IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON COMPLEX SYSTEM HEALTH MANAGEMENT BASED ON CONDITION MONITORING AND TEST DATA

Received June 19, 2017, accepted July 5, 2017, date of publication July 18, 2017, date of current version August 22, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2728010

Stacked Sparse Autoencoder-Based Deep Network for Fault Diagnosis of Rotating Machinery

YUMEI QI[1], CHANGQING SHEN[1], DONG WANG[2], JUANJUAN SHI[1], XINGXING JIANG[1], AND ZHONGKUI ZHU[1]

[1]School of Urban Rail Transportation, Soochow University, Suzhou 215131, China
[2]Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong

Corresponding author: Changqing Shen (cqshen@suda.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51505311 and Grant 51375322, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20150339, and in part by the Jiangsu Postdoctoral Science Foundation funded project under Grant 1601227C.

ABSTRACT As a breakthrough in the field of machine fault diagnosis, deep learning has great potential to extract more abstract and discriminative features automatically without much prior knowledge compared with other methods, such as the signal processing and analysis-based methods and machine learning methods with shallow architectures. One of the most important aspects in measuring the extracted features is whether they can explore more information of the inputs and avoid redundancy to be representative. Thus, a stacked sparse autoencoder (SAE)-based machine fault diagnosis method is proposed in this paper. The penalty term of the SAE can help mine essential information and avoid redundancy. To help the constructed diagnosis network further mine more abstract and representative high-level features, the collected non-stationary and transient signals are preprocessed with ensemble empirical mode decomposition and autoregressive (AR) models to obtain AR parameters, which are extracted based on the intrinsic mode functions (IMFs) and regarded as the low-level features for the inputs of the proposed diagnosis network. Only the first four IMFs are considered, because fault information is mainly reflected in high-frequency IMFs. Experiments and comparisons are complemented to validate the superiority of the presented diagnosis network. Results fully demonstrate that the stacked SAE-based diagnosis method can extract more discriminative high-level features and has a better performance in rotating machinery fault diagnosis compared with the traditional machine learning methods with shallow architectures.

INDEX TERMS Sparse autoencoder, ensemble empirical mode decomposition, autoregressive model, fault diagnosis.

I. INTRODUCTION

Rotating machinery is the most widely used mechanical equipment in the industrial field and of great importance to the economic development of the society. Faults that occur in rotating machinery may result in huge economic losses and even casualties. In order to monitor the operation condition and improve the security and reliability of the rotating machinery as well as avoid unexpected casualties and economic losses [1], many researchers have devoted to study fault detection and many methods have achieved successful applications. Signal processing and analysis-based methods are successfully used in machine fault diagnosis, such as the wavelet transform (WT), Wigner-Ville distribution, empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), and envelope analysis. Bouzida *et al.* [2] applied discrete WT to extract features from a wide range of frequencies and recognized faults in induction machinery. Yan *et al.* [3] provided a WT review in fault diagnosis. He *et al.* [4] proposed an EMD-based fault signature analysis method for rotating machinery. Feng *et al.* [5] utilized EEMD and Teager energy operator to extract the bearing fault characteristic frequency and then successfully detected the bearing faults. Lei *et al.* [6] provided a detailed description and application of EMD- and EEMD-based methods and their variations in rotating machinery fault diagnosis. These methods based on signal processing and analysis have

achieved great success in fault diagnosis; however, they all rely on a large amount of manpower to extract discriminative features and analyze these features for accurate fault diagnosis, which is time-consuming and requires abundant expertise about signal processing and analysis as well as fault diagnosis. Thus, it is difficult to deal with big datasets. Many intelligent fault diagnosis methods, such as support vector machine (SVM) as well as its variations and artificial neural networks (ANNs), have been successfully utilized in the field of fault diagnosis because of the development of machine learning techniques. Liu *et al.* [7] proposed a kernel joint approximate diagonalization of eigen-matrices-based feature fusion method and used SVM for bearing fault identification. Shen *et al.* [8] introduced a wavelet packet transform (WPT) and an SVM-based fault diagnosis method. Shen *et al.* [9] used multi-class transductive SVM to diagnose faults in gear reducers with artificially extracted features by EMD. Samanta and Nataraj [10] extracted statistical features in the time domain and utilized ANNs and SVM for bearing fault diagnosis. Although these machine learning methods can automatically discriminate faults based on artificially extracted features, which significantly reduce labor consumption for recognition, the shallow architecture limits the capability of automatically learning high-level essential information from the inputs. Thus, traditional machine learning methods highly depend on artificially extracted discriminative features as the inputs. It is difficult to determine the most suitable features to be extracted and different features may directly lead to different diagnosis results, which is time-consuming and unstable [11]. Deep learning is regarded as a significant breakthrough in machine learning, and their deep architectures can be exploited to further learn high-level essential features automatically based on the inputs with great potential [12]. Compared with traditional machine learning methods, the performance of the deep learning methods depends on their learning ability of the complex and non-linear relationship of the inputs instead of the inputs themselves. That is, the inputs are only performing as low-level features regardless whether they are preprocessed or not, and deep learning methods can further extract more discriminative and high-level features automatically based on the inputs. These extracted high-level features are the key to achieve a satisfying fault diagnosis performance. At present, deep learning methods have been successfully applied in speech recognition [13], [14], computer vision [15]–[19], natural language process [20], and medical application [21]. Hannun *et al.* [13] proposed a speech recognition system based on recurrent neural network. Ronao and Cho [15] applied convolution neural network (CNN) to extract robust features automatically and data-adaptively for human activity recognition. Ji *et al.* [16] successfully developed a 3D CNN for human action recognition. Liu *et al.* [17] applied discriminative deep belief networks (DBNs) to classify visual data. Johnson and Zhang [20] successfully applied CNN for text categorization. Koziol *et al.* [21] applied restricted Boltzmann machines to recognize hepatocellular carcinoma. An increasing number of researchers are motivated by the achievements of deep learning methods in the recognition fields aforementioned; they are devoted in applying deep learning methods to the machine fault diagnosis field. Tran *et al.* [22] successfully used DBN to detect faults in valves of reciprocating compressor with Teager-Kaiser energy operator for preprocessing. Tamilselvan and Wang [23] successfully used DBN for the health diagnosis of aircraft and electric power transformer. Jia *et al.* [24] successfully utilized a deep neural network (DNN) based on autoencoders (AE) for fault diagnosis of bearings and planetary gearboxes with the original signals in the frequency domain. Shao *et al.* [25] proposed an AE-based enhancement deep feature fusion method. Li *et al.* [26] combined deep Boltzmann machines and SVM to form a multimodal deep support vector classification method for gearbox fault diagnosis. In addition to recognize the fault types with deep learning methods, Gan *et al.* [27] and Guo *et al.* [28] presented a hierarchical fault diagnosis network respectively based on DBN and CNN to automatically classify the bearing fault types and fault severity.

One of the most important aspects in measuring the features extracted by deep learning methods is whether more input information can be explored and whether redundancy can be avoided to be representative. Thus, this study proposes a diagnosis network based on sparse autoencoder (SAE) with a penalty term that helps mine more abstract and essential features as well as avoid redundancy to be representative. Research shows that parameters of an autoregressive (AR) model reflect the most sensitively to the state change [29]. Thus, AR parameters are extracted as the low-level features for the inputs to help the proposed network further mine high-level features, which are more abstract, representative, and effective for recognition. However, an AR model is unsuitable for non-stationary signals; thus, EEMD is applied in this study to preprocess the collected non-stationary fault signals. Then, an AR model is established based on the intrinsic mode functions (IMFs) decomposed by EEMD. Because fault information is mainly reflected in high-frequency IMFs, only the first four IMFs are considered, and the extracted AR parameters are regarded as the inputs of the proposed diagnosis network, which also helps reduce the dimension of the inputs and simplify the calculation. Compared to EMD, EEMD can overcome mode mixing; thus, EEMD is used in this study for preprocessing.

In summary, this study proposes a diagnosis network based on stacked SAE that combines EEMD and AR models for rotating machinery. The signals are firstly preprocessed by EEMD to obtain IMFs. Subsequently, an optimal AR model based on each IMF is established to obtain AR parameters as the inputs of the diagnosis network. Two cases, namely, bearing and gearbox fault diagnoses, are complemented to validate the performance of the proposed diagnosis network because bearings and gearboxes are vital elements of rotating machinery, which directly influence the operating conditions of rotating machinery.

