

Modified Stacked Auto-encoder Using Adaptive Morlet Wavelet for Intelligent Fault Diagnosis of Rotating Machinery

Haidong Shao, Min Xia, *Member IEEE*, Jiafu Wan, and Clarence W. de Silva, *Fellow, IEEE*

Abstract—Intelligent fault diagnosis techniques play an important role in improving the abilities of automated monitoring, inference, and decision-making for the repair and maintenance of machinery and processes. In this paper, a modified stacked auto-encoder (MSAE) that uses adaptive Morlet wavelet is proposed to automatically diagnose various fault types and severities of rotating machinery. Firstly, the Morlet wavelet activation function is utilized to construct an MSAE to establish an accurate nonlinear mapping between the raw nonstationary vibration data and different fault states. Then, the nonnegative constraint is applied to enhance the cost function to improve sparsity performance and reconstruction quality. Finally, the fruit fly optimization algorithm (FOA) is used to determine the adjustable parameters of the Morlet wavelet to flexibly match the characteristics of the analyzed data. The proposed method is used to analyze the raw vibration data collected from a sun gear unit and a roller bearing unit. Experimental results show that the proposed method is superior to other state-of-the-art methods.

Index Terms—Modified stacked auto-encoder, Intelligent fault diagnosis, Adaptive Morlet wavelet, Nonnegative constraint, Fruit fly optimization.

I. INTRODUCTION

ROTATING machinery has been widely applied in high-speed trains, wind turbines, helicopters, and many other important industrial equipment [1]. During the long-term operation under harsh conditions such as severe impacts, high speeds, and excessive loading, the key components of rotating machinery, such as bearings and gears, usually become susceptible to various types of faults [2]. Aiming at automatically distinguishing the fault categories and severities, intelligent fault diagnosis techniques play an important role in

improving the abilities of auto-monitoring, inference, and decision-making of rotating machinery [3-5].

Vibration analysis is still the most popular approach to the health monitoring of rotating machinery [6-11]. Although the conventional intelligent diagnosis framework based on shallow learning models and vibration analysis has been studied for decades, it cannot avoid the tedious feature extraction and selection that is relied on rich domain experience and knowledge [12-13]. To overcome this inherent limitation, a new trend has emerged in the past several years where deep learning techniques are applied to the intelligent diagnosis of equipment, including stacked auto-encoder (SAE), deep belief network (DBN), convolutional neural network (CNN), and long short-term memory (LSTM) [14-20]. Compared with DBN, CNN, and LSTM, SAE possesses the properties of unsupervised learning, high-efficiency training, and easy implementation [21, 22], and has wide application in different fields, such as fault diagnosis, image classification, data denoising, and feature reduction.

However, the challenges exist when applying the basic SAE to the practical fault diagnosis task of rotating machinery. On one hand, various fault categories and severities may reduce the distinguishable characteristic differences hidden in the raw vibration data [23]. On the other hand, for some rotating machines such as planetary gearboxes, the coupled vibration of multiple components and complicated transmission paths will cause stronger non-stationarity and increased interference of the collected vibration signals [24]. Neural networks designed with general activation functions may have limitations in establishing accurate mappings between a nonlinear and nonstationary input data and various output patterns [25, 26]. Unlike the popular transformation functions, wavelet function has a special attribute of time-frequency localization. Inspired by the successful application of wavelet neural networks (WNNs) in fault diagnosis [27], there exists a strong motivation to modify the basic SAE by using a wavelet function. Because of the strong similarity to periodic impulse components of mechanical vibration signals [28, 29], Morlet wavelet has been successfully utilized to modify the basic SAE in recent work [25, 30]. However, the two waveform parameters of Morlet wavelet activation function are manually selected and fixed, which cannot flexibly match the characteristics of the analyzed data. Besides, Morlet wavelets with different parameters probably show different performance when dealing with the

Manuscript received May 11, 2020; revised October 8, 2020; accepted January 16, 2021. This work was partially supported by the EU Horizon 2020 (101007005), the National Natural Science Foundation of China (51905160), and The Joint Fund of the National Natural Science Foundation of China and Guangdong Province (No. U1801264). (*Corresponding author: Min Xia.*)

H. Shao is with the State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, College of Mechanical and Vehicle Engineering, Hunan University, China (e-mail: hdshao@hnu.edu.cn)

M. Xia is with the Department of Engineering, Lancaster University, United Kingdom (e-mail: m.xia3@lancaster.ac.uk).

J. Wan is with the Provincial Key Laboratory of Technique and Equipment for Macromolecular Advanced Manufacturing, South China University of Technology, China (e-mail: mejwan@scut.edu.cn)

C. W. de Silva is with the Department of Mechanical Engineering, The University of British Columbia, Canada (e-mail: desilva@mech.ubc.ca).

analyzed data [29]. Thus, there exists a strong motivation to design an adaptive Morlet wavelet through flexibly adjusting the two parameters to achieve the best match with the characteristics of the analyzed signals. The fruit fly optimization algorithm (FOA) can effectively carry out a global optimization with many advantages including faster convergence, stronger stability, higher precision, and easier implementation than many other optimization algorithms [31]. In the last two years, FOA has been gradually used to optimize the intelligent diagnosis model of rotating machinery and has shown good performance [32]. Thus, modified SAE (MSAE) with adaptive Morlet wavelet using FOA has the potential to properly match the characteristics of the analyzed data in different diagnosis tasks.

The other challenge is due to the increase of the width and the depth of the SAE model, which would require updating of a larger amount of weights, making it harder to train. Even though a weight decay term is usually added to the cost function of SAE to avoid over-fitting, numerous non-zero connection weights will lead to a reduction in sparsity and affect reconstruction quality [33, 34]. Thus, more effective weight decay strategies are crucial.

To address these challenges, in the present paper, MSAE that uses an adaptive Morlet wavelet and improved training algorithm is proposed to automatically diagnose various faults of rotating machinery. The developed method is verified by analyzing the raw vibration data collected from a sun gear unit and a roller bearing unit. Experimental results demonstrate the superiority of the proposed method over other state-of-the-art deep learning methods. The main contributions of this paper are:

- (1) Establishes an accurate nonlinear mapping between the raw nonstationary vibration data and various fault states through MSAE which is constructed by improving the basic SAE with Morlet wavelet activation function.
- (2) Achieves high-quality reconstruction with improved cost function by incorporating a nonnegative constraint in the algorithm.
- (3) A FOA-based method is developed to adaptively determine the adjustable parameters of Morlet wavelet activation function, which achieves a flexible matching of the characteristics of the analyzed data.
- (4) The developed approach can work directly on raw sensing data with noise and achieve more accurate fault diagnosis results on rotating machinery than state-of-art deep learning-based approaches.

The rest of the paper is organized as follows. Section II briefly reviews the basic auto-encoder. The proposed method is introduced in Section III. In Section IV, the effectiveness of the proposed method is evaluated by two experiments with vibration signals obtained from a sun gear unit and a roller bearing unit. The conclusions and future work are presented in Section V.

