IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 10 January 2023, accepted 8 February 2023, date of publication 14 February 2023, date of current version 17 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3244795

RESEARCH ARTICLE

Fault Diagnosis of Rotating Machinery Using Denoising-Integrated Sparse Autoencoder Based Health State Classification

JING YANG[1], GUO XIE[2], (Member, IEEE), AND YANXI YANG[2]
[1]School of Mechatronics and Automotive Engineering, Tianshui Normal University, Tianshui 741000, China
[2]School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China

Corresponding author: Jing Yang (JingYangTS@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 62063032, and in part by the Natural Science Foundation of Gansu Province under Grant 21JR1RE296.

ABSTRACT The diagnostic study on single-fault with distinguishing features based on monitoring data analysis is mature and fruitful in recent years. However, the early fault signals collected by practical monitoring systems often possess the following characteristics: 1) Fairly weak signal strength; 2) Submerged in powerful background noise; 3) Coupling of different fault data. These features not only increase the diagnostic difficulty, but also make the existing methods hardly to get the desired results. Consequently, the early compound faults diagnosis commonly in industrial systems is still a thorny and urgent problem. Therefore, in order to solve this problem and provide technical support for the practical industrial machinery fault diagnosis, a denoising-integrated sparse autoencoder (DISAE) model for early compound faults diagnosis is proposed in this paper. The innovation points of this study mainly include: 1) A feature-enhanced and denoising solution based on fault sensitivity degree (FSD) is designed, and the reconstructed diagnostic signals are acquired. 2) A disassociation framework is formulated, and the data coupling is solved. 3) A weight constraint term of SAE is constructed to improve the effectiveness and diversity of feature learning. 4) An adaptive loss function and a DISAE model is formed, and the early compound faults diagnosis is achieved. Finally, different trials and comparison results display the effectiveness and superiority of the designed DISAE based scheme.

INDEX TERMS Early compound faults, EEMD (Ensemble Empirical Mode Decomposition), fault diagnosis, fault sensitivity degree, signal denoising, sparse autoencoder.

I. INTRODUCTION

With the rise of various analysis tools, the rapid development of data-based mechanical fault diagnosis technology has been promoted. Currently, with benefit from data mining technology to achieve equipment health condition detection has played a decisive role in many key tasks [1], [2], [3], and has long been valued and studied [4], [5], [6], [7]. However, since large scale and complexity have become the characteristics of current applied machines, the composition and structure of the device are related and affected, and the monitoring signals characterizing the running state exhibit high dimension, sparse, hard-to-quantify and indistinguishable properties. Even more troublesome, these machines frequently run in harsh environments with intensive noise. The existence of early compound faults greatly increases the difficulty of condition monitoring. For the current situation, improper evaluation of early compound faults' occurrence has resulted in serious or even irreparable damages. So, on the basis of denoising and decoupling of monitoring data, it is of great significance to fully mine the feature information hidden in them, so as to achieve accurate early compound faults diagnosis.

Actually, most industrial machinery would produce vibration signal with different frequencies during working, one is the inevitable normal vibration inherent in the machine,

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

while the other is the undesired abnormal vibration caused by equipment failure. Therefore, using vibration signal analysis can effectively achieve reliable and accurate mechanical health condition recognition. Generally, feature extration and status recognition [8] are essential part of traditional fault diagnosis based on vibration signal analysis. Su et al. [9] adopted OSLLTSA (Orthogonal Supervised Linear Local Tangent Space Alignment) and LS-SVM (Least Square Support Vector Machine) to design a multi-fault diagnosis method for rotating machinery. The effectiveness of the method was verified by classifying the health condition of rolling bearing. Based on empirical wavelet transforms and EMD (Empirical Model Decomposition), Kedadouche et al. [10] achieved single-point fault diagnosis of bearings. Using maximum correlated kurtosis deconvolution, an improved spectral kurtosis scheme was formed in [11], which used feature parameter selection to achieve fault diagnosis under the condition of known fault feature frequency. A method based on complete EEMD with adaptive noise and improved multivariate MSE (Multiscale Sample Entropy) is proposed by Lv et al. [12], which extracted the characteristic frequency of fault signal to realize mechanical early fault diagnosis. Although fine fruits has been well applied, above measures were usually inseparable from artificial algorithms to extract the representative features hidden in the monitoring data. Therefore, not only does it requires high operational requirements and is difficult to implement, but also the diagnosis accuracy is restricted by non-objective matters, which seriously hinders the superiority of these designed measures. Noteworthily, the well studied measures suitable for single-structure devices cannot be applied to common early compound faults caused by the complexity and correlation of the structure [13].

Currently, deep learning [14] has been broadly used in different fields [15], [16], [17], [18]. Specially, as one of classical deep learning technology, autoencoder (AE) has been prevailing applied in a wide area such as image processing, speech recognition and medicine [19], [20]. Actually, the application of AE technology to the machinery fault diagnosis has also received desired achievements [21], [22], [23], [24], [25]. In [26] a stacked denoising AE based treatment measure was built, unfortunately, it was incapable when there were multiple mixed fault modes. A method based on stacked denoising AE to discriminate individual faults was given in [27], which ignored the fault level determination as well as the early compound faults identification.

Early compound faults signal itself is weak, interfered by strong background noise, and different fault data are coupled with each other, which makes the existing fault diagnosis methods hardly to apply. Therefore, a DISAE model for early compound faults diagnosis is proposed in this paper. The main innovation points of this work are as follows. 1) FSD-based feature-enhanced and denoising strategy is developed, and then the reconstructed high-quality diagnostic signals are produced. 2) Undesired data correlation in diagnosis are mitigated by the designed disassociation framework. 3) The constraint term for model hyper-parameters is proposed, and the effectiveness and diversity of feature learning are improved. 4) Adopting the designed adaptive cost function, early compound faults identification is achieved by the proposed DISAE.

