_p = I - I_d$, $P = D^{-1}w$, where $D$ is the diagonal matrix with the $i$th diagonal element being $D_{ii} = \sum w_{ij}$. According to Eq. (8), we have

$$F(t) = (I_a P)^t F(0) + \sum_{i=0}^{t-1} (I_a P)^i (I_a) y$$

(9)

By the iteration equation (9), the iterative process converges to

$$F = \lim_{t \to \infty} F(t) = (I - I_a P)^{-1}(I - I_a) y$$

(10)

where the sum of each row of $F$ is equal to 1. For the unlabeled sample $x_i \ (l + 1 \leq i \leq n)$, calculating the position of the maximum value in column of $F_y \ (1 \leq j \leq C + 1)$, if $j \leq C$, that’s to say, the unlabeled sample $x_i \ (l + 1 \leq i \leq n)$ is recognized as the known one and belongs to class $j$; if $j = C + 1$, it means that $x_i$ belongs to the new emerge class.

B. THE ADAPTIVE LABEL PROPAGATION ALGORITHM

In $k$-nearest neighbor method, the value of $k$ is usually determined by experience, which will affect the performance of the algorithm. In this paper, an adaptive algorithm based on manifold learning theory is proposed to solve the problem. Compared with Euclidean space, the nonlinear distribution can be better characterized by measuring local geometric features in manifold space. Thus, the manifold curvature index is used to determine the number of neighbors of each sample in manifold space. For sample $x_i$, the manifold curvature $r_i$ is described as follows.

$$r_i = \frac{\sum_{j \in X_i} d_s(x_i, x_j)}{\sum_{j \in X_i} d_s(x_i, x_j)}$$

(11)

where $k$ is initial neighborhood number, $X_i$ represents the neighborhood set of sample $x_i$. $d_s(x_i, x_j)$ denotes the Euclidean distance between sample $x_i$ and its neighborhood sample $x_j$. $d_g(x_i, x_j)$ represents the geodesic distance between sample $x_i$ and its neighborhood sample $x_j$, a detailed geodesic distance calculation method can be found in this reference [29].

The smaller the ratio of Euclidean distance among $X_i$ to its geodesic distance, the more curved the local manifold around sample $x_i$, in order to better characterize the nonlinear properties of sample $x_i$, the neighborhood set should have fewer neighbors. That is, the smaller the value of $r_i$, the smaller the value of $k_i$.

$$k_i = \text{ceil} \left( \frac{k \cdot r_i}{\frac{1}{2} (\min \{r_1, \cdots, r_i, \cdots, r_n\} + \max \{r_1, \cdots, r_i, \cdots, r_n\})} \right)$$

(12)

where the value of $k$ is the initial neighborhood number, $\text{ceil}(a)$ means to take the smallest integer greater than $a$.

IV. PROPOSED METHOD

A. INPUT DATA OF DCVAN-ALP MODEL

Vibration data are processed by multiscale feature extraction method, and high dimensional features are constructed as the input of the proposed model. The steps of vibration data preprocessing are as follows:

1) The raw vibration signals are decomposed to level 4 by using the DTCWPT [30] method, and therefore 16 frequency-band signals are obtained.

2) The 24 time-domain and frequency-domain features are extracted from each of the 16 frequency-band signals. Then the extracted 384 features from each sample are combined to construct the original high dimensional feature set.

In this work, those twenty-four feature parameters [30], [31] are selected as shown in Table 1. The eleven parameters ($p_1 - p_{11}$) are time domain statistical characteristics and the thirteen parameters ($p_{12} - p_{24}$) are frequency-domain statistical characteristics.

B. STRUCTURE OF DCVAN-ALP MODEL

The DCVAN-ALP model consists of three parts, as shown in Figure 3. Stage I is the training and fine-tuning process of the deep convolution variational autoencoder, Stage II is to reduce the dimension of $z$ hidden layer features by t-SNE, and in Stage III, the improved label propagation algorithm is applied to fault diagnosis.