II. THE PRINCIPLE OF AUTO-ENCODER

Auto-encoder (AE) belongs to unsupervised deep learning-based models, and its architecture is shown in Fig. 1. The training goal of AE is to achieve the reconstruction of the inputs as accurately as possible by adjusting the model parameters. The main formulas of AE are as follows [22]:

$$\mathbf{h} = s_g(\mathbf{W}\mathbf{x} + \mathbf{b}) \quad (1)$$

$$\mathbf{z} = s_f(\mathbf{W}'\mathbf{h} + \mathbf{b}') \quad (2)$$

$$C_1 = \frac{1}{2} \sum_{i=1}^m (z_i - x_i)^2 + \beta \left(\sum_{j=1}^p r \log \frac{r}{\hat{r}_j} + (1-r) \log \frac{1-r}{1-\hat{r}_j} \right) \quad (3)$$

where C_1 is the cost function of AE, β is a sparse penalty coefficient, r is a sparse constant, $\mathbf{x} = [x_1, x_2, \dots, x_m]$ is an unlabeled input sample whose feature vector and the reconstruction vector are $\mathbf{h} = [h_1, h_2, \dots, h_p]$ and $\mathbf{z} = [z_1, z_2, \dots, z_m]$, respectively, s_g and s_f are the activation functions in the hidden layer and the output layer, respectively, which are generally selected as Sigmoid (Sigm) or rectified linear unit (ReLU) activation functions, \mathbf{W} , \mathbf{W}' are weights, and \mathbf{b} , \mathbf{b}' are biases.

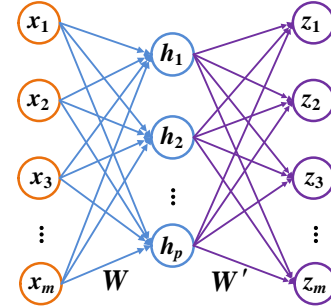


Fig. 1. The architecture of the AE model.

III. THE PROPOSED METHOD

A. MSAE construction

Base on the idea of wavelet transform, wavelet function has been a new choice of the activation functions applied in the neural network, which can make full use of the time-frequency localization characteristics. To establish an accurate nonlinear mapping between the collected nonstationary vibration data and various working states, this paper uses Morlet wavelet as the activation function of the hidden layer of the basic AE due to its greater similarity to the fault characteristic components hidden in the collected vibration signal than other wavelets [28, 29]. Morlet wavelet is expressed as

$$\psi(t) = \frac{1}{\sqrt{f_b \pi}} \cos(2\pi f_c t) \exp(-t^2 / f_b) \quad (4)$$

in which f_b is the bandwidth, f_c is the central frequency, and these two parameters influence the performance of Morlet wavelet function. Based on the Morlet wavelet activation

function, a new version of AE model called modified AE (MAE) is constructed as shown in Fig. 2.

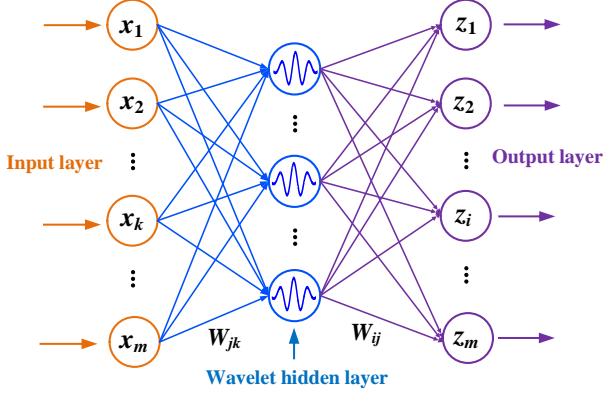


Fig. 2. MAE based on Morlet wavelet function.

For the input sample $\mathbf{x} = [x_1, x_2, \dots, x_m]$, the expression of the hidden layer output is given by

$$h_j = \frac{1}{\sqrt{f_b \pi}} \cos\left(2\pi f_c \left(\sum_{k=1}^m W_{jk} x_k - c_j\right) / d_j\right) \cdot \exp\left(-\left(\sum_{k=1}^m W_{jk} x_k - c_j\right) / d_j\right)^2 / f_b\right) \quad (5)$$

where h_j is the output of the hidden node j , d_j and c_j are the scale factor and the shift factor, respectively, W_{jk} is the weight between the hidden node j and the input node k , and W_{ij} is the weight between the hidden node j and the output node i . Set the nonlinear transformation of the output layer as a tanh function. Then the final reconstructed output is

$$z_i = s_g \left(\sum_{j=1}^p W_{ij} h_j\right) = \tanh\left(\sum_{j=1}^p W_{ij} h_j\right) \quad (6)$$

To avoid over-fitting, a weight decay term can be added to the cost function, as

$$C^T = \frac{1}{2} \sum_{i=1}^m (z_i - x_i)^2 + \frac{\lambda}{2} \sum_{i,k=1}^m \sum_{j=1}^p \left((W_{ij})^2 + (W_{jk})^2\right) + \beta \left(\sum_{j=1}^p r \log \frac{r}{\hat{r}_j} + (1-r) \log \frac{1-r}{1-\hat{r}_j}\right) \quad (7)$$

where λ is a weight decay factor and C^T is the traditional cost function. However, a large number of connecting weights in this decay strategy will lead to a reduction in sparsity. The nonnegative constraint of the connecting weights was first proposed in 2016 to further improve sparsity and reconstruction quality by reducing the negative weights [33]. In the last two years, SAEs integrated with the nonnegative constraint have been gradually used for the fault diagnosis of rotating machinery [35], and their diagnosis results are better than those using a conventional weight decay term in Eq. (7). Here, to achieve higher quality reconstruction, the nonnegative constraint is introduced into the cost function of the MAE, as

$$C^E = \frac{1}{2} \sum_{i=1}^m (z_i - x_i)^2 + \frac{\delta}{2} \sum_{L=1}^2 \sum_{l=1}^{s_L} \sum_{j=1}^{s_{L+1}} G(W_{jl}^{(L)}) + \beta \left(\sum_{j=1}^p r \log \frac{r}{\hat{r}_j} + (1-r) \log \frac{1-r}{1-\hat{r}_j}\right) \quad (8)$$

$$G(W_{jl}^{(L)}) = \begin{cases} (W_{jl}^{(L)})^2, & \text{if } W_{jl}^{(L)} < 0 \\ 0, & \text{if } W_{jl}^{(L)} \geq 0 \end{cases} \quad (9)$$

where the second term represents the nonnegative constraint, δ represents a penalty coefficient, C^E represents the enhanced cost function, and s_L represents the node dimension in layer L . The training task of the MAE is also to adjust the weights $W_{jl}^{(L)}$ so as to make C^E a minimum. Gradient descent with back propagation is a simple and fast way to update the weights as follows

$$W_{jl}^{(L)} = W_{jl}^{(L)} - \eta \frac{\partial C^M}{\partial W_{jl}^{(L)}} \quad L=1,2 \quad (10)$$

$$\frac{\partial C^M}{\partial W_{jl}^{(L)}} = \frac{\partial C_1}{\partial W_{jl}^{(L)}} + \delta g(W_{jl}^{(L)}) \quad (11)$$

$$g(W_{jl}^{(L)}) = \begin{cases} W_{jl}^{(L)}, & W_{jl}^{(L)} < 0 \\ 0, & W_{jl}^{(L)} \geq 0 \end{cases} \quad (12)$$

where η denotes the learning rate, $W_{jl}^{(1)} = W_{jk}$, $W_{jl}^{(2)} = W_{ij}$.

MMAE with stacked trained MAEs can further capture the valuable features hidden in the input samples, as shown in Fig. 3. After that, the learned deep features are used as the input of the *Softmax* classifier for fault classification.