Finally, this study is organized as follows: early compound faults as well as its diagnosis research status are elaborated in Section II. The constructed DISAE model for early compound faults diagnosis is demonstrated completely in Section III. The artificial and practical early compound faults diagnosis of gearbox and rolling bearings are conducted respectively to exhibit the effectiveness and superiority of the designed DISAE based scheme in Section IV. Conclusions are drawn in Section V.

II. EARLY COMPOUND FAULTS CHARACTERISTICS AND DIAGNOSIS RESEARCH STATUS

From the structure and working mechanism, running equipment would generally produce vibration. To be specific, in the normal operating condition, the behavior caused by the device itself is weak and stable, but such phenomenon would be destroyed in an abnormal condition. Actually, faults are mainly caused by internal defects of device (such as bearing pitting and gear wear). For a machine with failure symptoms, it would output error results during the active period of failure, while the system outputs the correct result again during the inactive period. In other words, the acquired monitoring signal from faulty part is deformed relative to the signal from normal location. Then, take the condition monitoring data of gearbox (two-stage cascade structure) which is widely used in the practical industrial system as an example to analyze, as shown in FIGURE 1 and FIGURE 2. FIGURE 1 is the vibration curve of the unloaded gearbox in normal. It can be seen that now this behavior conforms to the vibration under normal operating condition.

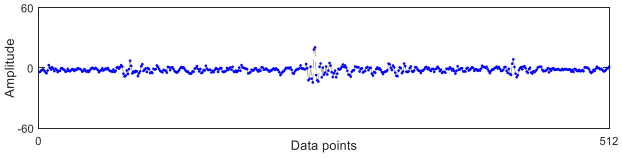

FIGURE 1. Vibration track of gearbox in normal condition.

FIGURE 2 is the vibration curve of the loaded gearbox in early fault condition. Specifically, FIGURE 2 (a) is the curve when only the secondary gear is in the early wear fault condition, FIGURE 2 (b) is the one when the primary gear is in the early pitting fault and the secondary gear is in the early wear fault (i.e. early compound fault) condition, and FIGURE 2 (c) is the one when the primary gear is in the early broken tooth fault and the secondary gear is in the early wear fault (i.e. early compound fault) condition. It shows small amplitude at the non-faulty part, and large and non-stationary amplitude at the faulty unit. In addition, although early compound faults in FIGURE 2 (b) and FIGURE 2 (c)

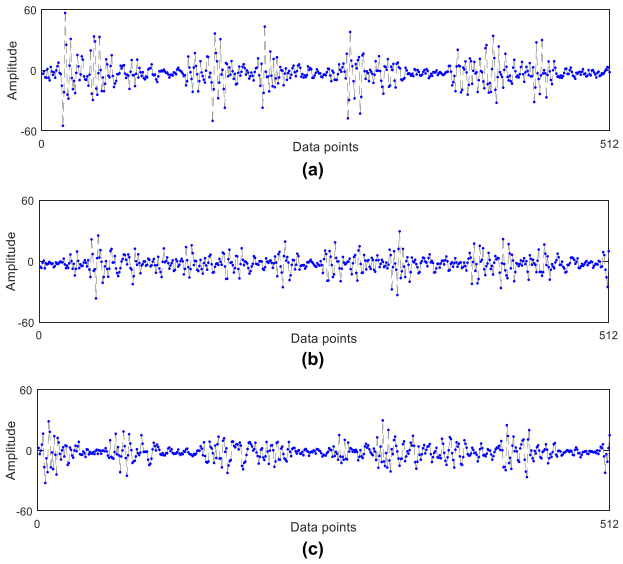

FIGURE 2. Vibration curves of gearbox in early fault conditions. (a) Gearbox in single point wear fault condition; (b) Gearbox in early compound faults condition (pitting fault & wear fault); (c) Gearbox in early compound faults condition (broken teeth fault & wear fault).

are dissimilar, the curves of them are hardly to discriminate. Therefore, compared with early single-point faults, the vibration of early compound faults is more complicated and the fault characteristics are not obvious. In fact, for the modern mechanical system, the complex operating conditions and harsh working environment determine the non unicity of fault characteristics, which means that early compound faults inevitably exist. However, the trouble of early compound faults diagnosis lies in non superposition of fault characteristics, submergence of background noise, and relevance of monitoring data.

Since early compound faults in equipment seriously block the practical industrial production process, researchers have designed multiple fault diagnosis methods about working reliability of current industry machinery. By means of deep AE, a fault identification model was formed in [27], which adaptively realized the feature mining and then the single-point fault recognition of rolling bearing based on the preprocessed vibration signals. However, the designed diagnosis network is multi hidden layer structure, and the traditional activation function and optimization algorithm make the training time longer. In [28] the classic sparse stacked autoencoders were used to construct hybrid feature pool, and achieved health status recognition of bearing. Unfortunately, the manual selection of the desired feature parameters for status recognition was indispensable. In addition, the structure of the two-stage model was complex and created inefficiencies. To make matters worse, the ability to identify rolling parts faults deteriorated significantly. Based on stacked SAE (Sparse Autoencoder) and signals preprocessed by EEMD, Qi et al. [29] designed a method for mechanical health status identification, and based on a small fraction of data to carry out verification tests. The design of these tests lacked practical considerations and ignored the differences in the sensitivity of IMF (Intrinsic Mode Function). Certainly, these findings are indeed applicable to non early single point of failure, but the unique characteristics of early compound faults have not been taken into account.