The rest of the paper is organized as follows. Section 2 introduces the background of EEMD and AR parameter extraction based on EEMD. Section 3 details the structure of the stacked SAE network. The proposed diagnosis procedure is detailed in Section 4. In Section 5, experiments and comparisons are implemented to validate the availability and superiority of the proposed diagnosis network. Important discussions and conclusions are respectively presented in Section 6 and Section 7.

II. EEMD AND AR PARAMETER EXTRACTION BASED ON EEMD

A. EEMD

EEMD decomposes any non-stationary signal $s(t)$ into several IMFs, as depicted in (1).

$$s(t) = \sum_{\varepsilon=1}^{I} c_\varepsilon(t) + r_I \tag{1}$$

where r_I is the residual term, and $c_\varepsilon(t)(\varepsilon = 1, 2 \cdots I)$ are IMFs that represent different frequency components from high to low. The decomposition process must follow the following definitions:

(1) The number of the zero-crossing points and the extreme must either be equal or different at no more than one for all datasets;

(2) The mean value of the envelope, which is defined by the local maxima and local minima, is equal to zero.

B. AR PARAMETER EXTRACTION BASED ON EEMD

The AR model is applicable for linear prediction with parameters that reflect the characteristic of the system and sensitive to condition change. However, the AR model is not applicable in analyzing non-stationary signals. Therefore, EEMD is necessary and the AR model based on the IMFs decomposed by EEMD is established. For a zero-mean IMF $c(t)$, the AR model can be established using (2), where P is the order, $w(t)$ is the white noise with the mean of zero and the variance of σ_w^2 , and a_k $(k = 1, 2 \cdots P)$ are the weighted coefficients.

$$c(t) = \sum_{k=1}^{P} a_k c(t-k) + w(t) \tag{2}$$

The AR parameters based on EEMD are extracted by following three main procedures.

(1) Decompose the original signal with EEMD to get IMFs.

(2) Select the optimal AR model order.

The optimal order is the key foundation in establishing an optimal AR model. A high order implies over-fitting and results in more calculation, whereas a small order cannot reach optimal fitting. The final prediction error (FPE) criterion is applied to determine the optimal AR order P, which is obtained when the criterion obtains the minimum value. N denotes the length of the signal to be modeled, as shown in (3).

$$FPE(P) = \frac{N+P}{N-P}\sigma_w^2 \tag{3}$$

(3) Establish an optimal AR model for each IMF and estimate the AR parameters including the weighted coefficients and one variance.

For each IMF, AR weighted coefficients are calculated with the least square method after determining the AR order P. On the basis of (2), $w(t)$ and σ_w^2 can be written as (4) and (5) respectively. After the AR weighted coefficients $a_k(k = 1, 2 \cdots P)$ have been obtained, the variance σ_w^2 can be calculated using (5). For a P-order AR model, P weighted coefficients and one variance can be obtained and converted into a $(P+1)$-dimensional vector.

$$w(t) = c(t) - \sum_{k=1}^{P} a_k c(t-k) \tag{4}$$

$$\sigma_w^2 = \frac{1}{N-P}\sum_{t=P+1}^{N}(c(t) - \sum_{k=1}^{P} a_k c(t-k))^2 \tag{5}$$

III. BRIEF INTRODUCTION TO THE STACKED SAE-BASED NETWORK

A. SAE

Fig. 1 shows that an AE is an unsupervised feature learning neural network with three layers, namely, the input layer that represents inputs, the hidden layer that represents learned features, and the output layer with the same dimension of the input layer that represents reconstruction. The input and hidden layers form the encoder network responsible for transforming the original inputs into hidden representation codes, whereas the hidden and output layers form the decoder network responsible for reconstructing the original inputs from the learned hidden representation codes.

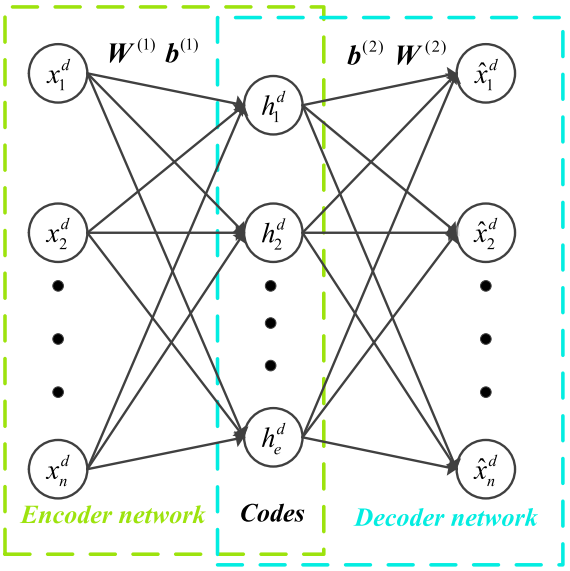


FIGURE 1. Architecture of AE.

For each input vector $\boldsymbol{x}^d$ from datasets $\{\boldsymbol{x}^d\}_{d=1}^{M}$, the representation vector $\boldsymbol{h}^d$ and the reconstructed vector $\hat{\boldsymbol{x}}^d$ can be defined as (6) and (7), respectively, where $\boldsymbol{W}^{(1)}$ and $\boldsymbol{W}^{(2)}$

are the weight matrices, $\boldsymbol{b}^{(1)}$ and $\boldsymbol{b}^{(2)}$ are the bias vectors, and f is the active function. The sigmoid function is used in this study. The reconstruction error $L(\boldsymbol{x}^d, \hat{\boldsymbol{x}}^d)$ between $\hat{\boldsymbol{x}}^d$ and $\boldsymbol{x}^d$ is defined as (8). The overall cost function of the M samples can be defined as (9). The first term of (9) denotes the reconstruction error of the whole datasets, and the second term is the regularization weight penalty term, which aims to prevent over-fitting by restraining the weights magnitude. λ is the weight decay parameter, n_l is the layer number of the network, s_l denotes the neuron number in layer l, and $\boldsymbol{W}_{ji}^{(l)}$ is the connecting weight between neuron i in layer $l+1$ and neuron j in layer l.

$$\boldsymbol{h}^d = f(\boldsymbol{W}^{(1)}\boldsymbol{x}^d + \boldsymbol{b}^{(1)}) \tag{6}$$

$$\hat{\boldsymbol{x}}^d = f(\boldsymbol{W}^{(2)}\boldsymbol{h}^d + \boldsymbol{b}^{(2)}) \tag{7}$$

$$L(\boldsymbol{x}^d, \hat{\boldsymbol{x}}^d) = \frac{1}{2}\left\|\boldsymbol{x}^d - \hat{\boldsymbol{x}}^d\right\|^2 \tag{8}$$

$$J(\boldsymbol{W}, \boldsymbol{b}) = \left[\frac{1}{M}\sum_{d=1}^{M} L(\boldsymbol{x}^d, \hat{\boldsymbol{x}}^d)\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{S_l}\sum_{j=1}^{S_{l+1}}(\boldsymbol{W}_{ji}^{(l)})^2 \tag{9}$$

However, the AE simply copies the inputs; that is, although the learned feature representations may perfectly reconstruct the original inputs, the features are redundant and are not representative enough for classification to some extent. Thus, the cost function of the AE is added with a sparsity penalty term, and the SAE, which has great potential to learn more abstract and representative compressed features of the inputs than the AE, can be obtained. Equation (10) shows the overall cost function of the SAE, where $J(\boldsymbol{W}, \boldsymbol{b})$ is shown before as (9), and the second term is the sparsity penalty term, where $\hat{\rho}_g (g = 1, 2 \cdots e)$ is the average activation value of the hidden unit g, which is defined as (11), ρ is an artificially given small value called the sparsity parameter, and β is the sparsity penalty term parameter used to control the relative importance between the first reconstruction term and the second penalty term.

$$J_{sparse}(\boldsymbol{W}, \boldsymbol{b}) = J(\boldsymbol{W}, \boldsymbol{b}) + \beta\sum_{g=1}^{e}\left(\rho\log\frac{\rho}{\hat{\rho}_g} + (1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_g}\right) \tag{10}$$

$$\hat{\rho}_g = \frac{1}{M}\sum_{d=1}^{M} h_g^d \tag{11}$$

The SAE aims to learn more representative and sparse features, which can remain and extract as much as more information of the inputs instead of simply copying the inputs, by minimizing the cost function $J_{sparse}(\boldsymbol{W}, \boldsymbol{b})$ using the back propagation (BP) algorithm [30]. The optimal parameter sets $\boldsymbol{W}^{(1)}$, $\boldsymbol{W}^{(2)}$, $\boldsymbol{b}^{(1)}$, and $\boldsymbol{b}^{(2)}$ can be learned simultaneously using the minimizing process, which is called the training process of the SAE, as shown in Fig. 2.