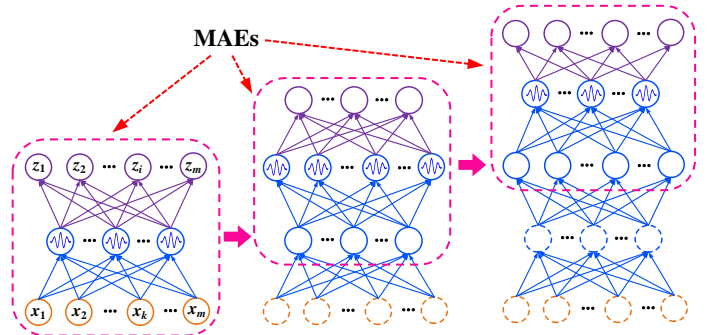


Fig. 3. Construction of an MMAE.

B. Adaptive Morlet wavelet design

According to Eq. (4), the performance of Morlet wavelet relies on the parameters f_b and f_c , as shown in Fig. 4. Different time-frequency resolutions can be acquired by adjusting these two parameters. Thus, it is important to design an adaptive Morlet wavelet through flexibly adjusting the two parameters to achieve the best match with the characteristics of the analyzed signals.

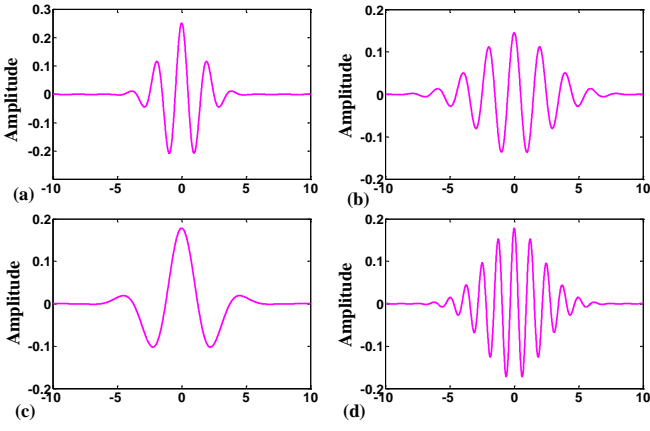


Fig. 4. Morlet wavelets with different parameters f_b and f_c : (a) $f_b = 5$ and $f_c = 0.5$; (b) $f_b = 15$ and $f_c = 0.5$; (c) $f_b = 10$ and $f_c = 0.2$; (d) $f_b = 10$ and $f_c = 0.8$.

FOA can effectively search global optimization with many advantages, and it is adopted for flexibly determining the adjustable parameters of Morlet wavelet in the present paper. More details of FOA can be found in [31]. The flowchart of the scheme is given in Fig. 5 and summarized as follows (The fitness is misclassification rate on validation samples).

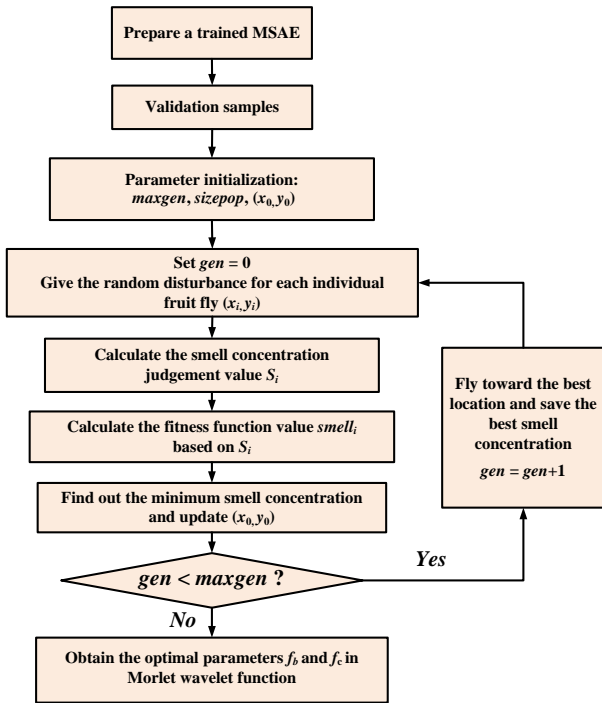


Fig. 5. Adaptive Morlet wavelet design of the MSAE using FOA.

- 1) Prepare an MSAE model with an initial Morlet wavelet that has already been trained by using the training samples.
- 2) Input the validation samples. Determine the maximum epoch number, population size, and initial location of the fruit fly swarm.
- 3) Give each fruit fly a random search direction and distance for foraging the food, based on the smell.
- 4) Calculate the smell concentration judgment value using the distance between each fruit fly and the origin.
- 5) Search the smell concentration of each location of the fruit fly through substituting the smell concentration judgment

- value into the fitness function, and then look for the minimum smell concentration among the fruit fly swarm.
- 6) The fruit fly swarm saves the best smell concentration value and will fly toward the best location using vision. Repeat step 3 to step 5, and continue the optimization until reaching the maximum epoch number.
- 7) The designed MSAE with adaptive Morlet wavelet is used to analyze the testing samples.

C. The overall framework of the proposed method

Fig. 6 gives the overall framework of the proposed method. The followings are its main steps:

Step 1: Collect the raw vibration data of the key parts of the rotating machine, which are divided into training, validation, and testing samples.

Step 2: Design MSAE with an adaptive Morlet wavelet.

2.1: Morlet wavelet is employed as the activation function to design MSAE based on Eqs. (4) - (6).

2.2: Nonnegative constraint is adopted to modify the cost function based on Eqs. (8) and (9).

2.3: Update the weights of the MSAE using Eqs. (10) - (12).

2.4: FOA is adopted for the adaptive Morlet wavelet design of the MSAE to minimize the misclassification rate of the validation samples.

Step 3: Verify the effectiveness of the developed fault diagnosis model by the testing samples.

IV. EXPERIMENTAL VALIDATION

CASE 1 Fault diagnosis of a sun gear unit

A. Fault data description of the sun gear

In Case 1, the effectiveness of the proposed approach is tested using the measured vibration data of sun gears from the Drivetrain Dynamics Simulator (DDS), University of Connecticut [36]. As shown in Fig. 7, the DDS mainly includes motor, motor controller, brake, parallel gearbox, and planetary gearbox. An accelerometer (model: PCB608A11) is installed to collect vibration signals at a sampling frequency of 20 kHz under the stable running of the monitored components. The specifications of the accelerometer including frequency range, measure range, and sensitivity are 0.5 Hz-10 kHz, ± 50 g, and 100 mV/g, respectively. A total of nine working conditions of sun gears in a planetary gearbox are employed as shown in Fig. 8. Specifically, the fault types include missing tooth, root crack, spalling, and chipping tip, which are common fault modes of gear. For chipping tip, there are five damage degrees.

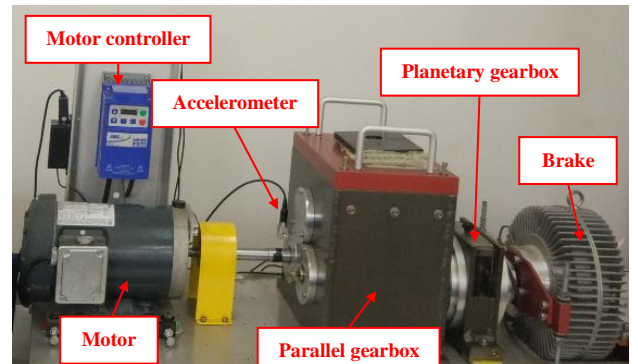


Fig. 7. The experimental setup of the planetary gearbox.