Stage I:

1) Training process

In the encoding stage of variational autoencoder, the structure of convolutional layers is applied to construct the VAE. Conv1 is the first convolutional layer which is connected to the input layer, the symbol $16@4 \times 1$ represents 16 feature surfaces, and the size of the convolution kernel is $(4, 1)$, besides, the stride is $(2, 1)$, which indicates that the sliding step is 2 in the longitudinal direction and 1 in the transverse direction of the feature surfaces. The trick of batch normalization [32] is added right after the convolutional layer of Conv1, that can reduce the shift of internal covariance and accelerate the training process, and Rectified Linear Unit (ReLU) is used as activation unit. Then a max-pooling layer is also added after the convolutional layer of Conv1, and the step size is 2. The output of the first pooling layer is used as the input of Conv2 which is the second convolution layer, similarly, batch normalization and max-pooling are all implemented right after the second convolution layer. The mean value ($\mu$) and logarithm of variance ($\log \sigma^2$) of the hidden layer are output through a full connection layer (Dense) with 100 neurons, and the features ($z$) of the hidden layer are obtained by re-parameterized sampling method. Since VAE is an unsupervised learning method, it needs to reconstruct the input data to complete the training by implementing the decoding process. The decoding process is the reverse operation of the encoding process, and the deconvolution is used to replace the convolution operation. Meanwhile, the local and
TABLE 1. The feature parameters.

| Num | Equation                                                                 |
|-----|--------------------------------------------------------------------------|
| 1   | \( p_1 = \frac{\sum x_i}{N} \)                                         |
| 2   | \( p_2 = \sqrt{\frac{\sum (x_i - p_1)^2}{N-1}} \)                       |
| 3   | \( p_3 = \sqrt{\frac{\sum e_i^2}{N}} \)                                |
| 4   | \( p_4 = \sqrt{\frac{\sum e_i^2}{N}} \)                                |
| 5   | \( p_5 = \max |x_i| \)                                                 |
| 6   | \( p_6 = \frac{\sum (x(i) - p_1)^2}{(N-1)p_1^2} \)                     |
| 7   | \( p_7 = \frac{\sum (x(i) - p_1)^4}{(N-1)p_1^4} \)                     |
| 8   | \( p_8 = \frac{p_3}{p_4} \)                                             |
| 9   | \( p_9 = \frac{p_5}{p_3} \)                                             |
| 10  | \( p_{10} = \frac{P_k}{1 + \sum |x(i)|} \)                              |
| 11  | \( p_{11} = \frac{P_k}{1 + \sum |x(i)|} \)                              |
| 12  | \( p_{12} = \frac{\sum s(k)}{K} \)                                      |
| 13  | \( p_{13} = \frac{\sum (s(k) - p_{12})^2}{K-1} \)                       |
| 14  | \( p_{14} = \frac{\sum (s(k) - p_{12})^4}{Kp_1^4} \)                    |
| 15  | \( p_{15} = \frac{\sum (s(k) - p_{12})^4}{Kp_1^4} \)                    |
| 16  | \( p_{16} = \frac{\sum s(k)}{K} \)                                      |
| 17  | \( p_{17} = \sqrt{\frac{\sum (f_k - p_{16})^2 s(k)}{K}} \)              |
| 18  | \( p_{18} = \sqrt{\frac{\sum f_k^2 s(k)}{K}} \)                         |
| 19  | \( p_{19} = \frac{\sum f_k^2 s(k)}{\sum f_k^2} \)                      |
| 20  | \( p_{20} = \frac{\sum f_k^2 s(k)}{\sum f_k^2} \)                      |

where \( x(i) \) is a signal series for \( n = 1, 2, \ldots, N \), \( N \) is the number of data points.
where \( s(k) \) is a spectrum for \( k = 1, 2, \ldots, K \), \( K \) is the number of spectrum lines; \( f_k \) is the frequency value of the \( k \)-th spectrum line.

FIGURE 3. The structure chart of DCVAN-ALP model.

(2) Fine-tuning process

non-local characteristics of samples are extracted to construct the objective function of VAE.
The hidden layer features ($z$) of training data are extracted as the input of Softmax layer, then, the labels of training data and the output of Softmax layer jointly construct the cross entropy loss function. Thus, the network parameters are updated through reverse fine-tuning to complete the training. For the convenience of comparison, the model proposed in Stage I is named deep convolution variational autoencoder network (DCVAN).

Stage II:
After completing the training of DCVAN model, the hidden layer features ($z$) of training data and testing data can be extracted, respectively. The t-distributed stochastic neighbor embedding (t-SNE) [33] algorithm is adopted to obtain the low dimensional features of training data and testing data.

Stage III:
The low dimensional features of training data and testing data are taken as the input of the label propagation algorithm, to predict the labels of testing data, so as to complete the diagnosis of the known fault and new emerge fault. Additionally, it is also for comparison in this paper, the diagnosis model with the value of parameter $k$ fixed in label propagation algorithm that is named DCVAN-LP, and the diagnosis model embedded with adaptive label propagation algorithm is called DCVAN-ALP.

C. DIAGNOSIS PROCEDURE
The proposed DCVAN-ALP method includes the following procedures.

Step1: the vibration signals of training samples and testing samples are decomposed to level 4 by using DTCWPT method, respectively, and those 24 feature parameters are extracted from each of the frequency-band signals to construct 384 initial features of each sample.