In summary, for finding a solution to early compound faults identification commonly in industrial machinery, a DISAE model for early compound faults diagnosis is proposed in this paper. The method is mainly composed of several critical steps: 1) original signal feature-enhanced and denoising; 2) decoupling constraint term and weight constraint term design; 3) adaptive loss function and diagnosis model construction; 4) model training and early compound faults identification. Specifically, the original signal feature-enhanced and denoising mainly use the FSD of the IMFs generated by EEMD to achieve the diagnosis signal preprocessing; Decoupling constraint and weight constraint are designed to alleviate the impact of data correlation on the diagnosis accuracy and realize the effectiveness and diversity of DISAE feature learning; The construction of adaptive loss function and diagnosis model is mainly to build DISAE diagnosis network suitable for early compound faults, and finally achieve early compound faults diagnosis accurately through effective training.

III. DISAE MODEL BASED ADAPTIVE DIAGNOSIS SCHEME

DISAE scheme for early compound faults diagnosis is comprehensively described here, and the implementation of the proposed method is shown in FIGURE 3.

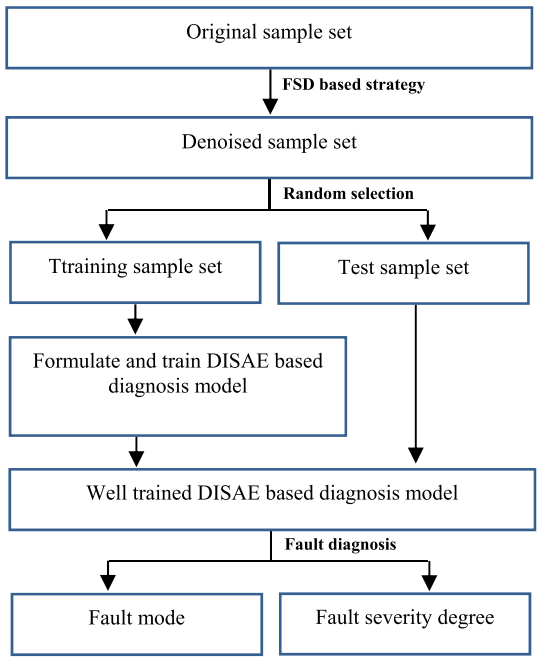

FIGURE 3. Block diagram of the proposed DISAE based early compound faults diagnosis method.

A. FSD BASED FEATURE-ENHANCED AND DENOISING

As is known to all, in order to suppress the mode mixing of empirical mode decomposition (EMD), Flandrin et al. [30] proposed an improved EMD method based on noise-assisted analysis, namely ensemble empirical mode decomposition (EEMD). Specifically, EEMD is a kind of multiple EMD superimposed with Gaussian white noise, which makes use of the statistical characteristics of uniform frequency distribution of Gaussian white noise. By adding different white noise of the same amplitude each time to change the extreme point characteristics of the signal, and then averaging the corresponding IMF obtained from multiple EMDs to offset the added white noise, so as to effectively suppress the generation of mode mixing. Actually, for the analysis and disposal of on-stationary and non-linear signals, EEMD is superior to other technologies, moreover, the signal-to-noise ratio (SNR) of the processed signals is relatively high.

For the monitoring data of fault signals, some of the IMF units obtained by EEMD technology especially exhibit a unique sensitivity to mechanical failures and contain abundant health condition information, while others only have invalid information for fault diagnosis [31], [32]. So, for noise elimination of monitoring signal and enhancement of hidden feature information, thereby effectively improving the accuracy and efficiency of early compound faults diagnosis, it is necessary to design an appropriate feature-enhanced and denoising measure, for selecting the IMF units with high FSD and eliminating ones with low FSD.

Therefore, an effective measure is designed to screen the IMF units with high FSD. The measure is to determine the IMF units with high FSD by surveying the correlation between the original vibration signal and its IMF units, also the relevance between the monitoring data in normal and those IMF units. Importantly, it is necessary to excavate the units with high FSD for the reson of highlighting the potential failure feature. Supposing that the monitored vibration signal in normal denoted by $s_{Normal}(t)$ and the original vibration signal to be analyzed recorded as $\mathrm{s}_{diagnosis}(t)$, then the operation process of the proposed FSD based feature-enhanced and denoising measure is as follows.

Step 1: Calculate the correlation coefficient cc_{i}^1 between $\mathrm{s}_{diagnosis}(t)$ and the i^{th} IMF component $c_i (i = 1, 2, \ldots, N)$ of $\mathrm{s}_{diagnosis}(t)$.

Step 2: Calculate the correlation coefficient cc_{i}^2 between $s_{Normal}(t)$ and the i^{th} IMF component c_i $(i = 1, 2, \ldots, N)$ of $\mathrm{s}_{diagnosis}(t)$.

Step 3: Define and calculate the early compound faults-related coefficient cc_{i}^3 according to cc_{i}^1 and cc_{i}^2 calculated by Step 1 and Step 2, as shown in formula (1),

$$cc_{\mathrm{i}}^3 = cc_{\mathrm{i}}^1 - cc_{\mathrm{i}}^2 \tag{1}$$

Step 4: Define and calculate the FSDs R_i of IMF units of the original signal according to cc_{i}^3 calculated by Step 1-Step 3, as shown in formula (2),

$$\begin{cases} R_i = \dfrac{cc_{\mathrm{i}}^3 - \min(cc^3)}{\max(cc^3) - \min(cc^3)} \\ cc^3 = \sum\limits_{i=1}^{N} cc_{\mathrm{i}}^3 \end{cases} \tag{2}$$

Step 5: Determine the IMF units component with high FSD according to the FSDs, and the detailed operation steps are described below.

1) IMF units of $\mathrm{s}_{diagnosis}(t)$ are ranked based on their possessed FSD descending order to obtain the IMF units sequence as given in (3),

$$\begin{cases} \{c_i'\}, & i = 1, 2, \ldots, N \\ c_1' \geq c_2', & \ldots, \geq c_i', \ldots, \geq c_N' \end{cases} \tag{3}$$

2) Differentiation factor df_i between those FSDs of two adjacency IMF units c_i' after sorting is calculated, as shown in formula (4),

$$\mathrm{df}_i = c_i' - c_{i+1}' \tag{4}$$

3) Seek out the index i belonging to the maximum differentiation factor df_i, subsequently, the first i IMF units $\{c_1', \ldots, c_i'\}$ in the ranked series $\{c_i'\}$ are the IMF units with high FSD.