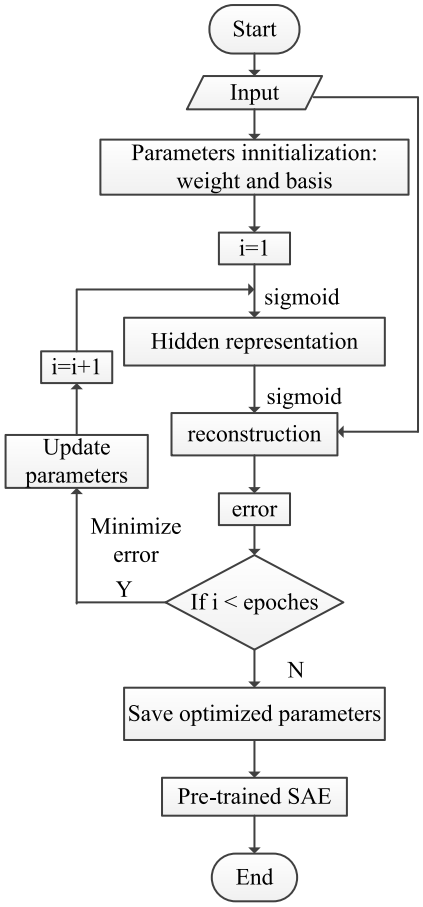


FIGURE 2. SAE training process.

B. SOFTMAX CLASSIFIER

Softmax classifier is commonly used in neural networks for multi-class classification. Suppose that the training datasets can be described as $\{(\boldsymbol{z}^{(1)}, y^{(1)}), \ldots, (\boldsymbol{z}^{(m)}, y^{(m)})\}$, where $\boldsymbol{z}^{(i)} \in R^b, i = 1, 2, \ldots, m$ is one of the input vectors of the softmax classifier and $y^{(i)}, i = 1, 2, \ldots, m$ is the corresponding label. Assume that the training datasets belong to q different classes; thus, the labels have q different values and $y^{(i)} \in \{1, 2, \ldots, q\}$. For each input vector $\boldsymbol{z}^{(i)}$, the hypothesis $h_{\boldsymbol{\theta}}(\boldsymbol{z}^{(i)})$ defined in (12) estimates the probability of $\boldsymbol{z}^{(i)}$ belonging to class v, which is the probability of $y^{(i)} = v, v = 1, 2, \ldots, q$. In (12), $\boldsymbol{\theta} \in R^{q\times b}$ is the parameter matrix of the softmax classifier defined in (13). To describe the error between the target and predicted class labels, the cost function of the softmax classifier $J(\boldsymbol{\theta})$ is defined as (14), where the second term $\frac{\gamma}{2}\|\boldsymbol{\theta}\|_2^2$ is the weight decay penalty term that prevents over-fitting, and γ is the penalty term parameter used to control the relative significance between the first and second terms. Similar to the training process of the SAE, BP optimization algorithm is also used to train the softmax classifier to obtain the optimal parameter matrix $\boldsymbol{\theta}$, and establish the suitable transformation between the inputs and the target label values by minimizing the error between

the target and predicted label values.

$$h_{\theta}(z^{(i)}) = \begin{bmatrix} p(y^{(i)}=1|z^{(i)};\theta) \\ p(y^{(i)}=2|z^{(i)};\theta) \\ \vdots \\ p(y^{(i)}=q|z^{(i)};\theta) \end{bmatrix} = \frac{1}{\sum_{\mu=1}^{q} e^{\theta_{\mu}^{T} z^{(i)}}} \begin{bmatrix} e^{\theta_1^T z^{(i)}} \\ e^{\theta_2^T z^{(i)}} \\ \vdots \\ e^{\theta_q^T z^{(i)}} \end{bmatrix} \tag{12}$$

$$\theta = \left[\theta_1^T \theta_2^T \cdots \theta_q^T\right]^T \tag{13}$$

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{v=1}^{q} 1\{y^{(i)}=v\} \log \frac{e^{\theta_v^T z^{(i)}}}{\sum_{\mu=1}^{q} e^{\theta_{\mu}^T z^{(i)}}}\right] + \frac{\gamma}{2}\|\theta\|_2^2 \tag{14}$$

C. STACKED SAE NETWORK

A stacked SAE network with U hidden layers can be viewed as a stack of U SAEs, as shown in Fig. 3. The input and the first hidden layers are viewed as the encoder network of SAE1, whereas the first and the second hidden layers are viewed as the encoder network of SAE2, and so on. The softmax classifier is added to the output layer of the neural network for classification.

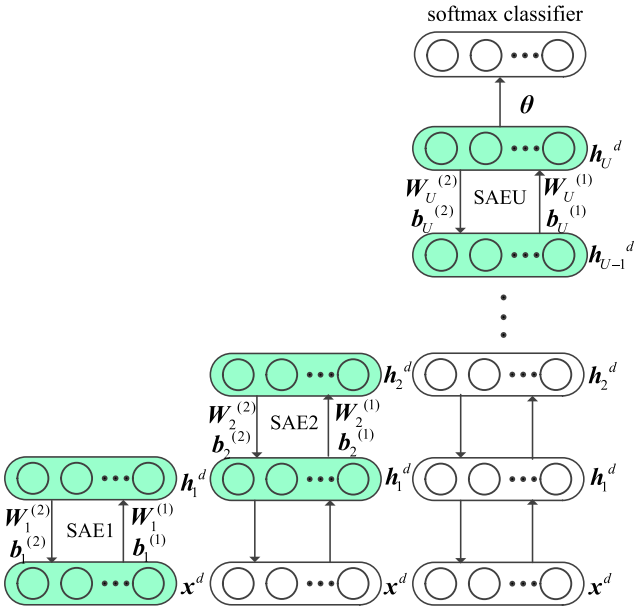


FIGURE 3. Architecture of the stacked SAE network.

Fig. 4 shows the training process of the stacked SAE network, which is composed of two main procedures, namely, pre-training and fine-tuning. The pre-training process aims to pre-train each SAE and the softmax classifier at a time to learn the intra-relationship of each model. First, the BP optimization algorithm is used to train SAE1 as introduced in Part A, and the optimal parameter sets $\boldsymbol{W}_1^{(1)}$, $\boldsymbol{W}_1^{(2)}$, $\boldsymbol{b}_1^{(1)}$ and $\boldsymbol{b}_1^{(2)}$ can be obtained by optimizing the cost function as

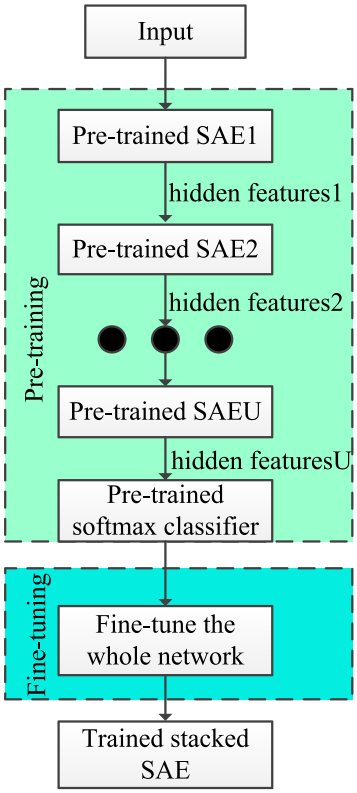


FIGURE 4. Training process of the stacked SAE network.

depicted in (10). After SAE1 is well trained, the first hidden representation code $\boldsymbol{h}_1^d$ can be calculated as (15), which is regarded as the input vector of SAE2. Correspondingly, SAE2 can also be well trained by minimizing the cost function with BP algorithm, and the second hidden representation $\boldsymbol{h}_2^d$ can be calculated as the input vector of SAE3. The process is repeated until the Uth SAE is well trained, and the Uth hidden representation $\boldsymbol{h}_U^d$ can be calculated as (16), which is regarded as the input vector of the softmax classifier. Once $\boldsymbol{h}_U^d$ is obtained, the softmax classifier is pre-trained through supervised learning as introduced in Part B, to obtain the optimal parameter matrix $\boldsymbol{\theta}$ by minimizing the error between the target and predicted class label values. In summary, the pre-training process aims to learn the intra-nonlinear transformation relationship of each single SAE and the relationship between the learned high-level features and target labels, which also helps the network obtain a set of optimal weight initialization for the stacked SAE network compared to the random initialization [31].