Step2: the initial high dimensional features of training samples and their labels are used as the input of DCVAN model in Stage I, after completing the training and fine tuning processes, then the hidden layer variables ($z$) of training samples and testing samples are extracted, respectively.

Step3: the t-SNE algorithm is applied to reduce the dimensions of the hidden layer variables ($z$) of training samples and testing samples.

Step4: the three-dimensional features of training samples and testing samples are processed by the adaptive label propagation algorithm to predict the labels of testing samples.

V. CASE STUDIES AND EXPERIMENTAL RESULTS
Two case studies are presented in this section to examine and validate the effectiveness of the proposed DCVAN-ALP model. Case 1 focuses on the benchmark data provided by the Case Western Reserve University (CWRU) bearing data center [34]. Case 2 is devoted to the damage data of ammunition swing mechanism from an artillery experimental platform.

A. CASE 1: ANALYSIS OF THE BENCHMARK DATA SET FROM CWRU

1) DATA SOURCES DESCRIPTION
The basic layout of the test rig is shown in Figure 4. The data sets were collected at the sampling frequency of 12kHz under four different loads and speeds (0 hp/1797 rpm, 1 hp/1772 rpm, 2 hp/1750 rpm and 3 hp/1730 rpm). In this study, data sets S1, T1, S2, T2, S3, T3 and S4, T4 are constructed by using the bearing data under the four different working conditions, as shown in Table 2, the purpose is to test the performance of the proposed model in diagnosing new emerge fault class under single working condition or variable working conditions. The training data includes 7 classes, and the testing data contains 8 classes, where 90 samples are taken as training samples of each class, and 25 samples are constructed as testing samples of each class. For instance, S1-T1 means to train the diagnostic model by using the training samples of data set S1 with 7 classes, and to test by using the testing samples of data set T1 with 8 classes under the same working condition (load 0 hp, speed 1797 rpm); S1-T2 means to train the diagnostic model by using the training samples of data set S1 (load 0 hp, speed 1797 rpm) with 7 classes, and to test by using the testing samples of data set T2 with 8 classes under different working condition (load 1 hp, rotating speed 1772 rpm).

To demonstrate the fault diagnosis capability of the proposed model, several methods are applied for comparison, and the input data of all models are the same, the brief introduction of each method is as follows.

(1) SVM: SVM is the representative of traditional classifiers, and is widely used in fault diagnosis.

(2) SDAE: SDAE is the sparse auto-encoder-based deep neural network [35], as a method of unsupervised learning, which is successfully applied for faults classification.

(3) CNN: CNN is an intelligent fault diagnosis method, in this work the model adopts five-layer convolution structure, and the characteristic surfaces of each convolution layer are 16, 32, 64, 64 and 64, respectively, the full connection layer contains 100 neurons, which is connected to the Softmax layer.
TABLE 2. Data set of the experiment.

| Load (hp) | Speed (rpm) | Bearing location | Normal | Inner ring | Outer ring | Rolling element |
|-----------|-------------|------------------|--------|-----------|------------|-----------------|
| 0         | 1797        | S1 Train         | 90     | 90        | 90         | 90              |
|           |             | T1 Test          | 25     | 25        | 25         | 25              |
| 1         | 1772        | S2 Train         | 90     | 90        | 90         | 90              |
|           |             | T2 Test          | 25     | 25        | 25         | 25              |
| 2         | 1750        | S3 Train         | 90     | 90        | 90         | 90              |
|           |             | T3 Test          | 25     | 25        | 25         | 25              |
| 3         | 1730        | S4 Train         | 90     | 90        | 90         | 90              |
|           |             | T4 Test          | 25     | 25        | 25         | 25              |

FIGURE 5. Diagnosis accuracy of different methods on CWRU dataset.

(4) **DCVAN**: DCVAN is proposed in this work. Similarly, Softmax is used as the classifier.

(5) **DCVAN-LP**: The neural network architecture of DCVAN-LP is the same as the proposed DCVAN-ALP model, but the difference is that the parameter $k$ in the label propagation algorithm is a fixed value determined by experience.

As the number of each class in the training samples is the same, the initial value of $k$ in DCVAN-LP and DCVAN-ALP models just takes the value that is less than the number of each class in the training samples. In this case, the initial value of $k$ is set to be 30.

2) EXPERIMENTAL RESULT AND ANALYSIS

Each diagnostic model is run for 10 times to reduce the influence of randomness, and the diagnostic accuracy is averaged as shown in Figure 5 and Table 3, we can find that the proposed DCVAN-ALP model significantly outperforms the other compared models in domain adaptation.