4) The IMF units with high FSD is determined to generate the denoised signal $\tilde{s}_{diagnosis}(t)$ in formula (5) for early compound faults recognition,

$$\tilde{s}_{diagnosis}(t) = \sum_{\mathrm{j}=1}^{i} c_j' \tag{5}$$

According to the Step 1-Step 5 above, the proposed FSD based feature-enhanced and denosing measure essentially classifies all IMF units into two groups based on the IMF's FSD of fault, one group contains the IMF units with high FSD, while another group contains the ones with low FSD. The definition of the IMF's FSD of fault in this paper is summarized as follows.

1) The IMF units revealing the original monitoring signal feature strongly possess correlations to the original monitoring signal itself. Instead, IMF units with low FSD own weakly relevance to the original monitoring signal. So, cc_i^1 between $\mathrm{s}_{diagnosis}(t)$ and the i^{th} IMF component $c_i (i = 1, 2, \ldots, N)$ of $\mathrm{s}_{diagnosis}(t)$ can be adopted as a criterion to reflect the FSD of IMF units.
2) Taking $s_{Normal}(t)$ as a reference to calculate the correlation coefficient cc_i^2 between it and the i^{th} IMF component $c_i (i = 1, 2, \ldots, N)$ of $\mathrm{s}_{diagnosis}(t)$ can really clear the public information hidden in $\mathrm{s}_{diagnosis}(t)$ and $s_{Normal}(t)$.
3) Determination of the IMF's FSD of fault both regards to the relevance between the original signal and its IMF, as well as the relevance between the normal signal and those IMF.

Therefore, the proposed FSD based feature-enhanced and denoising strategy not only enhances the fault features hiding

in the original monitoring signal, but also alleviates the influence of background noise interference, thus, it is more instrumental in early compound faults feature mining and equipment health status identification.

B. DISAE ADAPTIVE DIAGNOSIS MODEL

The differences between the DISAE-based intelligent diagnostic model designed in this study and traditional-SAE network are described as follows.

Difference 1: Sparsity constraint term. Traditional SAE generally uses the KL (Kullback-Leibler) divergence function as the sparsity constraint. However, there are two parameters of the function that need to be optimized during the model training process. Differently, the L1 norm implements weight sparsity, while it contains only one optimization parameter. Therefore, L1 norm is employed as the sparsity constraint term of DISAE in this paper. Specifically, the sparsity constraint on the feature vector $\boldsymbol{h}_{\text{pre}}^{1}$ of the primary hidden layer is given in (6),

$$\left\|\boldsymbol{h}_{\text{pre}}^{1}\right\|_{1}=\sum_{\text{pre}=1}^{\text{TN}}\left|h_{\text{pre}}^{1}\right| \tag{6}$$

Difference 2: Decoupling constraint term. The monitoring signals characterizing the running state of equipment show the correlation and interaction, which inevitably blocks the diagnostic effectiveness and credibility for early compound faults. Thus, in order to alleviate the adverse effects caused by data coupling, a decoupling constraint measure is designed in (7) below,

$$\underset{\boldsymbol{W},\boldsymbol{b}}{min}\frac{1}{2\text{TN}}\sum_{\text{pre}=1}^{\text{TN}}\left\|\boldsymbol{s}_{\text{pre}}^{\text{g}}.(\boldsymbol{s}_{\text{pre}}^{\text{g}})'-\bar{\boldsymbol{s}}_{\text{pre}}^{\text{g}}.(\bar{\boldsymbol{s}}_{\text{pre}}^{\text{g}})'\right\|_{2}^{2} \tag{7}$$

where $\boldsymbol{W}=\left\{\boldsymbol{W}_{1}^{\text{g}},\boldsymbol{W}_{2}^{\text{g}}\right\},\boldsymbol{b}=\left\{\boldsymbol{b}_{1}^{\text{g}},\boldsymbol{b}_{2}^{\text{g}}\right\}$, $\boldsymbol{W}_{1}^{\text{g}}$ and $\boldsymbol{W}_{2}^{\text{g}},\boldsymbol{b}_{1}^{\text{g}}$ and $\boldsymbol{b}_{2}^{\text{g}}$ are the weight matrices and bias vectors of the primary and secondary hidden layer respectively. $\left\{\boldsymbol{s}_{\text{pre}}^{\text{g}}\right\}_{\text{pre}=1}^{\text{TN}}$ is the sample set preprocessed by the proposed FSD based feature-enhanced and denoising strategy, where $\boldsymbol{s}_{\text{pre}}^{\text{g}}\in\mathbb{R}^{\text{N}\times 1}$, $\text{g}=1,2\ldots,\text{g}_{\text{max}}$, g_{max} is sample group size, TN is the number of samples, i.e. $\boldsymbol{s}_{\text{pre}}^{\text{g}}$ represents the pre^{th} sample in the g^{th} sample set. $\bar{\boldsymbol{s}}_{\text{pre}}^{\text{g}}$ is the reconstructed signal of $\boldsymbol{s}_{\text{pre}}^{\text{g}}$.

Noteworthily, when the formula (7) is optimized, in order to reduce the computational burden and improve the efficiency of early compound faults diagnosis, the secondary data relationship need to be removed and the relationship that has main function for early compound faults diagnosis be retained only. Thus, the threshold constraint function represented by formula (8) is designed in this paper to tune the weights of signal relevance.

$$\delta\left(\text{z}_{\varepsilon}\right)=\begin{cases}0, & \text{otherwise}\\ \text{z}, & \text{z}\geq\varepsilon\end{cases} \tag{8}$$

where ε is the threshold for filtering secondary data relationships. In fact, a reasonable configuration of ε can prevent additional computational burdens caused by secondary relationships under the premise of satisfying decoupling.