$$\boldsymbol{h}_1^d = f(\boldsymbol{W}_1^{(1)}\boldsymbol{x}^d + \boldsymbol{b}_1^{(1)}) \tag{15}$$

$$\boldsymbol{h}_U^d = f(\boldsymbol{W}_U^{(1)}\boldsymbol{h}_{U-1}^d + \boldsymbol{b}_U^{(1)}) \tag{16}$$

Compared to the pre-training process that only considers the intra-relationship of each layer, the fine-tuning process aims to learn the inter-relationship among layers and establish the suitable relationship between the learned

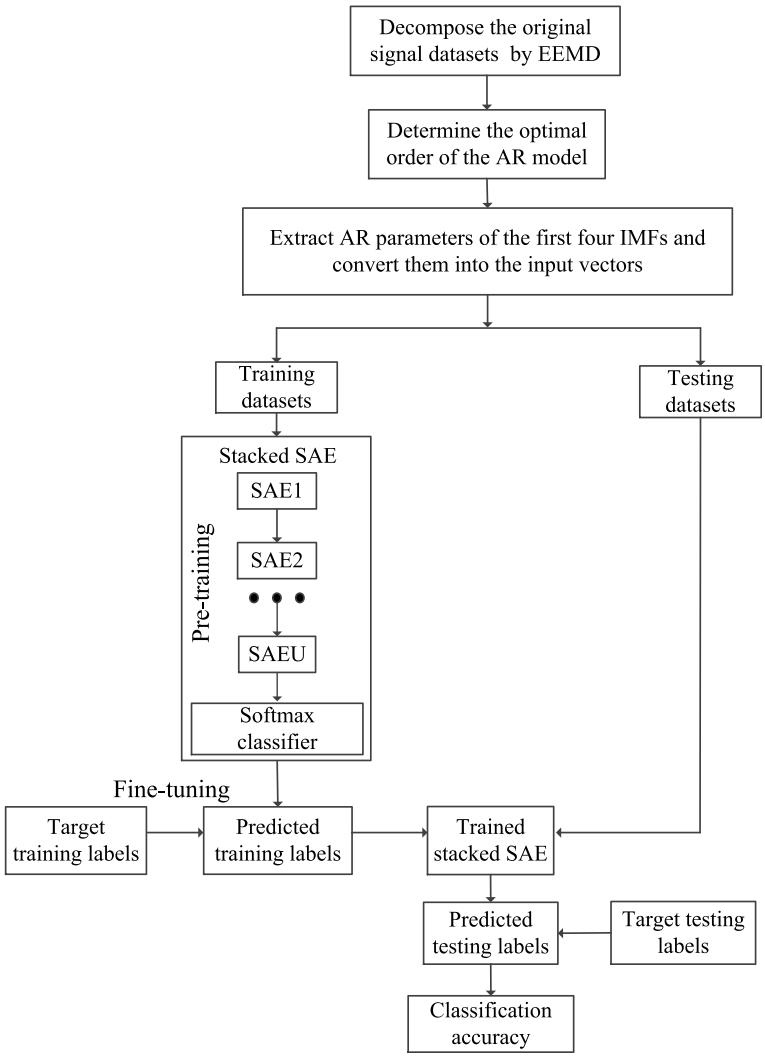


FIGURE 5. Diagnosis flow chart of the proposed stacked SAE network.

Algorithm 1 Fine-Tuning Process

Step 1: Perform a feed forward pass and compute the activation vectors of each layer.
Step 2: Compute the residual error of the network according to the error between the target and practical outputs.
Step 3: Propagate the residual error back from the softmax classifier layer to the input layer.
Step 4: Update the parameters of each layer with the gradient descent.
Step 5: Repeat Steps 1 to 4 until the maximum iteration is achieved.

high-level intrinsic features and the health conditions, which is useful for classification. Algorithm 1 presents the fine-tuning process.

IV. PROPOSED STACKED SAE-BASED FAULT DIAGNOSIS METHOD

Fig. 5 details the flowchart of the proposed stacked SAE-based diagnosis method. The original datasets are initially decomposed into several IMFs. Then, the FPE criterion is used to select the optimal AR model order P and establish the optimal AR model for each IMF. Only the first four IMFs are considered, because fault information is mainly reflected in high-frequency IMFs. Thus, AR parameters, including P weighted coefficients and one variance can be obtained for each IMF. A $4 \times (p + 1)$-dimensional vector can be converted as the input vector of the network for each original signal. Subsequently, the entire input vectors are randomly divided into training and testing two datasets, in which the former is responsible for training the constructed stacked SAE network, whereas the latter is responsible for validating the performance of the stacked SAE-based diagnosis network after being well-trained.

TABLE 1. Detailed information of the bearing datasets.

Health condition	Fault severity (inch)	Training sample size	Testing sample size	Label
IF	0.007/0.014/0.021	400/400/400	200/200/200	1
BF	0.007/0.014/0.021	400/400/400	200/200/200	2
OF	0.007/0.014/0.021	400/400/400	200/200/200	3
N		400	200	4

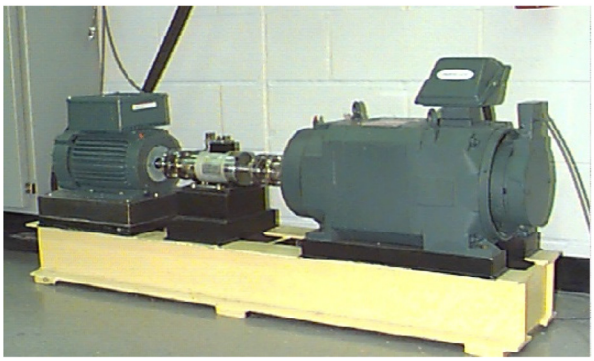


FIGURE 6. Test rig for bearing signals.

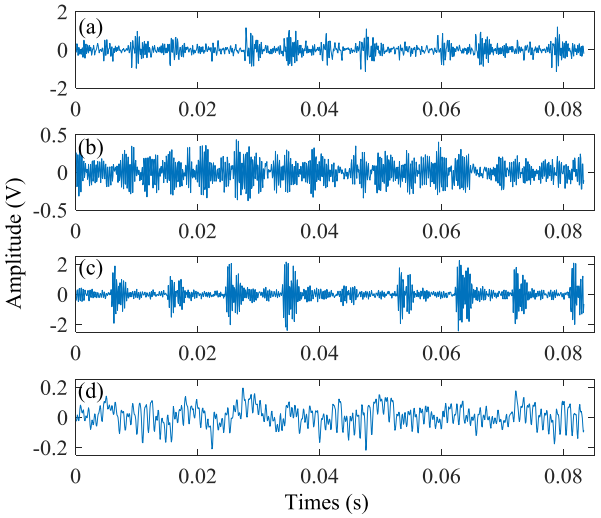


FIGURE 7. Bearing signals of different health conditions in the time domain: (a) IF, (b) BF, (c) OF, (d) N.

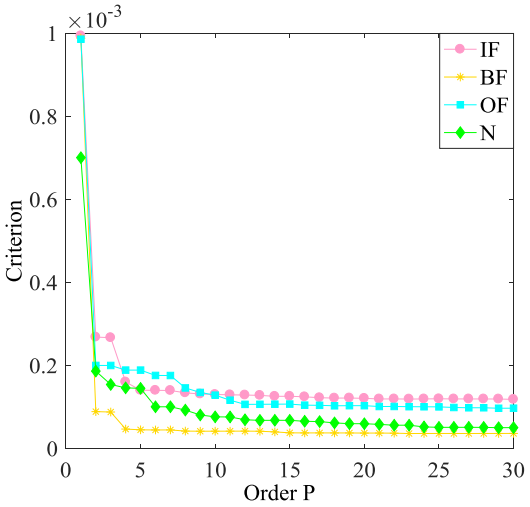


FIGURE 8. Varying curve of the FPE criterion under different health conditions.

V. EXPERIMENTS FOR VALIDATION

A. CASE 1: VALIDATION WITH BEARING DATASETS

1) DATA INTRODUCTION

The bearing datasets obtained from the Case Western Reserve University are used to validate the effectiveness of the constructed stacked SAE-based fault diagnosis network. Fig. 6 shows the test rig.

When the data sampling points are set to 1000 and sampled at 12 KHz, datasets containing four different health conditions, namely inner race fault (IF), ball fault (BF), outer race fault (OF), and normal condition (N) are collected, which are labeled as 1, 2, 3, and 4, respectively. In addition, each type of fault signals contains three fault sizes (0.007, 0.014, and 0.021 inches). A total of 600 samples for each fault size under the same fault type are collected, in which 400 samples randomly selected and the remaining 200 samples are respectively responsible for training and testing the network. Thus, IF, BF, and OF all contain 1800 samples, in which 1200 and 600 samples are respectively regarded as the training and testing samples. A total of 600 samples in normal condition are collected, in which 400 samples randomly selected and the remaining 200 samples are responsible for training and testing. Fig. 7 shows the bearing signals in the time domain with a fault size of 0.007 inch. Table 1 details the bearing datasets information.