The diagnostic results are basically divided into two parts, one is less than 87.5% and the other is greater than 87.5%. For example, when adapting from Domain S1 to T2, the average accuracy of SVM, SDAE, CNN, DCVAN, DCVAN-LP, DCVAN-ALP reaches 85.5%, 87.5%, 87.5%, 95.6%, 98.9%, respectively. To further show the diagnostic information clearly, the confusion matrix of SVM, DCVAN and DCVAN-ALP is displayed in Figure 6, Figure 7 and Figure 8, respectively. As shown in Figure 6, few samples of inner ring fault and all new emerge fault class are wrongly diagnosed as the outer ring fault, because each diagnostic model is run for 10 times, the average sample size of misdiagnosis is 29 in 10 trials, then the average accuracy is 85.5%. Besides, as shown in Figure 7, all seven existing fault classes can be correctly diagnosed by DCVAN, however, all the new emerge fault class of rolling element are wrongly diagnosed as inner ring fault, then the average accuracy is 87.5%. In contrast, as shown in Figure 8, all the new emerge fault class of rolling element can be correctly diagnosed by DCVAN, however, all the new emerge fault class of rolling element are wrongly diagnosed as inner ring fault, then the average accuracy is 98.9%. Diagnostic result less than 87.5% indicates that the method can not only accurately diagnose the existing fault classes, but also has no ability to recognize new emerge fault class, because the samples of new emerge fault class are all wrongly diagnosed as the existing classes among training data. In other words, traditional classifier and deep learning method such as SVM, SDAE, CNN, DCVAN can not detect the new emerge...
TABLE 3. Diagnostic results of rolling bearing for 8 state classes under various operating conditions (%).

| Methods     | SVM | SDAE | CNN | DCVAN | DCVAN-LP(k=30) | DCVAN-ALP(k=30) |
|-------------|-----|------|-----|-------|----------------|-----------------|
| S1→T1       | 87  | 87.5 | 87.5| 87.5  | 98.55          | 99.45           |
| S1→T2       | 85.5| 87.5 | 87.5| 87.5  | 95.6           | 98.9            |
| S1→T3       | 65  | 86.95| 87.5| 87.5  | 94.1           | 98.4            |
| S1→T4       | 57  | 86.2 | 87.1| 87.2  | 90             | 96.25           |
| S2→T1       | 85  | 87.5 | 87.5| 87.5  | 95.1           | 98.55           |
| S2→T2       | 87  | 87.5 | 87.5| 87.5  | 98.85          | 99.55           |
| S2→T3       | 87  | 87.5 | 87.5| 87.5  | 97.55          | 99.7            |
| S2→T4       | 83.5| 87.1 | 87.4| 87.4  | 94.95          | 98.45           |
| S3→T1       | 63.5| 86.8 | 87.05| 87.25 | 92.1           | 97.7            |
| S3→T2       | 86.5| 87.25| 87.35| 87.45 | 97.95          | 99.5            |
| S3→T3       | 87.5| 87.5 | 87.5| 87.5  | 98             | 99.5            |
| S3→T4       | 86  | 87.15| 87.5| 87.5  | 97.85          | 99.3            |
| S4→T1       | 57  | 86.95| 86.55| 86.85 | 88.05          | 95.05           |
| S4→T2       | 85  | 86.65| 87.1 | 87.3  | 94.35          | 98.85           |
| S4→T3       | 85.5| 87.5 | 87.5| 87.5  | 96.7           | 99.1            |
| S4→T4       | 87  | 87.5 | 87.5| 87.5  | 98.8           | 99.55           |
| Average     | 79.69| 87.19| 87.35| 87.4  | 95.53          | 98.61           |

FIGURE 6. Confusion matrix of SVM in the domain adaptation task of S1 to T2.

FIGURE 7. Confusion matrix of DCVAN in the domain adaptation task of S1 to T2.

fault class. On the contrary, DCVAN-ALP and DCVAN-LP perform better in diagnosing new emerge fault class by using label propagation algorithm.