Difference 3: Initial weight constraint term. Traditional SAE with randomly initializing weights may cause the extremely small gradient in the process of back propagation, which brings the issues of gradient diffusion and efficiency droop. So, a restraint condition for initial weights is developed to promote the diagnostic model to extract more useful features.

Specifically, suppose the input layer size of connection weight $\boldsymbol{W}_{1}^{\text{g}}$ is x and the output layer size is y, and define the function $\varphi_{xy}=\sqrt{6/x+y}$. Then the initial weight constraint term designed in this paper is shown in formula (9).

$$\begin{cases}\boldsymbol{W}_{1}^{\text{g}}=2\varphi_{xy}*\text{r}\left(x,y\right)-\varphi_{xy}\\ \boldsymbol{W}_{2}^{\text{g}}=(\boldsymbol{W}_{1}^{\text{g}})'\end{cases} \tag{9}$$

where $r\left(\cdot\right)$ represents a random function.

Difference 4: Weight regularization term. On the basis that the classical weight regularization term remains unchanged, the restraint condition for primary hidden layer's weights is developed to improve the discrimination of input signal feature extraction, as shown in formula (10),

$$\left\|\boldsymbol{W}_{1}^{\text{g}}\right\|=\sum_{i=1}^{nn}\sqrt{\sum_{j=1}^{mm}\left[\left(W_{1}^{\text{g}}\right)_{ij}\right]^{2}} \tag{10}$$

where nn and mm respectively denote the input dimension and characteristic dimension.

So, the cost function of designed DISAE model is finally expressed in (11) below,

$$\begin{cases}\underset{\boldsymbol{W},\boldsymbol{b}}{min}\frac{1}{2\text{TN}}\sum_{\text{pre}=1}^{\text{TN}}\left\|\boldsymbol{s}_{\text{pre}}^{\text{g}}.(\boldsymbol{s}_{\text{pre}}^{\text{g}})'-\bar{\boldsymbol{s}}_{\text{pre}}^{\text{g}}.(\bar{\boldsymbol{s}}_{\text{pre}}^{\text{g}})'\right\|_{2}^{2}+\\ a\sum_{\text{pre}=1}^{\text{TN}}\left\|\boldsymbol{h}_{\text{pre}}^{1}\right\|_{1}+\frac{b}{2}\sum_{W}\left(\sum\left(W_{1}^{\text{g}}\right)^{2}+\sum\left(W_{2}^{\text{g}}\right)^{2}\right)\\ \boldsymbol{W}_{1}^{\text{g}}=2\varphi_{xy}*\text{r}\left(x,y\right)-\varphi_{xy}\\ \boldsymbol{W}_{2}^{\text{g}}=(\boldsymbol{W}_{1}^{\text{g}})'\\ \left\|\boldsymbol{W}_{1}^{\text{g}}\right\|=\sum_{i=1}^{n}\sqrt{\sum_{j=1}^{m}\left[\left(W_{1}^{\text{g}}\right)_{ij}\right]^{2}}\end{cases} \tag{11}$$

where a and b are regulating parameters. When the well-trained DISAE model completes the feature mining operation, the model employs softmax classifier to achieve early compound faults diagnosis.

IV. EARLY COMPOUND FAULTS DIAGNOSIS USING DISAE-DASED MODEL

A. CASE STUDY1: EARLY COMPOUND FAULTS DIAGNOSIS FOR GEARBOX ARTIFICIAL DAMAGE DATASET

1) DIAGNOSTIC CONDITIONS

This proposed DISAE-based identification scheme is proved based on the gearbox dataset, and this test-rig is manufactured by Qianpeng Company [33]. The signal data in the set are sampled at a frequency of 5.12 kHz. The loaded gearbox runs in six working conditions: normal condition (NC), primary gear corrosive pitting failure condition (PFC), primary gear tooth-breaking failure condition (BFC), primary gear tooth-breaking and secondary gear wear-out failure condition

TABLE 1. Description of gearbox early compound faults diagnosis sample set.

Loaded (hp)	Set size	Health status	Health label
3	104	PFC	1
	104	PWFC	2
	104	BFC	3
	104	BWFC	4
	104	WFC	5
	104	NC	6

(BWFC), primary gear corrosive pitting and secondary gear wear-out failure condition (PWFC), secondary gear wear-out failure condition (WFC), and TABLE 1 shows detailed description of the dataset. Specifically, every health status contains 104 sample data, while every 512 points forms a sample data, therefore, the diagnostic-set of the loaded gearbox contains 624 samples.

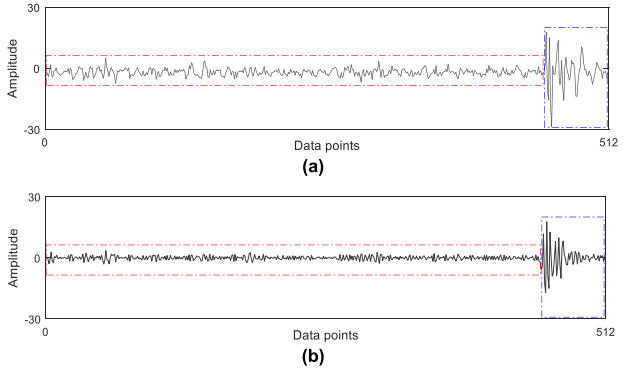

FIGURE 4. Sample data curves of gearbox before and after preprocessing. (a) Initial sample data curve; (b) Preprocessed sample data curve.