2) DATA PREPROCESSING

The non-stationary bearing signals are initially preprocessed with EEMD to obtain several IMFs. An AR model is then established on each of the IMFs to obtain AR parameters as the inputs of the stacked SAE network. Only the first four IMFs are considered in this study because fault information is mainly reflected in high-frequency IMFs. Then, the FPE criterion is applied to determine the optimal AR order. Fig. 8 shows the varying curve of the FPE criterion of the IF, BF, OF, and N considering the first IMF.

TABLE 2. Detailed parameters of the proposed diagnosis network.

No. of hidden layers	No. of hidden layer neurons	No. of input layer neurons
3	100-60-10	100
No. of output layer neurons	Learning rate of each layer	Learning rate of fine-tuning
4	0.3-0.3-0.3-3	0.2

It is noticed that when the order P gets 24, the criterion almost remains unchanged; thus, the optimal order P is selected as 24. Then, a 24-order AR model is established, and 25 AR parameters, including 24 weighted coefficients and one variance are obtained for each IMF utilizing the least square method. For each signal considering the first four IMFs, 100 AR parameters are obtained and converted into the input vector. Compared with the original signals of 1000 dimension, the preprocessed signals of 100 dimension as inputs help reduce the computation significantly.

3) VALIDATION RESULTS

The proposed stacked SAE-based diagnosis network utilized in this case contains three hidden layers, and each hidden layer consists of 100, 60, and 10 neurons, respectively. The input layer consists of 100 neurons representing the 100-dimensional AR parameters, and the output layer has 4 neurons representing 4 different health conditions. The learning rate of each SAE, softmax classifier and the fine-tuning process is 0.3, 0.3, 0.3, 3, and 0.2 respectively. Table 2 shows the detailed parameters of the proposed diagnosis network.

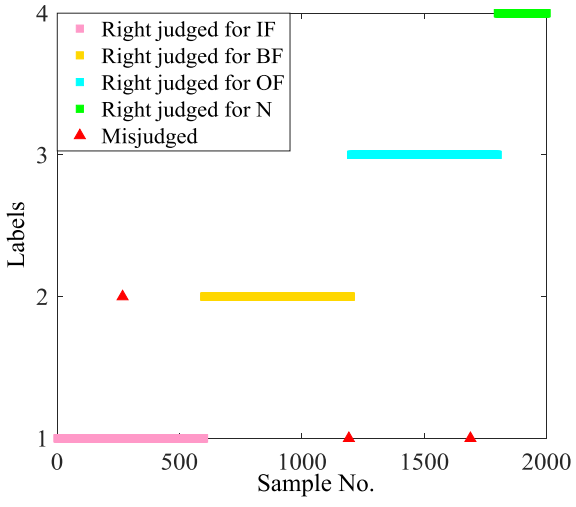


FIGURE 9. Fault classification results of the proposed stacked SAE network for bearing datasets.

Fig. 9 shows the fault classification results of the proposed diagnosis network for bearing datasets when the iterations of each SAE, softmax classifier, and the fine-tuning process is set to be 10000, 10000, 10000, 10000, and 40000. By observation, there are all 3 samples misclassified, including one sample of IF is misclassified to BF, one sample of BF is misclassified to IF, and one sample of OF is misclassified to IF, thereby achieving a diagnosis accuracy of 99.85%.

To further demonstrate the superiority of the proposed stacked SAE-based method in bearing fault detection, another commonly used fault diagnosis methods, such as the stacked AE network with the same architecture of the stacked SAE network, SVM and ANN are applied for comparison. The type of SVM is set to be epsilon-SVR, and RBF kernel function is used. The cross-validation parameter is set as 3. The ANN has one hidden layer, which consists of 100 neurons. The detailed testing accuracy of each method is shown in Table 3. The first four methods are applied with the inputs that are preprocessed by EEMD and AR models or EMD and AR models, whereas the rest of the methods directly handle the original signals in the time domain.

By observation, regardless whether the signals are preprocessed or not, the performances of the SAE or AE-based methods are better than those of the traditional intelligent methods with shallow architectures using the same inputs. Moreover, the superiority is significant, especially when handling the original time-domain signals, which fully validates the ability of deep learning methods with deep architectures in automatically mining discriminative high-level features. In addition, the SAE-based method achieves a higher diagnosis accuracy than the AE-based method, which illustrates the superiority of the sparse penalty term of SAE in mining more representative features that are useful for classification than the AE. Meanwhile, the method based on SAE or SVM dealing with the preprocessed inputs performs better than the method based on the same SAE or SVM dealing with the original time-domain inputs, which demonstrates that the preprocessing based on AR models is able to help mine more representative and useful features. It is noticed that the method based on SAE or SVM using EEMD for preprocessing performs better than that using EMD for preprocessing, which can be explained that EMD has the deficiency of mode mixing, whereas EEMD can overcome this shortcoming. In general, the proposed stacked SAE-based method that applies EEMD and AR models performs the best.

B. CASE 2: VALIDATION WITH GEARBOX DATASETS

1) DATA INTRODUCTION

Another case about gearbox fault detection is analyzed in this study to demonstrate the effectiveness of the presented stacked SAE network. Fig. 10 shows the test rig of gearbox signals.

The speed of the automobile transmission gearbox is 1600 rpm, and the gears run five cycles, namely, running in, normal (N), slight wear (SW), medium wear (MW), and broken tooth (BT), as described in Table 4. Four health condition datasets, including SW, MW, BT, and N labeled as 1, 2, 3, and 4, respectively, can be obtained with the sampling frequency of 3 KHz and a data sampling point of 1000. Fig. 11 shows the different health conditions of gearbox signals in the time domain. Each health condition datasets contain

TABLE 3. Testing accuracy of the comparative methods.

Methods	IF	BF	OF	N	Total
EEMD+AR+SAE	99.83	99.83	99.83	100	99.85
EMD+AR+SAE	99.83	99.50	99.83	100	99.75
EEMD+AR+SVM	98.00	100	97.50	100	98.65
EMD+AR+SVM	97.33	99.83	98.50	97.83	98.05
SAE	87.50	91.33	95.17	100	92.20
AE	85.33	91.83	89.67	100	90.05
SVM	52.00	92.00	32.00	93.00	62.10
ANN	72.33	96.17	67.00	97.00	80.35

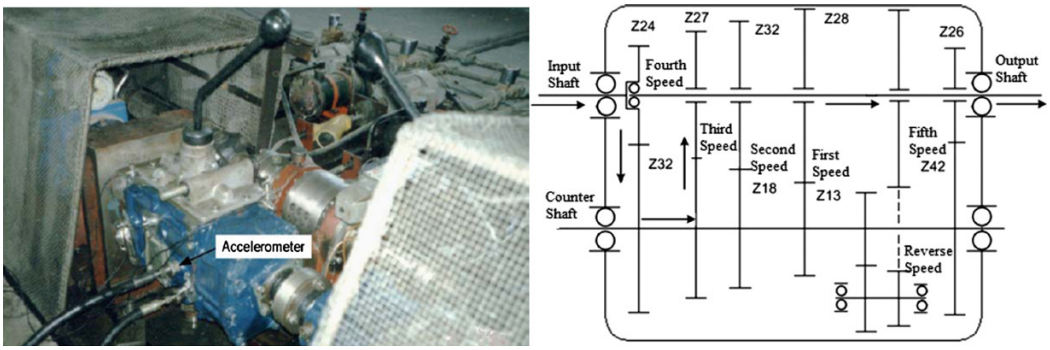


FIGURE 10. Test rig of the gearbox signals.

TABLE 4. Description of the gearbox running cycles.

Running cycle	Health condition	Meshing time (thousand)
1	Running-in	0-700
2	N	700-2800
3	SW	2800-5600
4	MW	5600-6300
5	BT	6300-7000

TABLE 5. Detailed information of gearbox datasets.