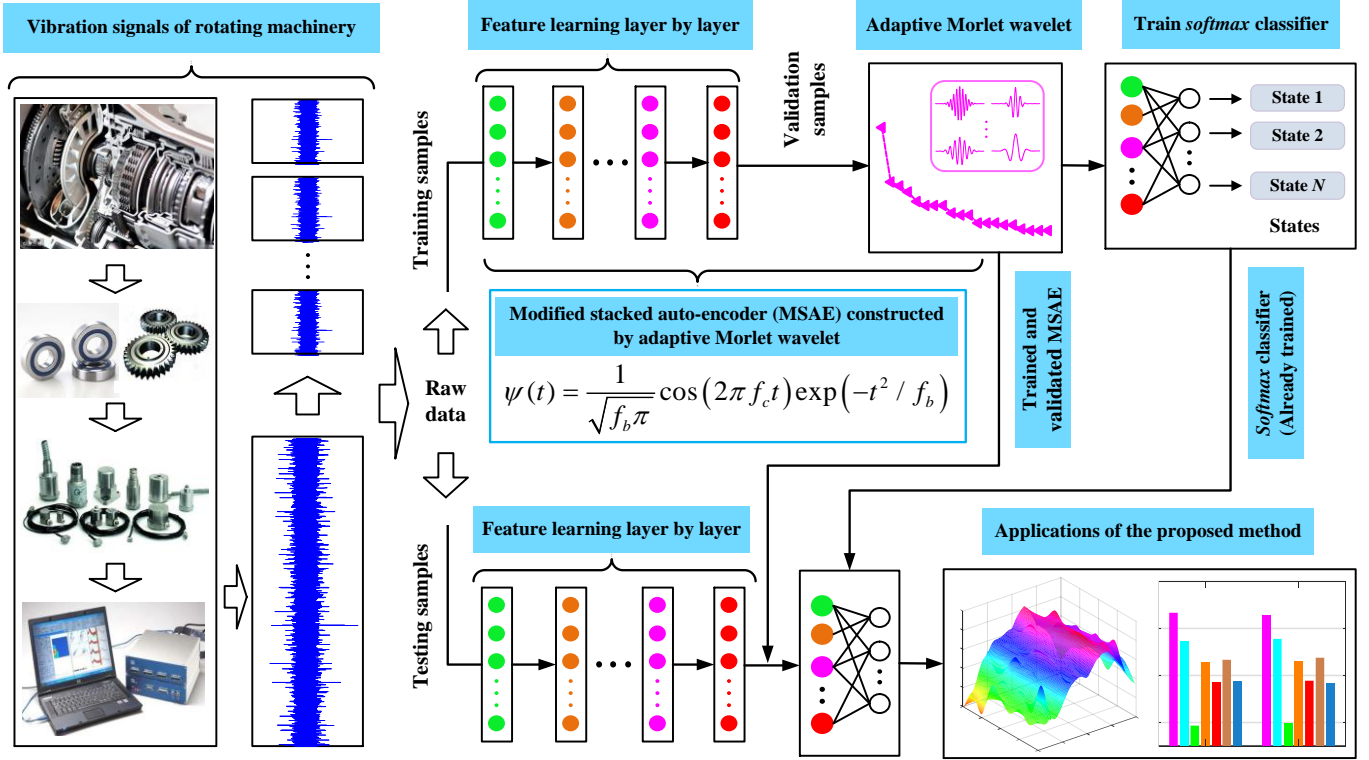


Fig. 6. The overall framework of the proposed method.



Fig. 8. Pictures of the nine types of sun gears.

Detailed sample distributions of the nine states of sun gears are listed in **Table I**. Every state has 360 data samples, and each sample is a vibration signal segment with 1024 data points. Randomly selected 200 out of 360 samples are employed for training, while the rest 60 and 100 samples are used for validation and testing, respectively. Time-domain waveforms of the nine working states of sun gears are plotted in **Fig. 9**.

TABLE I
THE NINE WORKING STATES OF SUN GEARS

Working states of sun gears	Number of the training/validation/testing samples	Labels of the states
Healthy	200 / 60 / 100	1
Missing tooth	200 / 60 / 100	2
Root crack	200 / 60 / 100	3
Spalling	200 / 60 / 100	4
Chipping tip (Severe 1)	200 / 60 / 100	5
Chipping tip (Severe 2)	200 / 60 / 100	6
Chipping tip (Severe 3)	200 / 60 / 100	7
Chipping tip (Severe 4)	200 / 60 / 100	8
Chipping tip (Severe 5)	200 / 60 / 100	9

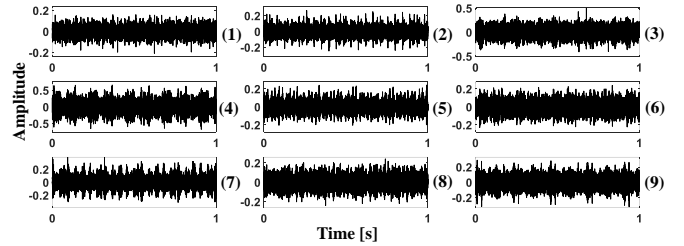


Fig. 9. Time-domain waveforms of the nine working states of sun gears

B. Comparison with state-of-the-art deep learning methods

The proposed method is compared to popular methods that use deep learning, including four kinds of SAEs, two kinds of DBNs, and a classical CNN model called LeNet-5. Their inputs are all selected as the 1024-dimensional raw vibration data.

To reduce the impact of contingency on the diagnosis results, ten trials for each method are run and the detailed results are shown in **Fig. 10**. For each run, the overall classification accuracy is the ratio of the total number of correctly classified samples to the total number of the testing samples whose true labels are known. As listed in **Table II**, the average value of the overall testing accuracy of the proposed method ($f_b = 0.621$ and $f_c = 3.114$) reaches 98.86% (8897/9000, 8897 is the total number of correctly classified samples during the ten runs and 9000=10*9*100 is the total number of testing samples during the ten runs), and it is higher than for the eight contrastive methods, which are 96.20%, 94.90%, 91.18%, 89.02%, 94.07%, 91.98%, 92.27%, and 90.22%. For the first run, the testing accuracy of the proposed method is 98.89% (890/900), and F -measure is used to evaluate its diagnosis performance for each working state, as shown in **Fig. 11**.

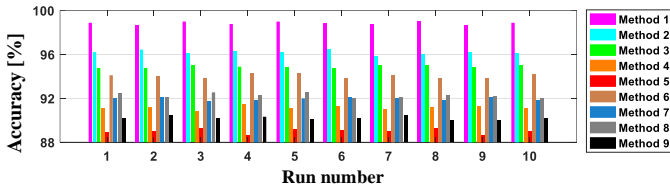


Fig. 10. The repeated diagnosis results based on the nine methods in Case 1. ($f_b = 0.621$ and $f_c = 3.114$)

TABLE II

STATISTICAL DIAGNOSIS RESULTS OF THE NINE METHODS IN CASE 1

Diagnosis methods	Average testing accuracies
Method 1 (Proposed method)	98.86% (8897/9000)
Method 2 (SAE: Morlet with C^T)	96.20% (8658/9000)
Method 3 (CNN: LeNet-5)	94.90% (8541/9000)
Method 4 (Gaussian DBN)	91.18% (8206/9000)
Method 5 (Basic DBN)	89.02% (8012/9000)
Method 6 (SAE: ReLU with C^E)	94.07% (8466/9000)
Method 7 (SAE: ReLU with C^T)	91.98% (8278/9000)
Method 8 (SAE: Sigm with C^E)	92.27% (8304/9000)
Method 9 (SAE: Sigm with C^T)	90.22% (8120/9000)

Notes: C^T means the traditional cost function in Eq. (7) and C^E means the enhanced cost function in Eq. (8).