Firstly, the FSD based feature-enhanced and denoising measure formed in this study is adopted to conduct noise reduction in the initial gearbox sample data, and also enhance those hidden feature information. A PFC sample in the diagnostic set is employed to test the implementation effect, as shown in FIGURE 4 below. FIGURE 4(a) and FIGURE 4(b) are the initial sample data curve and the curve preprocessed by the proposed FSD based feature-enhanced and densing measure, respectively. Obviously, the proposed measure have the ability of signal denoising and feature enhancement, which can be seen from the red and blue boxes drawn in the figure respectively. So, it is necessary to adopt these preprocessed sample data as the model input for feature mining and status recognition.

2) VERIFICATION TESTS AND ANALYSIS

The structure of the proposed DISAE-based intelligent fault diagnostic model is: $L_{\text{Input}} + L_{\text{Primaryhidden}} + L_{\text{Secondaryhidden}} + L_{\text{Output}}$, where L denote the layer of the model. The neurons of L_{Input} and L_{Output} respectively consistent with dimension of input sample data and number of mechanical working condition, and the neurons of $L_{\text{Primaryhidden}}$ and $L_{\text{Secondaryhidden}}$ are respectively determined as 200 and 100 based on tests.

TABLE 2. Quantified Diagnosis Accuracy Of Loaded Gearbox With Early Compound Faults.

Test times	Accuracy (%)	Test times	Accuracy (%)	Average accuracy (%)
1	99.038	11	99.038	99.022
2	99.038	12	98.718	
3	98.718	13	99.359	
4	99.038	14	99.038	
5	99.038	15	99.038	
6	98.718	16	98.718	
7	99.038	17	99.038	
8	99.359	18	99.038	
9	99.038	19	99.359	
10	99.038	20	99.038	

Formula (9) and formula (10) are respectively employed to initialize the weights in $L_{\text{Primaryhidden}}$ and $L_{\text{Secondaryhidden}}$, and initial biases are set to zero vectors. Moreover, other parameters of DISAE-based model are determined experimentally. For avoiding contingency as well as randomness, twenty tests are carried out continuously. In every trial, fifty percent of input sample data is randomly picked as the training object, and then the rest of fifty percent ones as test object, finally the diagnostic result is given in FIGURE 5. It displays that the designed DISAE-based scheme can diagnose early compound faults excellently even for gearbox loaded, and the average recognition accuracy is over 99%.

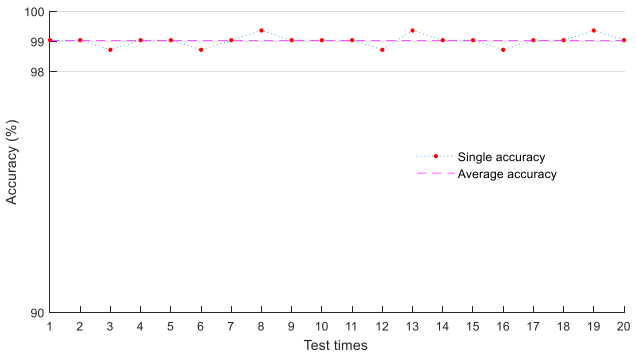

FIGURE 5. Early compound faults diagnosis results of loaded gearbox.

Furthermore, the numerical accuracy of every test from FIGURE 5 is listed in TABLE 2, which displays that the proposed FSD based feature-enhanced and denoising measure can well handle the issues caused by noise interference and load changes. So, for gearbox with workload, the average test result can reach 99.022% while the accuracy of each test shows slight fluctuations.

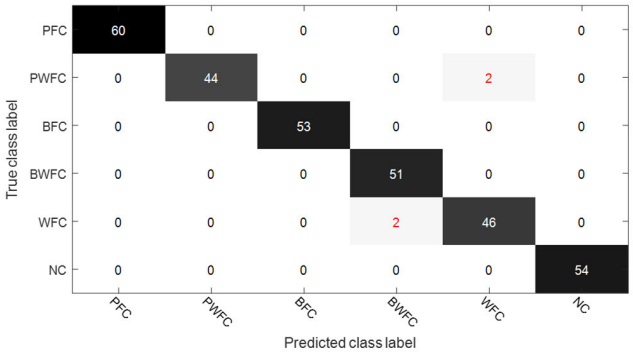

FIGURE 6. Confusion matrix of loaded gearbox early compound faults diagnosis for each health condition (3rd test).

TABLE 3. Diagnostic performance comparison of various models for gearbox early compound faults.

Model framework	Prediction results
Model 1: DWT+SVM+ANN [34]	87.949%
Model 2: CS+SAE [20]	96.058%
Model 3: SAE-based deep neural networks[28]	92.404%
Model 4: SAE with two hidden layer	92.003%
Model 5: The DISAE-based model	**99.022%**

DWT: DISCRETE WAVELET TRANSFORM; SVM: SUPPORT VECTOR MACHINES; ANN: ARTIFICIAL NEURAL NETWORK; CS: COMPRESSED SENSING.

Next, for the recognition test of different health states by the DISAE-based model, one of the test results is randomly chosen for sharing and studying, and is given in FIGURE 6. Those correctly identified target size is displayed on diagonal of the matrix, while the incorrectly predicted target size is marked on non diagonal, which shows that only a small amount of misdiagnosis occurs among the tested fault samples (marked in red), while other tested samples are well predicted (marked in the diagonal line). Specifically, due to such correlation between single-point fault features and its related early compound faults features, misdiagnosis mainly occurs between PWFC and WFC, BWFC and WFC. Therefore, compared with single-point fault, the features of early compound faults are more complex and the diagnosis is more difficult. Especially, the designed DISAE-based model does not cause missed diagnosis at all when diagnosing early compound faults of loaded gearbox.