Health condition	Training sample No.	Testing sample No.	Label
SW	300	300	1
MW	300	300	2
BT	300	300	3
N	300	300	4

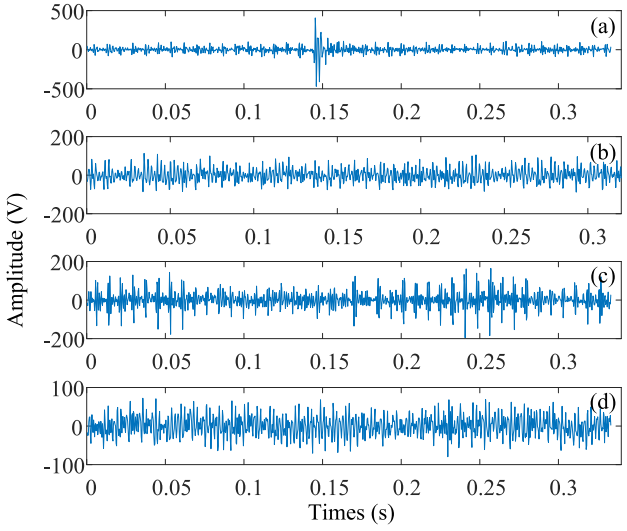


FIGURE 11. Gearbox signals of different health conditions in the time domain: (a) SW, (b) MW, (c) BT, (d) N.

600 samples, in which 300 samples randomly selected and the remaining 300 samples are respectively responsible for training and testing the network, as shown in Table 5.

2) DATA PREPROCESSING

The original gearbox signals in the time domain are also decomposed into several IMFs by EEMD. Then the FPE criterion is applied to determine the optimal order of the AR model and extract the AR parameters as the input vectors of the constructed network. Fig. 12 shows the varying curve of the FPE criterion with the gearbox signals considering the first IMF. The optimal order P is selected as 24 because when the order P gets 24, the criterion almost remains

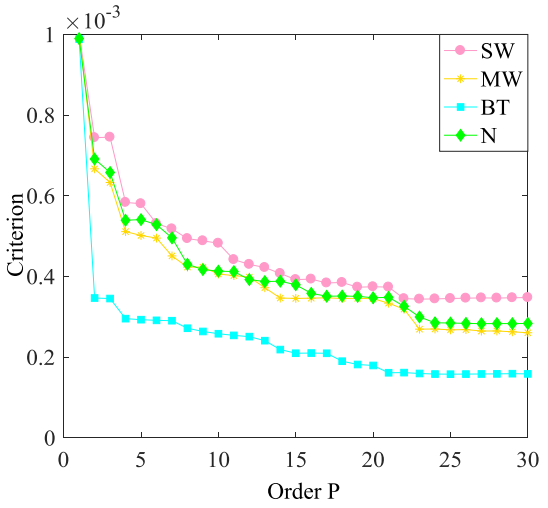


FIGURE 12. Varying curve of the FPE criterion under different health conditions.

unchanged. Then a 24-order AR model is established and 25 AR parameters, including 24 weighted coefficients and one variance are obtained for each IMF. For each signal considering the first four IMFs, 100 AR parameters are obtained and converted into the input vector of the diagnosis network. Compared with the original signals of 1000 dimension, the preprocessed signals of 100 dimension as inputs significantly help reduce the computation.

3) VALIDATION RESULTS

The structure of the constructed stacked SAE-based diagnosis network used in this case is shown in Table 6, which also contains three hidden layers. When the iterations of each SAE, softmax classifier and the fine-tuning process is set as 8000, the fault classification results of the proposed diagnosis network for gearbox datasets is shown in Fig. 13. By observation, there are all two samples misclassified, and are all misclassified to the normal, thereby achieving a diagnosis accuracy of 99.83%.

TABLE 6. Detailed parameters of the proposed diagnosis network.

No. of hidden layers	No. of hidden layer neurons	No. of input layer neurons
3	100-60-10	100
No. of output layer neurons	Learning rate of each layer	Learning rate of fine-tuning
4	0.3-0.3-0.3-0.2	0.2

The same comparative methods used in Case 1 are also applied here to further demonstrate the superiority of the presented method in gearbox fault detection. Table 7 shows the detailed testing accuracy of each method. The first four methods are applied with the inputs preprocessed by EEMD and AR models or EMD and AR models, whereas the rest of the methods directly deal with the original signals in the time domain.

TABLE 7. Testing accuracy of the comparative methods.

Methods	SW	MW	BT	N	Total
EEMD+AR+SAE	99.67	100	99.67	100	99.83
EMD+AR+SAE	98.67	100	100	100	99.67
EEMD+AR+SVM	96.67	100	100	100	99.17
EMD+AR+SVM	96.67	100	99.67	97.67	98.50
SAE	100	100	96.67	100	99.17
AE	100	100	93.33	100	98.33
SVM	78.33	98.67	73.00	100	87.50
ANN	38.00	33.67	36.00	24.00	32.92

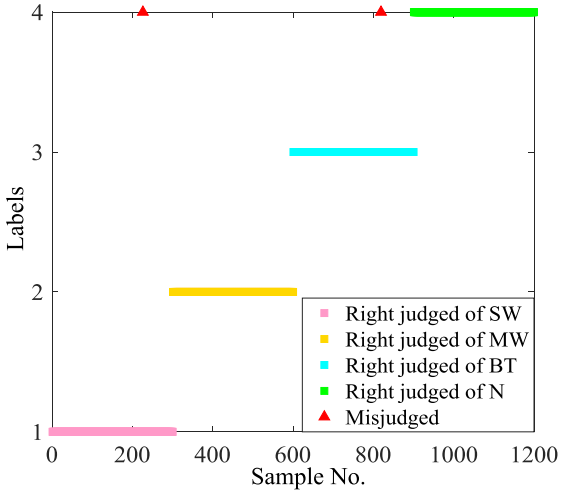


FIGURE 13. Fault classification results of the proposed stacked SAE network for gearbox datasets.

By observation, the comparison results are similar with that shown in Table 3, thereby validating that regardless whether the signals are preprocessed or not, deep learning methods are more superior than traditional intelligent methods with shallow architectures in fault diagnosis due to their ability in mining discriminative high-level features; its superiority is especially significant when handling original time-domain signals. In addition, the sparse penalty term of SAE helps mine more representative features and achieve a higher diagnosis accuracy compared to the AE. Meanwhile, the preprocessing based on AR models is able to help the network learn more essential and useful features compared with the original time-domain inputs. It is noticed that the method based on the SAE or SVM using EEMD for preprocessing performs better than that using EMD for preprocessing, which can be explained that EMD has the deficiency of mode mixing, whereas the EEMD can overcome this shortcoming.

In general, the proposed stacked SAE-based method that applies EEMD and AR models performs the best.

VI. DISCUSSION

A. SELECTION OF THE TRAINING SAMPLES

Comparisons are conducted with different training sample sizes to explore the relationship between the training sam-

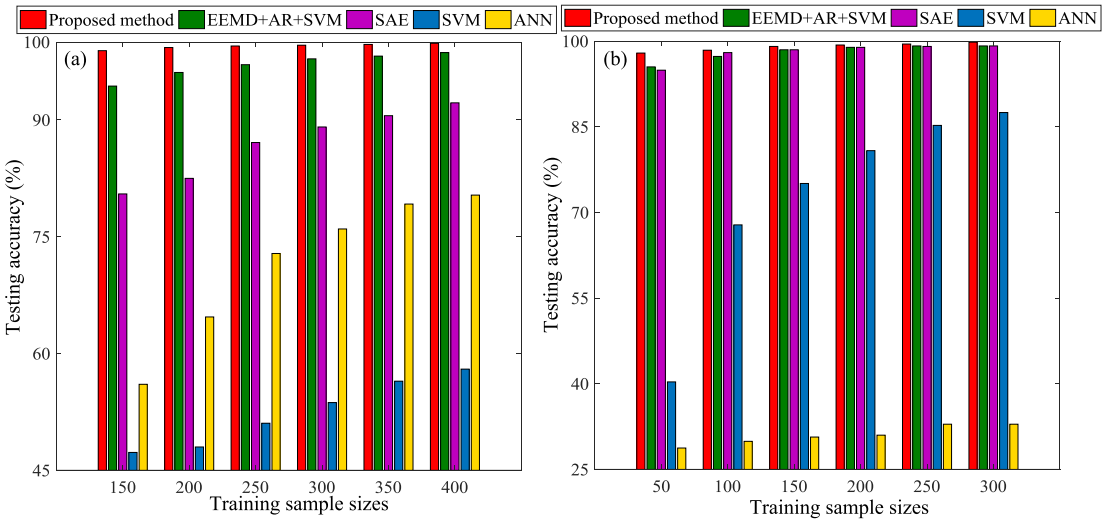


FIGURE 14. Diagnosis accuracy of different methods with different training sample sizes: (a) bearing datasets, (b) gearbox datasets.

ple sizes and the diagnosis accuracy when applying different methods. Fig. 14 shows the testing accuracy of bearing and gearbox datasets with different training sample sizes.