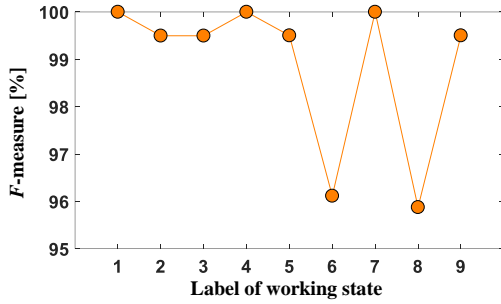


Fig. 11. F -measures of the proposed method for different gear sun states.

Based on the contrastive results, it can be concluded that the proposed method exhibits better diagnosis performance than other existing deep learning methods for fault diagnosis of a sun gear. Specifically, the superiority of the nonnegative constraint is demonstrated by the comparison results provided by Method 1 and Method 2. Also, the comparisons between the enhanced cost function and the traditional cost function of the first MAE are given in Fig. 12, which also shows the effectiveness of the former.

Many related studies have shown that SAEs with three hidden layers are often deep enough to perform high fault diagnosis accuracies. The model structure of the proposed MSAE is constructed as “1024-450-250-100-9” by setting the number of neurons in the hidden layers in descending order and about half of the neuron number of the previous layer. In this case study, the other hyper-parameters are given in Table III, and most of them are determined by experimental experience.

TABLE III

PARAMETERS OF THE PROPOSED METHOD IN CASE 1

Description	Value
Maximum epoch number of the MSAE	60
Sparsity constant / Sparse penalty coefficient	0.08 / 5
Weight decay factor / Learning rate	0.003 / 0.01
The optimal parameter f_b of Morlet wavelet	0.621 (Given by FOA)
The optimal parameter f_c of Morlet wavelet	3.114 (Given by FOA)
The maximum generation number of FOA	20
The population size of FOA	25

The importance of adaptive Morlet wavelet design is shown by the validation accuracies based on different combinations of parameters f_b and f_c , as shown in Fig. 13. From Fig. 13, it is seen that the validation accuracies are seriously affected by the parameters f_b and f_c . To show the superiority of FOA, two other algorithms are used for comparisons, which are genetic algorithm (GA) and particle swarm optimization (PSO). Fig. 14 shows the misclassification rates on the validation samples using different algorithms. The misclassification rates using FOA converge after about only 20 epoch times, while GA and PSO need more while starting from the same initial conditions. Besides, the minimum misclassification rate using FOA is smaller than the other three methods.

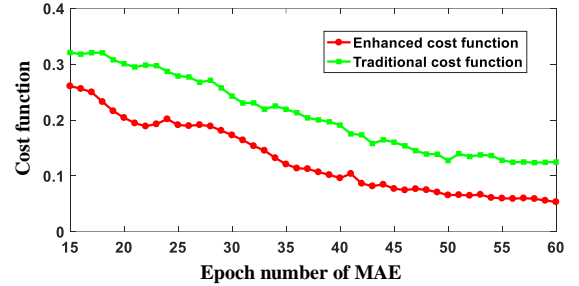


Fig. 12. Comparisons between the enhanced cost function and the traditional cost function in Case 1.

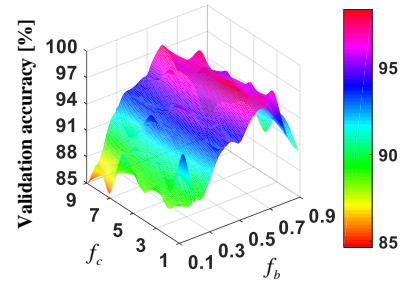


Fig. 13. The effect of parameters f_b and f_c on the validation accuracies in Case 1.

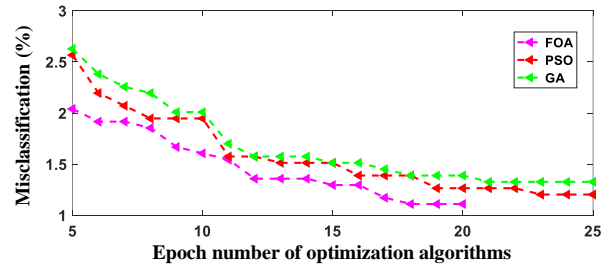


Fig. 14. Misclassification rate of the validation samples using three types of optimization algorithms in Case 1.

In addition to those methods listed in Table II, the proposed method is also compared with state-of-the-art deep learning methods published in recent years [37-41]. The influence of noise and comparisons between the other four types of popular wavelets [42] holding explicit expressions are discussed here at the same time. The average testing accuracies of the ten runs given under different white noise levels are shown in Table IV. Here, all the SAEs are designed with the enhanced cost functions, and in order to make fair comparisons, the input of each method is raw time-domain vibration data with no signal

pre-processing or feature extraction. Some conclusions can be drawn from **Table IV** as follows. (1) The proposed method based on Morlet wavelet shows higher diagnosis accuracies than Gaussian wavelet and Mexican Hat wavelet, while the results of Haar wavelet and Shannon wavelet are much worse. (2) The proposed method is more effective compared with the five state-of-the-art methods in analyzing the raw non-stationary vibration data under the influence of noise. (3) With the increase of noise, although the accuracies of all the methods decrease (from 96.30% to 88.16%), the proposed method shows the best anti-noise capability. This paper mainly focuses on diagnosing faults based on the raw vibration data, however, it is believed that using specialized signal processing techniques for de-noising can improve accuracy.

Diagnosis methods	Signal-Noise Ratio (SNR)			
	24 dB	20 dB	16 dB	12 dB
Method 1 (Proposed method)	96.30%	94.48%	92.01%	88.16%
Method 10 (SAE: Gaussian)	94.52%	93.64%	89.34%	85.86%
Method 11 (SAE: Haar)	81.88%	75.92%	71.50%	64.09%
Method 12 (SAE: Mexican Hat)	95.06%	93.97%	91.00%	86.42%
Method 13 (SAE: Shannon)	88.62%	85.07%	79.98%	72.06%
Method 14 proposed in [37]	88.29%	85.83%	81.38%	75.93%
Method 15 proposed in [38]	94.00%	91.94%	88.26%	84.03%
Method 16 proposed in [39]	93.38%	90.86%	87.90%	84.86%
Method 17 proposed in [40]	88.94%	86.50%	83.00%	78.34%
Method 18 proposed in [41]	87.14%	84.94%	81.06%	76.86%

CASE 2 Fault diagnosis of a roller bearing unit

C. Fault data description of roller bearing

In Case 2, the proposed method is used to analyze the vibration data collected from a bearing fault diagnosis test rig, Anhui University of Technology, China, as shown in **Fig. 15**. The rotating speed is set at 900 rpm and the load is 2kN.m. Vibration signals during the stable run are collected with a sampling frequency of 10 kHz using ICP INV9822 accelerometer. The specifications of the accelerometer including frequency range, measure range, and sensitivity are 0.5 Hz-8 kHz, ± 50 g, and 100 mV/g. The model number of the tested bearing is 6205-2RS.