In order to evaluate the influence of the value of parameter \( k \) in the label propagation algorithm on the diagnosis results, as the number of each class in the training samples is 90, DCVAN-LP model and DCVAN-ALP model are compared while the value of \( k \) varies from 10 to 90. The diagnosis accuracies of DCVAN-LP and DCVAN-ALP in the domain adaptation tasks of S1 to T4 and S4 to T2 can be seen from Table 4. For example, in task S1 to T4, when the initial value of \( k \) is equal to 30, then the parameter \( k_i(k_i \text{ represents the number of neighbors of the } i\text{th sample}) \text{ keeps the value of } 30\text{ unchanged in DCVAN-LP model, and then the average accuracy is } 90\%\text{. However, according to formula (12), the value of parameter } k_i\text{ varies in DCVAN-ALP model, which will help construct domain space for each sample, in fact, the value of parameter } k_i\text{ is } 60, 58, 60, 54, 59, 57, 53, 48, 51 \text{ and so on, respectively, then the average accuracy is } 96.25\%\text{. Moreover, in task S4 to T2, when the initial value of } k \text{ is equal to } 40, \text{ and the parameter } k_i \text{ keeps the value of } 40\text{ unchanged in DCVAN-LP model, then the average accuracy is } 95.1\%.}
However, the value of parameter \( k_i \) in DCVAN-ALP model is 77, 78, 80, 39, 76, 63, 39, 37 and so on, respectively, then the average accuracy is 99.1%. As shown in Figure 9 and Figure 10, although the initial value of parameter \( k \) is determined empirically, the proposed DCVAN-ALP model based on improved label propagation algorithm performs more accurate and stable than DCVAN-LP model in the domain adaptation tasks while the initial value of parameter \( k \) varies. In this case, the initial value of parameter \( k \) in DCVAN-ALP model can be set between 30 and 90, and the the average accuracy is higher than 95%.

For the sake of improving the identification of the features in hidden layer, the local and non-local information are embedded into the loss function of DCVAN-ALP model. In order to better understand the performance of the added local and non-local information, then the hidden layer variables (\( z \)) distribution of testing data are shown in Figure 11 and Figure 12, where the same color of the scatter points represents the samples come from the same class. Obviously, the model considers the local and non-local information can make the samples of the same class gradually gather, and the samples of different class gradually scatter, which will be conducive to fault diagnosis.

**B. CASE 2: ANALYSIS OF THE AMMUNITION SWING MECHANISM DATA SET**

1) DATA SOURCES DESCRIPTION

The experimental data comes from the bench test of ammunition swing mechanism, as shown in Figure 13. Acceleration sensors are arranged near the vulnerable parts such as the press plate and the roller to collect the vibration acceleration
The experimental data include three classes: the normal state of the swing mechanism, the damage of the press plate, and the crack of the roller. In this section, each class data set for training contains 40 samples, and 20 samples for testing, as shown in Table 5. As we can see, data set S1 contains 2 classes: normal and press plate damage, data set T includes 3 classes: normal, press plate damage, and roll crack. For example, S1-T means to train the diagnostic model by using the training samples of data set S1, and to test by using the data set T.

The time domain of the vibration signal in three states of ammunition swing mechanism is shown in Figure 16, and the time domain diagram of vibration signal in the dotted box of Figure 16 is amplified, which can obtain the processes of down swing and up swing in an action cycle as shown for training contains 40 samples, and 20 samples for testing, as shown in Table 5. As we can see, data set S1 contains 2 classes: normal and press plate damage, data set T includes 3 classes: normal, press plate damage, and roll crack. For example, S1-T means to train the diagnostic model by using the training samples of data set S1, and to test by using the data set T.

The time domain of the vibration signal in three states of ammunition swing mechanism is shown in Figure 16, and the time domain diagram of vibration signal in the dotted box of Figure 16 is amplified, which can obtain the processes of down swing and up swing in an action cycle as shown...
in Figure 17, and the impact is huge during the operation of ammunition swing mechanism.

2) EXPERIMENTAL RESULT AND ANALYSIS
As the number of each class in the training samples is the same, the initial value of \( k \) in DCVAN-LP and DCVAN-ALP models just takes the value that is less than the number of each class in the training samples. In this case, the initial value of \( k \) is set to be 30. Each diagnostic model is repeat 10 times to measure average accuracy as shown in Figure 18 and Table 6.

The diagnostic results are basically divided into two parts, one is less than 66.67% and the other is greater than 66.67%. For example, in task S2 to T, the average accuracy of SVM, SDAE, CNN, DCVAN, DCVAN-LP, DCVAN-ALP reaches 33.33%, 66.67%, 66.67%, 66.67%, 96.33%, 97.5%, respectively. To further show the diagnosis information clearly, the confusion matrix of SVM, DCVAN and DCVAN-ALP is displayed in Figure 19, Figure 20 and Figure 21, respectively. As shown in Figure 19, all samples of the normal state and new emerge fault class are wrongly diagnosed as the roller crack state, the average sample size of misdiagnosis is 40 in 10 trials, then the average accuracy is 33.33%. Besides, as shown in Figure 20, the two existing classes can be correctly diagnosed by DCVAN, however, all the new emerge fault class of press plate damage are wrongly diagnosed as the normal state or roller crack state, then the average accuracy is 66.67%. In contrast, as shown in Figure 21, all the new emerge fault class of press plate damage can be correctly diagnosed by DCVAN-ALP, while few samples of the normal state and roller crack state are wrongly diagnosed as the new emerge fault class, the average sample size of misdiagnosis is 1.5 in 10 trials, then the average accuracy is 97.5%.