It is shown that, for all comparative methods, the diagnosis accuracy decreases when the size of the training samples decreases, which is mainly because limited training samples cannot provide enough information that are useful for classification. For deep learning method based on SAE, it is difficult to extract as much as more representative information with limited datasets, because SAE is an unsupervised learning network. Thus, it is indicated that the SAE-based deep learning methods are suitable for dealing with large sample size of training datasets. It deserves to be mentioned that, the method based on SAE performs better than that based on SVM or ANN with different training sample sizes using the same inputs, regardless whether these inputs are preprocessed or not. Moreover, the superiority of the SAE-based method is more significant with small training sample sizes. This result demonstrates the ability of SAE-based deep learning method in extracting discriminative features and achieving high diagnosis performance compared with other traditional intelligent methods with shallow architectures.

In general, the proposed method that applies EEMD and AR models for preprocessing always performs the best with different training sample sizes, and it is suggested to select enough training samples to get a better diagnosis performance.

B. SELECTION OF THE SPARSITY PARAMETER

The gearbox datasets are considered as an example to analyze the effect of different sparsity parameter ρ on fault diagnosis accuracy. Table 8 shows the testing accuracy of the SAE with different ρ from 0 to 0.2 with a step of 0.05

TABLE 8. Testing accuracy of the SAE with different sparsity parameter.

Sparsity parameter ρ	0	0.05	0.10	0.15	0.20
Testing accuracy (%)	36.08	99.08	99.17	99.00	98.92

using the original gearbox signals in the time domain. It is shown that the best performance of the SAE is obtained when the sparsity parameter ρ is equal to 0.1, and the accuracy decreases with a smaller or a bigger ρ. This is because a bigger ρ limits the ability of SAE in learning sparse and essential features, whereas a smaller ρ results in excessive hidden neurons are inactive, which makes it difficult for SAE to extract enough useful information from the inputs. Thus, it is of vital importance to select an optimal sparsity parameter ρ for better diagnosis performance and 0.1 is selected as the sparsity parameter in this study. Further work about the selection of the sparsity parameter would be performed in the future.

C. FEATURES VISUALIZATION

Principal component analysis (PCA) is applied to visualize features that are extracted by the proposed diagnosis network and demonstrate its ability in mining high-level representative features that are useful for classification. Fig. 15 shows the first three principal components (PCs) of the original signals in the time domain and the extracted AR parameters for bearing and gearbox datasets. Fig. 16 shows the first three PCs of the features extracted by the proposed method with the bearing and gearbox datasets. In Figs. 16 (a) and 16 (b), the features that represent the same health condition are gathered better, whereas the features that represent different health conditions are separated better compared with the situation in Figs. 15 (a) and (b), as well as in Figs. 15 (c) and (d), which demonstrates the ability of deep

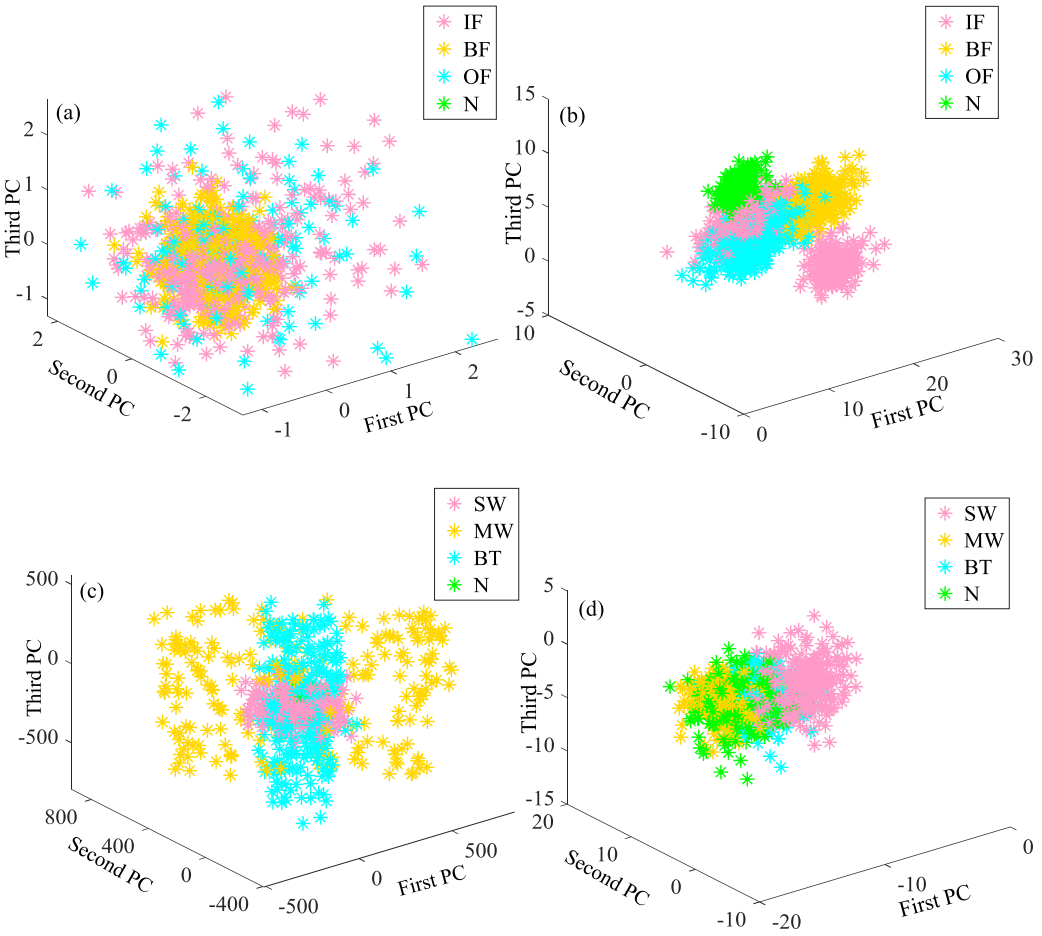


FIGURE 15. Features visualization: (a) raw data features for bearing, (b) AR-based features for bearing, (c) raw data features for gearbox, (d) AR-based features for gearbox.

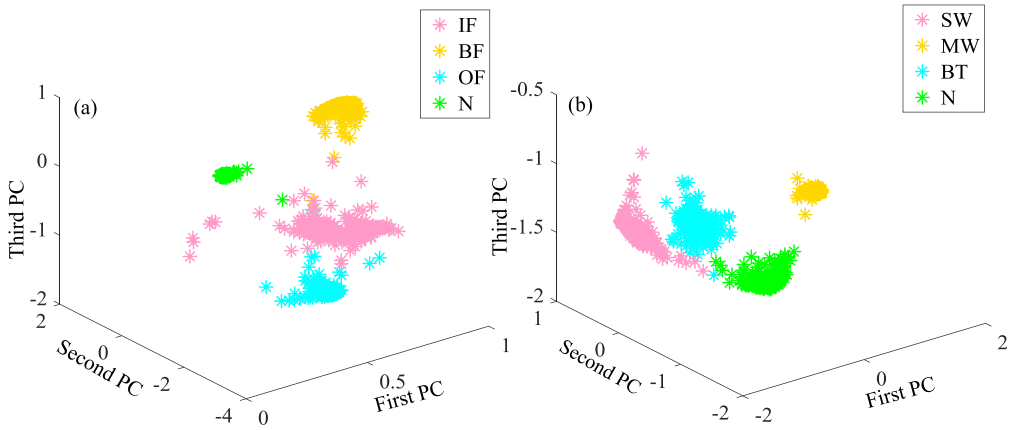


FIGURE 16. Visualization of features mined by the proposed method: (a) bearing datasets, (b) gearbox datasets.

learning methods in automatically mining representative features and explains why the proposed method achieves a higher diagnosis accuracy. It deserves to be mentioned that in Figs. 15 (b) and 15 (d), all features extracted by AR model are mixed and difficult to classify, which fully demonstrates that the excellent diagnosis performance of the proposed method is not mainly due to the preprocessing, but the learning ability of the stacked SAE. The preprocessing just obtains

good low-level features, which help the stacked SAE learn more abstract and discriminative high-level features; these learned high-level features are the key to achieve an excellent diagnosis performance.

Generally, the proposed stacked SAE-based method that combines EEMD and AR models is able to further extract abstract and discriminative high-level features automatically based on the preprocessed inputs. The extracted features can remain as much as more information of the inputs and avoid redundancy, which are effective for fault classification.