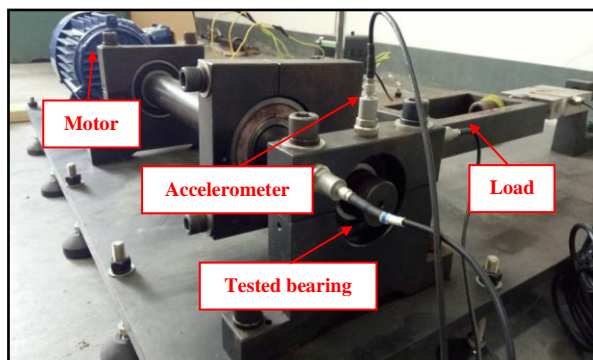


Fig. 15. Fault diagnosis test rig of roller bearings.

Nine working states of the roller bearing are collected and their samples are listed in **Table V**, including different fault types and different damage degrees. Each sample consists of

600 sampling data points, and the time-domain waveforms of the nine working states are plotted in **Fig. 16**.

TABLE V
THE NINE WORKING CONDITIONS OF ROLLER BEARINGS

Working states of roller bearings	Number of the training/validation/testing samples	Labels of the states
Normal	150 / 50 / 125	1
Ball (0.2 mm)	150 / 50 / 125	2
Ball (0.3 mm)	150 / 50 / 125	3
Inner race (0.2 mm)	150 / 50 / 125	4
Inner race (0.3 mm)	150 / 50 / 125	5
Inner race (0.4 mm)	150 / 50 / 125	6
Inner race (0.5 mm)	150 / 50 / 125	7
Outer race (0.2 mm)	150 / 50 / 125	8
Outer race (0.4 mm)	150 / 50 / 125	9

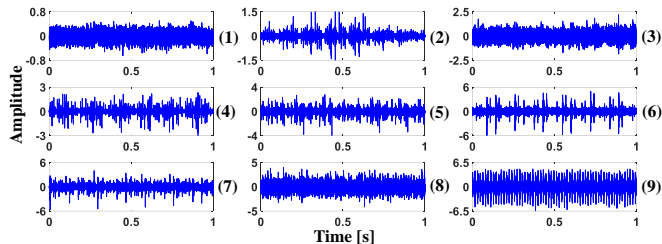


Fig. 16. Time-domain waveforms of the nine working states of roller bearings

D. Comparison with state-of-the-art deep learning methods

In this case study, LSTM, another two types of SAEs based on Leaky ReLU (LReLU) and exponential linear unit (ELU) [23], and the five state-of-the-art deep learning methods listed in **Table IV** are used for comparison.

As before, each method runs ten trials. The detailed testing diagnosis accuracies of the proposed method, LSTM, and two SAEs (LReLU, ELU) are given in **Fig. 17**, and their statistical results are listed in **Table VI**. The average testing accuracy provided by the proposed method ($f_b = 0.835$ and $f_c = 2.421$) is 97.16% (10930/11250, 11250=9*125*10), and it is higher than those of the six contrastive methods, which are 94.99%, 87.84%, 93.09%, 91.43%, 93.43%, and 91.54%, respectively. For the first run, the specific testing accuracy provided by the proposed method is 97.33% (1095/1125).

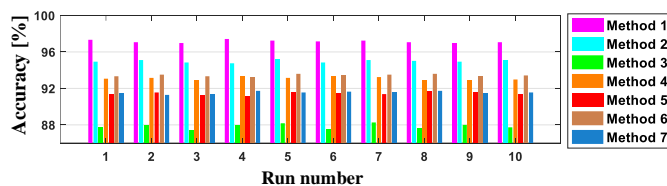


Fig. 17. The repeated diagnosis results based on the nine methods in Case 2. ($f_b = 0.835$ and $f_c = 2.421$)

TABLE VI
STATISTICAL DIAGNOSIS RESULTS OF THE SEVEN METHODS IN CASE 2

Diagnosis methods	Average testing accuracies
Method 1 (Proposed method)	97.16% (10930/11250)
Method 2 (SAE: Morlet with C^T)	94.99% (10686/11250)
Method 3 (Deep LSTM)	87.84% (9882/11250)
Method 4 (SAE: LReLU with C^E)	93.09% (10473/11250)
Method 5 (SAE: LReLU with C^T)	91.43% (10286/11250)
Method 6 (SAE: ELU with C^E)	93.43% (10511/11250)
Method 7 (SAE: ELU with C^T)	91.54% (10298/11250)

Notes: C^T means the traditional cost function in Eq. (7) and C^E means the enhanced cost function in Eq. (8).

The confusion matrix is shown in **Fig. 18** (two decimal places), in which the horizontal and vertical axes are the predicted and true state labels, respectively. These values located in the diagonal represent the correct classification rate of each state, and the others represent the misclassification rates. The diagnosis accuracies for most of the classes are higher than 0.97. The result on the fifth condition is slightly lower where some testing samples are misclassified to conditions 2, 4, and 6. **Table VII** shows the comparison results of the proposed method with the five state-of-the-art deep learning methods under different levels of white noise, which further confirms the superiority of the proposed method in the direct analysis of raw nonstationary vibration data under the influence of noise.

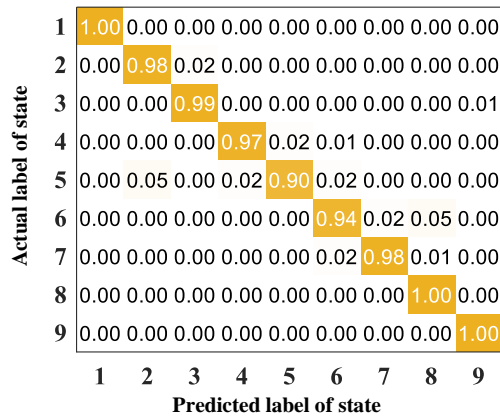


Fig. 18. Confusion matrix of the proposed method for the first run in Case 2.

TABLE VII
COMPARISONS BETWEEN OTHER WAVELETS AND STATE-OF-THE-ART METHODS UNDER DIFFERENT WHITE NOISE IN CASE 2

Diagnosis methods	Signal-Noise Ratio (SNR)			
	24 dB	20 dB	16 dB	12 dB
Method 1 (Proposed method)	95.04%	93.85%	90.37%	85.25%
Method 8 (SAE: Gaussian)	93.69%	91.28%	87.20%	82.08%
Method 9 (SAE: Haar)	80.29%	74.40%	68.67%	61.81%
Method 10 (SAE: Mexican Hat)	93.96%	91.52%	88.15%	83.25%
Method 11 (SAE: Shannon)	86.95%	84.75%	78.30%	70.38%
Method 12 proposed in [37]	88.01%	84.90%	80.08%	75.16%
Method 13 proposed in [38]	93.85%	91.00%	87.44%	82.68%
Method 14 proposed in [39]	92.79%	90.10%	86.95%	82.04%
Method 15 proposed in [40]	88.56%	85.48%	81.99%	76.95%
Method 16 proposed in [41]	86.48%	84.36%	80.20%	76.08%

TABLE VIII
PARAMETERS OF THE PROPOSED METHOD IN CASE 2

Description	Value
Maximum epoch number of the MSAE	70
Sparsity constant / Sparse penalty coefficient	0.10 / 5
Weight decay factor / Learning rate	0.003 / 0.01
The optimal parameter f_b of Morlet wavelet	0.835 (Given by FOA)
The optimal parameter f_c of Morlet wavelet	2.421 (Given by FOA)
The maximum generation number of FOA	20
The population size of FOA	25

The hyper-parameters of the MSAE are given in **Table VIII**. According to **Table III** in Case 1 and **Table VIII** in Case 2, the optimal parameters f_b and f_c have changed. Therefore, it is meaningful to adaptively determine these two adjustable parameters to achieve the best match with the analyzed data. The comparison results in Case 1 and Case 2 show that the proposed method is more effective than other state-of-the-art deep learning methods for fault diagnosis of rotating machinery

key parts. Adaptive Morlet wavelet activation function enables the MSAE to establish an accurate nonlinear mapping between various working states and the raw nonstationary vibration data. Besides, the nonnegative constraint of connection weights helps to achieve the high-quality reconstruction of the MSAE.