----

**FIGURE 17.** Time domain diagram of cyclic vibration signal.

**FIGURE 18.** Diagnosis accuracy of different methods.

**TABLE 6.** Diagnostic results of ammunition swing mechanism for 3 state classes (%).

| Methods       | S1→T | S2→T | S3→T | AVG  |
|---------------|------|------|------|------|
| SVM           | 33.33| 33.33| 33.33| 33.33|
| SDAE          | 65.0 | 66.67| 65   | 65.56|
| CNN           | 66.67| 66.67| 65.0 | 66.11|
| DCVAN         | 66.67| 66.67| 65.0 | 66.11|
| DCVAN-LP(k=30)| 98.33| 96.33| 94.33| 96.33|
| DCVAN-ALP(k=30)| 98.33| 97.5 | 96.17| 97.33|

**FIGURE 19.** Confusion matrix of SVM in the task S2 to T.

**FIGURE 20.** Confusion matrix of DCVAN in the task S2 to T.

Diagnostic result less than 66.67% indicates that the method can not only accurately diagnose the existing fault classes, but also has no ability to recognize new emerge fault class. In other words, traditional classifier and deep
learning method such as SVM, SDAE, CNN, DCVAN can not detect the new emerge fault class. However, the proposed DCVAN-LP and DCVAN-ALP models have the ability to detect new emerge fault class, and the performance of DCVAN-ALP is superior than DCVAN-LP.

In order to evaluate the influence of the value of parameter $k$ in the label propagation algorithm on the diagnosis results, as the number of each class in the training samples is 40, DCVAN-LP model and DCVAN-ALP model are compared while the value of $k$ varies from 5 to 40. The diagnosis accuracies of DCVAN-LP and DCVAN-ALP in the tasks of S2 to T and S3 to T can be seen from Table 7. For example, in task S2 to T, when the initial value of $k$ is equal to 10, then the parameter $k_i$($k_i$ represents the number of neighbors of the $i$th sample) keeps the value of 10 unchanged in DCVAN-LP model, and then the average accuracy is 95.67%. However, according to formula (12), the value of parameter $k_i$ varies in DCVAN-ALP model, which will help construct domain space for each sample, in fact, the value of parameter $k_i$ is 12, 18, 20, 14, 16, 5, 10, 19, 17 and so on, respectively, then the average accuracy is 97.17%. Moreover, in task S3 to T, when the initial value of $k$ is equal to 30, and the parameter $k_i$ keeps the value of 30 unchanged in DCVAN-LP model, then the average accuracy is 94.33%. However, the value of parameter $k_i$ in DCVAN-ALP model is 32, 31, 30, 29, 30, 32, 28, 33, 34 and so on, respectively, then the average accuracy is 96.17%. As shown in Figure 22 and Figure 23, the proposed DCVAN-ALP model based on improved label propagation algorithm performs more accurate than DCVAN-LP model while the initial value of parameter $k$ varies. In this case, the initial value of parameter $k$ in DCVAN-ALP model can be

| $k$ | S2→T DCVAN-LP | S2→T DCVAN-ALP | S3→T DCVAN-LP | S3→T DCVAN-ALP |
|-----|---------------|---------------|---------------|---------------|
| 5   | 89.67         | 91            | 85.83         | 88.83         |
| 10  | 95.67         | 97.17         | 90.67         | 93            |
| 15  | 96.83         | 97.67         | 92.33         | 94.17         |
| 20  | 96.67         | 97.67         | 94            | 95.34         |
| 25  | 96.67         | 97.33         | 94.33         | 95.84         |
| 30  | 96.33         | 97.5          | 94.33         | 96.17         |
| 35  | 97.17         | 98            | 94.17         | 95.84         |
| 40  | 96.33         | 97.67         | 94.33         | 95.5          |
| AVG | 95.67         | 96.75         | 92.62         | 94.34         |
set between 20 and 40, and the the average accuracy is higher than 95%.

In order to better understand the performance of the added local and non-local information in DCVAN-ALP model, the task S1 to T is taken as an example, then the visualizations of the hidden layer variables \( z \) distribution of testing data are shown in Figure 24 and Figure 25. Obviously, the model considers the local and non-local information can make the samples of the same class gradually gather, and the samples of different class gradually scatter. However, there are still few samples from different classes overlapped in the feature space as shown in Figure 24, which may lead to misdiagnosis, so the average accuracy is 98.33% as shown in Table 6.