VII. CONCLUSION

This study proposed a stacked SAE-based machine fault diagnosis method that combined EEMD and AR models. The time-domain original signals were preprocessed by EEMD and AR models to get AR parameters as inputs of the diagnosis network, which performed as the low-level features of the original signals and helped the stacked SAE mine more abstract and sparse high-level features automatically. The high-level features extracted by the stacked SAE were representative enough for fault classification. In addition, the preprocessing transformed the high-dimensional signals in the time domain into low-dimensional AR parameters, which significantly helped reduce the calculation. Two fault diagnosis cases with bearing and gearbox datasets were successfully analyzed to validate the effectiveness of the proposed method. Meanwhile, the comparisons were complemented using other intelligent fault diagnosis methods to validate the superiority of the proposed method. The validation results fully showed the better diagnosis performance of the proposed method compared with other methods, which demonstrated the ability of deep leaning methods in automatically extracting representative features.

ACKNOWLEDGMENT

The authors would like to thank Professor K. A. Loparo of Case Western Reserve University for his kind permission to use their bearing data.

REFERENCES

[1] W. He, Q. Miao, M. Azarian, and M. Pecht, "Health monitoring of cooling fan bearings based on wavelet filter," *Mech. Syst. Signal Process.*, vols. 64–65, pp. 149–161, Dec. 2015.

[2] A. Bouzida, O. Touhami, R. Ibtiouen, A. Belouchrani, M. Fadel, and A. Rezzoug, "Fault diagnosis in industrial induction machines through discrete wavelet transform," *IEEE Trans. Ind. Electron.*, vol. 58, no. 9, pp. 4385–4395, Sep. 2011.

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

[4] Q. He, Y. Liu, and F. Kong, "Machine fault signature analysis by midpoint-based empirical mode decomposition," *Meas. Sci. Technol.*, vol. 22, no. 1, p. 015702, 2010.

[5] Z. Feng, M. J. Zuo, R. Hao, and J. Lee, "Ensemble empirical mode decomposition-based Teager energy spectrum for bearing fault diagnosis," *J. Vibrat. Acoust.*, vol. 135, no. 3, p. 031013, 2013.

[6] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mech. Syst. Signal Process.*, vol. 35, nos. 1–2, pp. 108–126, 2013.

[7] Y. B. Liu, B. He, F. Liu, S. Lu, and Y. Zhao, "Feature fusion using kernel joint approximate diagonalization of eigen-matrices for rolling bearing fault identification," *J. Sound Vibrat.*, vol. 385, pp. 389–401, Dec. 2016.

[8] C. Shen, D. Wang, F. Kong, and P. W. Tse, "Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier," *Measurement*, vol. 46, no. 4, pp. 1551–1564, 2013.

[9] Z. Shen, X. Chen, X. Zhang, and Z. He, "A novel intelligent gear fault diagnosis model based on EMD and multi-class TSVM," *Measurement*, vol. 45, no. 1, pp. 30–40, 2012.

[10] B. Samanta and C. Nataraj, "Use of particle swarm optimization for machinery fault detection," *Eng. Appl. Artif. Intell.*, vol. 22, no. 2, pp. 308–316, 2009.

[11] L. Chen, Z. Wang, and B. Zhou, "Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification," *Adv. Eng. Inform.*, vol. 32, pp. 139–151, Apr. 2017.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[13] A. Hannun *et al.* (Dec. 2014). "Deep Speech: Scaling up end-to-end speech recognition," pp. 1–12. [Online]. Available: http://arxiv.org/abs/1412.5567

[14] A. van den Oord, S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2013, pp. 2643–2651.

[15] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Syst. Appl.*, vol. 59, pp. 235–244, Oct. 2016.

[16] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.

[17] Y. Liu, S. Zhou, and Q. Chen, "Discriminative deep belief networks for visual data classification," *Pattern Recognit.*, vol. 44, nos. 10–11, pp. 2287–2296, 2011.

[18] A. Toshev and C. Szegedy, "DeepPose: Human pose estimation via deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1653–1660.

[19] R. Girshick, J. Donahue, T. Darrel, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[20] R. Johnson and T. Zhang, "Effective use of word order for text categorization with convolutional neural networks," in *Proc. Hum. Lang. Technol., Annu. Conf. North Amer. Chapter*, Jun. 2015, pp. 103–112.

[21] J. A. Koziol, E. M. Tan, L. Dai, P. Ren, and J.-Y. Zhang, "Restricted Boltzmann machines for classification of hepatocellular carcinoma," *Comput. Biol. J.*, vol. 2014, pp. 1–5, 2014.

[22] V. T. Tran, F. Althobiani, and A. Ball, "An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks," *Expert Syst. Appl.*, vol. 41, no. 9, pp. 4113–4122, 2014.

[23] P. Tamilselvan and P. Wang, "Failure diagnosis using deep belief learning based health state classification," *Rel. Eng. Syst. Safety*, vol. 115, pp. 124–135, Jul. 2013.

[24] 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," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 303–315, May 2016.

[25] H. Shao, H. Jiang, F. Wang, and H. Zhao, "An enhancement deep feature fusion method for rotating machinery fault diagnosis," *Knowl.-Based Syst.*, vol. 119, pp. 200–220, Mar. 2017.

[26] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, and R. E. Vásquez, "Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis," *Neurocomputing*, vol. 168, pp. 119–127, Nov. 2015.

[27] M. Gan, C. Wang, and C. Zhu, "Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 72, pp. 92–104, May 2016.

[28] 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, Nov. 2016.

[29] L. Jiang, Y. Liu, X. Li, and A. Chen, "Degradation assessment and fault diagnosis for roller bearing based on AR model and fuzzy cluster analysis," *Shock Vibrat.*, vol. 18, nos. 1–2, pp. 127–137, 2011.

[30] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, 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.
[31] Y. Liu, X. Feng, and Z. Zhou, "Multimodal video classification with stacked contractive autoencoders," *Signal Process.*, vol. 120, pp. 761–766, Mar. 2016.

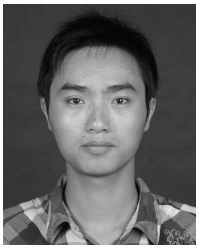
YUMEI QI received the B.S. degree in electrical engineering and automation from Soochow University in 2015, where she is currently pursuing the M.S. degree in measurement techniques and instruments. Her research interests include fault diagnosis and intelligent machine learning.

CHANGQING SHEN received the B.S. and Ph.D. degrees in precision machinery and precision instruments from the University of Science and Technology of China in 2009 and 2014, respectively, and the Ph.D. degree in systems engineering and engineering management from the City University of Hong Kong in 2014. He was appointed as a Research Associate with the City University of Hong Kong in the past several years. He is currently an Associate Professor with the School of Urban Rail Transportation, Soochow University, China. His research interests include mechanical dynamics, machinery data mining, and fault diagnosis.

DONG WANG received the B.S. and M.S. degrees in mechatronics engineering with the University of Electronic Science and Technology of China in 2007 and 2010, respectively, and the Ph.D. degree in systems engineering and engineering management from the City University of Hong Kong in 2015. He was appointed as a Research Assistant, a Research Associate, a Senior Research Assistant, and a Post-Doctoral Fellow with the City University of Hong Kong in the past several years. He is currently a Post-Doctoral Fellow with the City University of Hong Kong. His research interests include mechanical signal processing, machine fault diagnosis, prognostics and health management, statistical modeling, data mining, and non-destructive testing.

JUANJUAN SHI received the B.S. and M.S. degrees in mechanical engineering from Northwest A&F University, Shaanxi, China, in 2008 and 2011, respectively, and the Ph.D. degree in mechanical engineering from the University of Ottawa, Ontario, Canada, in 2015.

From 2012 to 2015, she was a Teaching Assistant with the University of Ottawa. Since 2016, she has been an Associate Professor with the School of Urban Rail Transportation, Soochow University, China. Her research interests include rotating machinery condition monitoring, vibration control, and digital signal processing.



XINGXING JIANG received the B.S. and Ph.D. degrees in vehicle engineering from the Nan University of Aeronautics and Astronautics, Nanjing, China, in 2008 and 2016, respectively. His current research interests include machinery condition monitoring and fault diagnosis, and time-frequency analysis.

ZHONGKUI ZHU received the B.S. degree in automobile and tractor (automobile) and the M.S. degree in vehicle engineering from Hefei Polytechnic University in 1997 and 2002, respectively, and the Ph.D. degree in instrument science and technology from the University of Science and Technology of China in 2005. He is currently a Professor with the School of Urban Rail Transportation, Soochow University. His research interests include fault diagnosis of mechanical equipment, vehicle system dynamics and control, and vibration measurement and signal processing.