V. CONCLUSIONS

To boost the diagnosis performance of intelligent fault diagnosis methods for various fault categories and severities of rotating machinery, this paper presented a new method that used MSAE and incorporated adaptive Morlet wavelet. In this method, Morlet wavelet was used to construct the MSAE to establish an accurate mapping between the raw nonstationary vibration data and various fault states. The nonnegative constraint was used to enhance the cost function to achieve high-quality reconstruction. FOA was used to determine the adjustable parameters of Morlet wavelet to flexibly match the characteristics of the analyzed data.

The effectiveness of the proposed method was validated by the experimental raw vibration signals collected from a sun gear unit and a roller bearing unit. The analysis of the results demonstrated that the proposed method was superior to other state-of-the-art intelligent diagnosis methods. Future work includes how to systematically design the deep neural network; how to solve the issue of limited fault condition data; and how to transfer the learned knowledge in diagnosing machinery with different specifications and working environment.

VI. REFERENCES

- [1] C. Ding, M. Zhao, J. Lin, J. Jiao and K. Liang, "Sparsity based Algorithm for Condition Assessment of Rotating Machinery Using Internal Encoder Data", *IEEE Trans. Ind. Electron.*, vol. 67, no. 9, pp. 7982-7993, Sept. 2020.
- [2] S. Lu, R. Yan, Y. Liu and Q. Wang, "Tachless Speed Estimation in Order Tracking: A Review With Application to Rotating Machine Fault Diagnosis", *IEEE Trans. Instrum. Meas.*, vol. 68, no. 7, pp. 2315-2332, Jul. 2019.
- [3] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li and A. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap", *Mech. Syst. Signal Process.*, vol. 138, 106587, Apr. 2020.
- [4] R. Huang, J. Li, W. Lei and L. Cui, "Deep Ensemble Capsule Network for Intelligent Compound Fault Diagnosis Using Multisensory Data", *IEEE Trans. Instrum. Meas.*, vol. 69, no. 5, pp. 2304-2314, May 2020.
- [5] L. Sun, H. Wang and P. Chen, "Vibration-Based Intelligent Fault Diagnosis for Roller Bearings in Low-Speed Rotating Machinery", *IEEE Trans. Instrum. Meas.*, vol. 67, no. 8, pp. 1887-1899, Aug. 2018.
- [6] E. T. Esfahani, S. Wang, and V. Sundararajan, "Multisensor wireless system for eccentricity and bearing fault detection in induction motors", *IEEE-ASME Trans. Mech.*, vol. 19, no. 3, pp. 818-826, June 2014.
- [7] Y. Hu, W. Bao, X. Tu, F. Li and K. Li, "An Adaptive Spectral Kurtosis Method and its Application to Fault Detection of Rolling Element Bearings", *IEEE Trans. Instrum. Meas.*, vol. 69, no. 3, pp. 739-750, Mar. 2020.
- [8] Z. Huo, M. Martínez-García, Y. Zhang, R. Yan and L. Shu, "Entropy Measures in Machine Fault Diagnosis: Insights and Applications", *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 2607-2620, June 2020.
- [9] L. Wang, G. Cai, J. Wang, X. Jiang and Z. Zhu, "Dual-Enhanced Sparse Decomposition for Wind Turbine Gearbox Fault Diagnosis", *IEEE Trans. Instrum. Meas.*, vol. 68, no. 2, pp. 450-461, Feb. 2019.
- [10] Y. Hao, L. Song, B. Ren, H. Wang, and L. Cui, "Step-by-step compound faults diagnosis method for equipment based on majorization-minimization and constraint sca", *IEEE-ASME Trans. Mech.*, vol. 24, no. 6, pp. 2477-2487, Dec. 2019.