**VI. CONCLUSION**

This study proposed a novel method called DCVAN-ALP for mechanical fault diagnosis. The local and non-local information among samples are embedded into the loss function for training, which is able to compress intra-class samples and separate inter-class samples in feature space. Then, the extracted discriminant features are fed to adaptive label propagation algorithm for fault diagnosis including new emerge fault class. What should be noted is that, no information of the new emerge fault class is provided in advance for training. Thus, compared with the traditional classifier and other deep learning methods, the effectiveness and superiority of the proposed model are verified by bearing fault data under multiple working conditions and damage data of ammunition swing mechanism under a single working condition.

In the future, it is necessary to study how to improve the robustness and the domain adaptability of the model under continuous changing working conditions or imbalanced number of training samples, and reduce the demand for time cost. In addition, we will further investigate to apply this model to other different engineering applications to improve the safety of industrial process and reduce risks.

**ACKNOWLEDGMENT**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

**REFERENCES**

[1] Y. G. Lei, B. Yang, X. W. Jiang, F. Jia, N. P. Li, and A. K. Nandi, “Applications of machine learning to machine fault diagnosis A review and roadmap,” *Mech. Syst. Signal Process.*, vol. 138, pp. 1–39, Jan. 2020.

[2] J. Yu and X. Yan, “A new deep model based on the stacked autoencoder with intensified iterative learning style for industrial fault detection,” *Process Saf. Environ. Protection*, vol. 153, pp. 47–59, Sep. 2021.

[3] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, “Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data,” *IEEE Trans. Ind. Electron.*, vol. 66, no. 9, pp. 7316–7325, Sep. 2019.

[4] W. Chen, J. N. Li, Q. Wang, and K. Han, “Fault feature extraction and diagnosis of rolling bearings based on wavelet thresholding denoising with CEEMDAN energy entropy and PSO-LSSVM,” *Measurement*, vol. 172, pp. 1–13, Feb. 2021.

[5] C. Nan, F. Khan, and M. T. Iqbal, “Real-time fault diagnosis using knowledge-based expert system,” *Process Saf. Environ. Protection*, vol. 86, no. 1, pp. 55–71, Jan. 2008.

[6] Y. Huang, C.-H. Chen, and C.-J. Huang, “Motor fault detection and feature extraction using RNN-based variational autoencoder,” *IEEE Access*, vol. 7, pp. 139086–139096, 2019.

[7] L. Jing, M. Zhao, P. Li, and X. Xu, “A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox,” *Measurement*, vol. 111, pp. 1–10, Dec. 2017.

[8] J. Jiao, M. Zhao, J. Lin, and K. Liang, “A comprehensive review on convolutional neural network in machine fault diagnosis,” *Neurocomputing*, vol. 417, pp. 36–63, Dec. 2020.

[9] Z. Y. Chen, A. Mauricio, W. H. Li, and K. Gryllias, “A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks,” *Mech. Syst. Signal Process.*, vol. 140, pp. 1–13, Jun. 2020.

[10] Y. Zhang, K. S. Xing, R. D. Y. Bai Sun, and Z. Meng, “An enhanced convolutional neural network for bearing fault diagnosis based on time-frequency image,” *Measurement*, vol. 157, pp. 1–22, Jun. 2020.

[11] L. Wen, X. Y. Li, L. Gao, and Y. Y. Zhang, “A new convolutional neural network-based data-driven fault diagnosis method,” *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5590–5598, Jul. 2018.

[12] Y. Xu, Z. X. Li, S. Q. Wang, W. H. Li, T. S. Gyan, and S. Z. Feng, “A hybrid deep-learning model for fault diagnosis of rolling bearings,” *Measurement*, vol. 169, pp. 1–14, Feb. 2021.

[13] Z. Wang, W. Zhao, W. Du, N. Li, and J. Wang, “Data-driven fault diagnosis method based on the conversion of erosion operation signals into images and convolutional neural network,” *Process Saf. Environ. Protection*, vol. 149, pp. 591–601, May 2021.

[14] M. Zhao, M. Kang, B. Tang, and M. Pecht, “Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes,” *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4290–4300, May 2018.

[15] J. Liu, Y. M. H., Y. Wang, B. Wu, J. K. Fan, and Z. X. Hu, “An integrated multi-sensor fusion-based deep feature learning approach for rotating machinery diagnosis,” *Meas. Sci. Technol.*, vol. 23, pp. 1–12, Mar. 2018.

[16] H. Shao, H. Jiang, F. Wang, and Y. Wang, “Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet,” *ISA Trans.*, vol. 69, pp. 187–201, Jul. 2017.