3) COMPARATIVE TESTS AND ANALYSIS

For surveying the diagnostic performance about various models when handling early compound faults issues of the loaded gearbox above, contrast tests are developed by using those models generally applied at the mechanical health monitoring. To demonstrate the stability of each model, 20 consecutive diagnoses are made. TABLE 3 lists those comparison accuracy, which reveals that the deep models (Model 2-Model 5) generally possess higher diagnosis results on dealing with mass sample data contrasted to the general ANN model (Model 1). However, Model 2-Model 4 above ignored the trouble caused by noise interference and data relevance in the process of early compound faults diagnosis, bringing about unsatisfactory prediction results. Conversely, the designed DISAE-based diagnosis scheme possessed well diagnostic performance and uncomplicated framework.

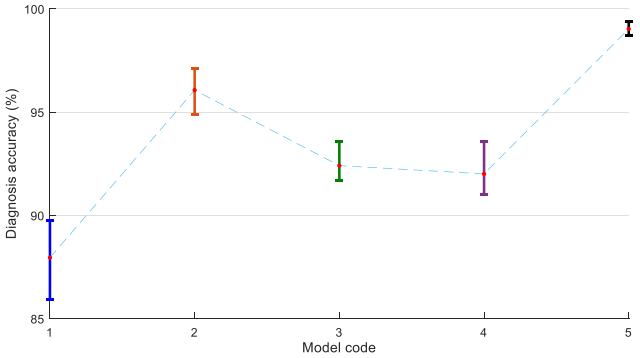

FIGURE 7. Diagnosis result fluctuation comparison of different models for gearbox early compound faults.

Further, a fluctuation curve of 20 tests for every diagnostic model is shown in FIGURE 7. In the figure, the line segments of different colors are the fluctuation range of the results, and the red dots represent the average accuracy. Compared with other diagnostic models, it can be seen that the model built in this study owns highest accuracy and smallest fluctuation even under the loaded condition of gearbox, reflecting the excellent diagnostic performance of the model.

B. CASE STUDY2: EARLY COMPOUND FAULTS DIAGNOSIS FOR ROLLING BEARINGS PARACTICAL FAULT DATASET

1) DIAGNOSTIC CONDITIONS

In order to fully demonstrate the effectiveness and superiority of the designed DISAE based scheme, those practical fault dataset of loaded rolling bearings provided by the University of Paderborn in Germany [35] is adopted for early compound faults identification test. The relevant parameters of the signal data in the set are as follows: 64 kHz (sampling frequency), 1500 rpm (running speed), 0.1Nm (load torque), and 1000 N (radial force). The loaded rolling bearings runs in three working conditions: Normal status (NS), Inner ring fault status (IRS) and Outer ring fault status (ORS), and the detailed information of the set is listed in TABLE 4. It shows that these fault types in dataset are diverse, containing a variety of practical early compound faults. Therefore, the intractability of this case is particularly higher contrasted to the loaded gearbox early compound faults prediction, thus the treat demand is inevitably raised.

2) VERIFICATION TESTS AND ANALYSIS

Firstly, the proposed FSD based feature-enhanced and denoising measure is still adopted to conduct noise reduction

TABLE 4. Dataset description of rolling bearings with practical fault.

Bearing code	Damage type	Damage position	Fault mode	Fault type	Fault degree
RB01	-	--	-	-	-
RB02	P	ORS	R	S	2
RB03	I	ORS	R	D	1
RB04	P	IRS	R	S	1
RB05	P	IRS	M	S	1
RB06	P	ORS	S	S	1
RB07	I	ORS	S	S	1
RB08	P	IRS	S	S	1
RB09	P	IRS	S	S	2
RB10	P	IRS	S	S	3
RB11	I	ORS+IRS	M	D	1
RB12	P	IRS+(ORS)	M	S	2
RB13	P	IRS+(ORS)	M	D	3

I: INDENTATIONS, P: PITTING, R: REPETITIVE, M: MULTIPLE, S: SINGLE POINT, D: DISTRIBUTED.

in the initial rolling bearings sample data, and also enhance those hidden feature information. In order to exhibit the superiority of the scheme, those hyper-parameters of the designed DISAE-based model in the loaded gearbox early compound faults diagnosis experiment are invariably used in Case study2. Moreover, the neurons of L_{Input} and L_{Output} are respectively 2560 and 13.

Subsequently, for avoiding contingency as well as randomness of test results, 20 consecutive diagnoses are made. In every trial, fifty percent of input sample data is randomly picked as the training object, and then the rest of fifty percent ones as test object, finally the diagnostic result is given in FIGURE 8. It displayed that the designed DISAE-based scheme can diagnose practical early compound faults excellently even for rolling bearings loaded, and the average recognition accuracy is more than 99%.

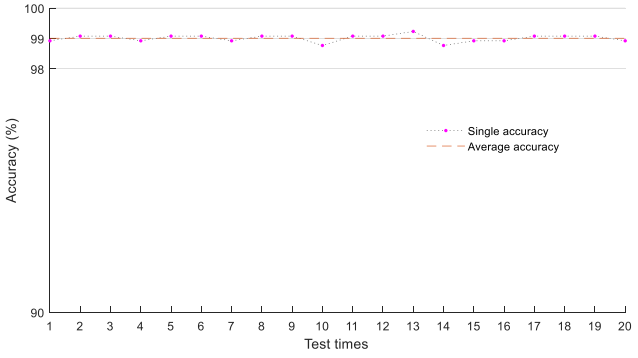

FIGURE 8. Diagnosis results of rolling bearings practical early compound faults sample set.

TABLE 5. Quantified diagnosis accuracy of loaded rolling bearings with practical early compound faults.

Test times	Accuracy (%)	Test times	Accuracy (%)	Average accuracy (%)
1	98.923	11	99.077	99.008
2	99.077	12	99.077	
3	99.077	13	99.231	
4	98.923	14	98.769	
5	99.077	15	98.923	
6	99.077	16	98.923	
7	98.923	17	99.077	
8	99.077	18	99.077	
9	99.077	19	99.077	
10	98.769	20	98.923	

Furthermore, the numerical accuracy of every test from FIGURE 8 is listed in TABLE 5, which displays that the proposed FSD based feature-enhanced and denoising measure can well handle the issues caused by noise interference and load changes for practical early compound faults recognition. So, for rolling bearings with workload, the average test result can reach 99.008% while the accuracy of each test shows slight fluctuations, reflecting the well superiority of the designed DISAE based scheme.