- [11] H. Shao, H. Jiang, H. Zhang and T. Liang, "Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network", *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2727-2736, Mar. 2018.
- [12] 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.*, vol. 72-73, pp. 303-315, May. 2016.
- [13] S. Shao, S. McAleer, R. Yan and P. Baldi, "Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning", *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446-2455, Apr. 2019.
- [14] M. Xia, T. Li, T. Shu, J. Wan, C. W. de Silva and Z. Wang, "A Two-Stage Approach for the Remaining Useful Life Prediction of Bearings Using Deep Neural Networks," *IEEE Trans. Ind. Informat.*, vol. 15, no. 6, pp. 3703-3711, June 2019.
- [15] M. Xia, T. Li, L.Xu, L.Liu, and C. W. de Silva, "Fault Diagnosis for Rotating Machinery Using Multiple Sensors and Convolutional Neural Networks", *IEEE-ASME Trans. Mech.*, vol. 23, no. 1, pp. 101-110, Feb. 2018.
- [16] S. Chen, Y. Meng, H. Tang, Y. Tian, N. He and C. Shao, "Robust Deep Learning-Based Diagnosis of Mixed Faults in Rotating Machinery" , *IEEE-ASME Trans. Mech.*, vol. 25, no. 5, pp. 2167-2176, Oct. 2020.
- [17] G. Jiang, P. Xie, H. He, and J. Yan, "Wind turbine fault detection using a denoising autoencoder with temporal information", *IEEE-ASME Trans. Mech.*, vol. 23, no. 1, pp. 89-100, Feb., 2018.
- [18] J. Wang, P. Fu, L. Zhang, R. Gao and R. Zhao, "Multilevel Information Fusion for Induction Motor Fault Diagnosis", *IEEE-ASME Trans. Mech.*, vol. 24, no. 5, pp. 2139-2150, Oct. 2019.
- [19] X. Pang, X. Xue, W. Jiang and K. Lu, "An Investigation into Fault Diagnosis of Planetary Gearboxes using a Bispectrum Convolutional Neural Network", *IEEE-ASME Trans. Mech.*, 2020, DOI 10.1109/TMECH.2020.3029058.
- [20] X. Zhao, M. Jia, P. Ding, C. Yang, D. She and Z. Liu, "Intelligent Fault Diagnosis of Multi-Channel Motor-Rotor System based on Multi-manifold Deep Extreme Learning Machine", *IEEE-ASME Trans. Mech.*, 2020, DOI 10.1109/TMECH.2020.3004589.
- [21] M. Xia, T. Li, L.Xu, L.Liu, and C. W. de Silva, "An Intelligent Fault Diagnosis Approach with Unsupervised Feature Learning by Stacked Denoising Autoencoder", *IET Science, Measurement & Technology*, vol. 11, no. 6, pp. 687-695, Sep. 2017.
- [22] 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.
- [23] H. Shao, H. Jiang, L. Ying and X. Li, "A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders", *Mech. Syst. Signal Process.*, vol. 102, pp. 278-297, Mar. 2018.
- [24] Y. Li, K. Feng, X. Liang and Ming J. Zuo, "A fault diagnosis method for planetary gearboxes under non-stationary working conditions using improved Vold-Kalman filter and multi-scale sample entropy", *J. Sound Vib.*, vol. 439, pp. 271-286, Jan. 2019.
- [25] H. Shao, H. Jiang, X. Li and S. Wu, "Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine", *Knowl-Based Syst.*, vol. 140, pp. 1-14, Jan. 2018.
- [26] X. Wen, Q. Miao, J. Wang and Z. Ju, "A multi-resolution wavelet neural network approach for fouling resistance forecasting of a plate heat exchanger", *Appl. Soft. Comput.*, vol. 57, pp. 177-196, Aug. 2017.
- [27] Y. Lei, Z. He and Y. Zi, "EEMD method and WNN for fault diagnosis of locomotive roller bearings", *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7334-7341, Jun. 2011.
- [28] B. Tang, W. Liu and T. Song, "Wind turbine fault diagnosis based on Morlet wavelet transformation and Wigner-Ville distribution", *Renew. Energ.*, vol. 35, no. 12, pp. 2862-2866, Dec. 2010.
- [29] Y. Jiang, B. Tang, Y. Qin and W. Liu, "Feature extraction method of wind turbine based on adaptive Morlet wavelet and SVD", *Renew. Energ.*, vol. 36, no. 8, pp. 2146-2153, Aug. 2011.
- [30] S. Hassairi, R. Ejbali and M. Zaied, "A deep stacked wavelet auto-encoders to supervised feature extraction to pattern classification", *Multimed. Tools Appl.*, vol. 4, pp. 1-17, Mar. 2017.
- [31] W.T. Pan, "A new Fruit Fly Optimization Algorithm: Taking the financial distress model as an example", *Knowl-Based Syst.*, vol. 26, pp. 69-74, Feb. 2012.
- [32] Y. Pan, R. Hong, J. Chen and W. Wu, "A hybrid DBN-SOM-PF-based prognostic approach of remaining useful life for wind turbine gearbox", *Renew. Energ.*, vol. 152, pp. 138-154, Jun. 2020.
- [33] E. Hosseini-Asl, J.M. Zurada and O. Nasraoui, "Deep Learning of Part-Based Representation of Data Using Sparse Autoencoders with Nonnegativity Constraints", *IEEE Trans. Neur. Net. Lear.*, vol. 27, no. 12, pp. 2486-2498, Dec. 2016.
- [34] A. Afan and Y. Fan, "Automatic Modulation Classification Using Deep Learning Based on Sparse Autoencoders With Nonnegativity Constraints", *IEEE Signal Proc. Let.*, vol. 24, no. 11, pp. 1626-1630, Nov. 2017.
- [35] X. Li, H. Jiang, K. Zhao and R. Wang, "A Deep Transfer Nonnegativity-Constraint Sparse Autoencoder for Rolling Bearing Fault Diagnosis With Few Labeled Data", *IEEE Access*, vol. 7, pp. 91216-91224, Jul. 2019.
- [36] P. Cao, S. Zhang and J. Tang, "Pre-Processing-Free Gear Fault Diagnosis Using Small Datasets with Deep Convolutional Neural Network-Based Transfer Learning", *IEEE Access*, vol. 6, pp. 26241-26253, May. 2018.
- [37] W. Deng, H. Liu, J. Xu, H. Zhao and Y. Song, "An Improved Quantum-Inspired Differential Evolution Algorithm for Deep Belief Network", *IEEE Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 7319-7327, Oct. 2020.
- [38] M. Zhao, S. Zhong, X. Fu, B. Tang, S. Dong and M. Pecht, "Deep Residual Networks with Adaptively Parametric Rectifier Linear Units for Fault Diagnosis", *IEEE Trans. Ind. Electron.*, 2020, DOI 10.1109/TIE.2020.2972458.
- [39] F. Jia, Y. Lei, J. Lin, N. Lu and S. Xing, "Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization", *Mech. Syst. Signal Process.*, vol. 15, pp. 349-367, Sep. 2018.
- [40] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen and J. Wang, "Machine Health Monitoring Using Local Feature-Based Gated Recurrent Unit Networks", *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539-1548, Feb. 2018.
- [41] Z. Zhao, T. Li, J. Wu, C. Sun, S. Wang, R. Yan and X. Chen, "Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study", *ISA T.*, vol. 107, pp. 224-255, Dec. 2020.
- [42] G. Wang, L. Guo and H. Duan, "Wavelet neural network using multiple wavelet functions in target threat assessment", *Sci. World J.*, vol. 1, 632437, Jan. 2013.



Haidong Shao received the B.S. degree in Electrical Engineering and Automation and the Ph.D. degree in Vehicle Operation Engineering from Northwestern Polytechnical University, Xi'an, China, in 2013 and 2018, respectively. He is currently an Assistant Professor in the College of Mechanical and Vehicle Engineering at Hunan University, Changsha, China. From 2019 to 2021, he was a Postdoctoral Fellow with the Division of Operation and Maintenance Engineering, Luleå University of Technology, Luleå, Sweden. His current research interests include fault diagnosis, and prognostics and health management.



Min Xia (M'11) is currently a lecturer in the Department of Engineering at the Lancaster University, UK. He received B.S. degree from Southeast University, China (2009); M.S. degree from the University of Science and Technology of China, China (2012); and Ph.D. degree from the University of British Columbia, Canada (2017). His research interests include smart manufacturing, machine diagnostics and prognostics, deep neural networks, wireless sensor network and sensor fusion.



Jiafu Wan has been a Professor in School of Mechanical & Automotive Engineering at South China University of Technology (SCUT) since Sep 2015. He joined in SCUT in May 2014. He received the Ph.D. degree in Mechatronic Engineering from SCUT in Jun 2008. His research interests include Cyber Physical Systems, Intelligent Manufacturing, Big Data Analytics, Industry 4.0, Smart Factory and Cloud Robotics.



Clarence W. de Silva (F'98) is a Fellow of: IEEE, ASME, Canadian Academy of Engineering, and Royal Society of Canada. He received Ph.D. degrees from Massachusetts Institute of Technology (1978); and University of Cambridge, U.K. (1998); and honorary D.Eng. degree from University of Waterloo, Canada (2008), and an earned higher doctorate (ScD) from University of Cambridge (2020). He has been a Professor of Mechanical Engineering and Senior Canada Research Chair and NSERC-BC Packers Chair in Industrial Automation, at the University of British Columbia, Vancouver, Canada since 1988. He has authored 25 books and over 550 papers, approximately half of which are in journals. His recent books published by Taylor & Francis/CRC are: *Modeling of Dynamic Systems—with Engineering Applications* (2018); *Sensor Systems* (2017); *Sensors and Actuators—Engineering System Instrumentation, 2nd edition* (2016); *Mechanics of Materials* (2014); *Mechatronics—A Foundation Course* (2010); *Modeling and Control of Engineering Systems* (2009); *VIBRATION—Fundamentals and Practice, 2nd Ed.* (2007); and by Addison Wesley: *Soft Computing and Intelligent Systems Design—Theory, Tools, and Applications* (with F. Karray, 2004).