[17] H. Tang, Z. Liao, P. Chen, D. Zuo, and S. Yi, “A novel convolutional neural network for low-speed structural fault diagnosis under different operating condition and its understanding via visualization,” *IEEE Trans. Meas.*, vol. 70, pp. 1–11, May 2021.

[18] Z. Xu, C. Li, and Y. Yang, “Fault diagnosis of rolling bearings using an improved multi-scale convolutional neural network with feature attention mechanism,” *ISA Trans.*, vol. 110, pp. 379–393, Apr. 2021.

[19] Y. S. Zou, Y. Z. Liu, J. L. Deng, Y. L. Jiang, and W. H. Zhang, “A novel transfer learning method for bearing fault diagnosis under different working conditions,” *Measurement*, vol. 171, pp. 1–11, Feb. 2021.

[20] C. C. Che, H. W. Wang, X. M. Ni, and R. G. Lin, “Hybrid multimodal fusion with deep learning for rolling bearing fault diagnosis,” *Measurement*, vol. 173, pp. 1–10, Mar. 2021.

[21] X. R. Ma, Y. Z. Lin, and Z. H. H. Nie Ma, “Structural damage identification based on unsupervised via variational auto-encoder,” *Measurement*, vol. 160, pp. 1–28, Aug. 2020.

[22] C. Chen, F. Shen, J. Xu, and R. Yan, “Domain adaptation-based transfer learning for gear fault diagnosis under varying working conditions,” *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.
[23] M. T. Amin, F. Khan, S. Ahmed, and S. Imtiaz, “Risk-based fault detection and diagnosis for nonlinear and non-Gaussian process systems using R-vine copula,” Process Saf. Environ. Protection, vol. 150, pp. 123–136, Jun. 2021.

[24] C. J. Wang, C. Xin, and Z. L. Xu, “A novel deep metric learning model for imbalanced fault diagnosis and toward open-set classification,” Knowl.-Based Syst., vol. 220, no. 23, pp. 1–14, May 2021.

[25] A. S. Yoon, T. Lee, Y. Lim, D. Jung, P. Kang, D. Kim, K. Park, and Y. Choi, “Semi-supervised learning with deep generative models for asset failure prediction,” 2017, arXiv:1709.00845.

[26] D. P. Kingma and M. Welling, “Auto-encoding variational Bayes,” 2013, arXiv:1312.6114.

[27] F. Nie, S. Xiang, Y. Jia, and C. Zhang, “Semi-supervised orthogonal discriminant analysis via label propagation,” Pattern Recognit., vol. 42, no. 11, pp. 2615–2627, 2009.

[28] S. Jia, X. Deng, M. Xu, J. Zhou, and X. Jia, “Superpixel-level weighted label propagation for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 58, no. 7, pp. 5077–5091, Jul. 2020.

[29] J. B. Tenenbaum, V. de Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” Science, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[30] J. Qu, Z. Zhang, and T. Gong, “A novel intelligent method for mechanical fault diagnosis based on dual-tree complex wavelet packet transform and multiple classifier fusion,” Neurocomputing, vol. 171, pp. 837–853, Jan. 2016.

[31] Y. Lei, Z. He, Y. Zi, and Q. Hu, “Fault diagnosis of rotating machinery based on multiple ANFIS combination with gas,” Mech. Syst. Signal Process., vol. 21, no. 5, pp. 2280–2294, Jul. 2007.

[32] S. Loffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” Proc. 32nd Int. Conf. Mach. Learn., pp. 448–456, 2015.

[33] L. van der Maaten and G. Hinton, “Visualizing data using t-SNE,” J. Mach. Learn. Res., vol. 9, pp. 2579–2605, Nov. 2008.

[34] W. A. Smith and R. B. Randall, “Rolling element bearing diagnostics using the case western reserve university data: A benchmark study,” Mech. Syst. Signal Process., vol. 29, pp. 400–418, Dec. 2015.

[35] W. Sun, S. Shaou, R. Zhaob, R. Yana, X. Zhangc, and X. Chen, “A sparse auto-encoder-based deep neural network approach for induction motor faults classification,” Measurement, vol. 89, pp. 171–178, Jul. 2016.

BO SHE received the Ph.D. degree from the Naval University of Engineering, Wuhan, China, in 2018. He is currently a Lecturer with the Naval University of Engineering. His research interests include signal processing, machinery condition monitoring, and intelligent fault diagnosis.

XUAN WANG received the M.Sc. degree from the Naval University of Engineering, Wuhan, China, in 2020, where he is currently pursuing the Ph.D. degree. His research interests include machinery condition monitoring, intelligent fault diagnostics, and remaining useful life prediction of rotating machinery.