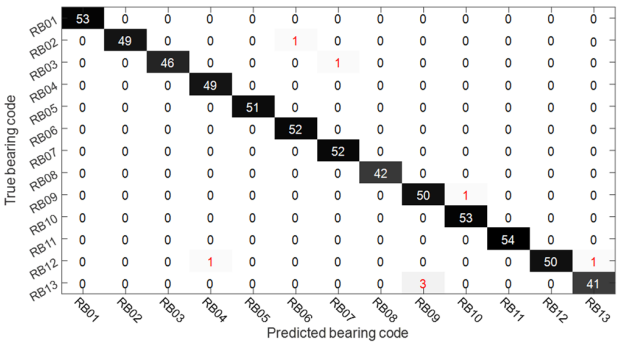

FIGURE 9. Confusion matrix of practical early compound faults diagnosis for each loaded rolling bearing.

Next, for the recognition test of different rolling bearings by the DISAE-based model, one of the test results is randomly chosen for sharing and studying, and is given in FIGURE 9. The abscissa and the ordinate represent the predicted code and the true code, respectively. Similarly, the correctly predicted target size is displayed on diagonal of the matrix, while the incorrectly predicted target size is marked on non diagonal, which shows that only a small amount of misdiagnosis occurs among the tested fault samples (marked in red), while other tested samples are well predicted (marked in the diagonal line).

Specifically, misdiagnosis mainly occurred in the following situations: 1) Between RB02 and RB06 with different

TABLE 6. Diagnostic performance comparison of various models for rolling bearings practical early compound faults.

Diagnosis model	Model 1: DWT+SVM+ANN [34]	Model 2: CS+SAE [20]	Model 3: SAE-based deep neural networks [28]	Model 4: SAE with two hidden layer	Model 5: The DISAE-based model
Average test accuracy	84.323%	95.054%	93.408%	86.278%	**99.008%**

Combination and Degree; 2) Between RB07 and RB03 with different Combination; 3) Between RB09 and RB10 with different Degree; 4) Between the single fault and its related compound fault (i.e. RB12 and RB04, RB13 and RB09, RB12 and RB13). Especially, the designed DISAE-based model does not cause missed diagnosis at all when diagnosing practical early compound faults of loaded rolling bearings, and it can even achieve the fault mode recognition and the severity degree determination simultaneously.

3) COMPARATIVE TESTS AND ANALYSIS

For reflecting the diagnostic performance about various models when handling practical early compound faults issues of the loaded rolling bearings above, contrast tests are developed. Moreover, to demonstrate the stability of each model, 20 consecutive diagnoses are made. TABLE 6 lists those comparison accuracy, apparently, similar to the early compound faults diagnosis of loaded gearbox, contrasted to the general ANN model (Model 1), the deep models (Model 2-Model 5) generally has superior diagnosis effect for the practical early compound faults. However, other deep learning models (Model 2-Model 4) are designed for equipment artificial damage diagnosis, and ignore the adverse effects of background noise and monitoring data coupling on mechanical health condition recognition. Therefore, the DISAE-based model proposed in this paper has higher diagnosis accuracy for the practical early compound faults.

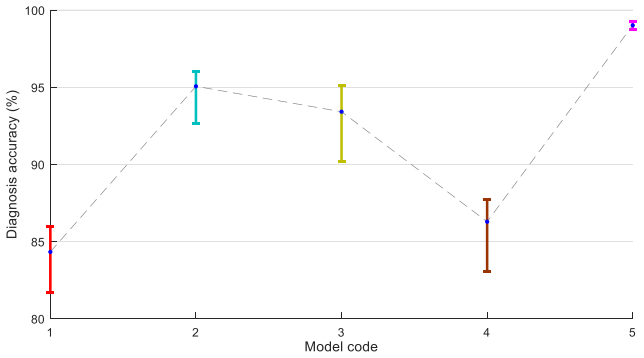

FIGURE 10. Diagnosis result fluctuation comparison of different models for rolling bearings practical early compound faults.

Finally, the accuracy fluctuation curve of 20 consecutive tests for each diagnosis model is shown in FIGURE 10. In the figure, the different segments indicate the diagnostic results variation, and blue solid points on the segment are the average diagnostic accuracy. It displays that although under the influence of load and background noise, compared with other models, the diagnosis accuracy of the proposed model is high and the fluctuation is slight. Therefore, for diagnosing practical early compound faults, the proposed model has excellent stability, powerful adaptivity and generalization performance, thus, it has an edge on early compound faults diagnosis in actual mechanical applications.

V. CONCLUSION

A denoising-integrated based sparse autoencoder (DISAE) model for early compound faults diagnosis was proposed in this paper, which can diagnose early compound faults in actual mechanical applications with high function (even in the case of multiple single-point fault and repetitive fault coexist). The tests displayed that the proposed DISAE model was significantly ahead of other diagnosis models on artificial and practical faults. Specifically, aiming at the difficulty in feature extration produced by signal weak-strength and background noise interference, by using the characteristic that the IMFs generated by EEMD have differential sensitivity to early fault, a feature-enhanced and denoising measure based on fault sensitivity degree (FSD) was designed, which screened IMFs sensitive to early faults to reconstruct the original diagnosis signal; for the obstacles caused by fault data coupling to diagnosis, a decoupling constraint was designed to reduce the adverse impact of data association; in view of the feature learning blindness of traditional-SAE, a restraint condition for initial weights and the primary hidden layer weight was designed, which achieved the effectiveness and diversity of DISAE feature learning; Finally, combined with the above strategies, an adaptive loss function was designed, and a DISAE model suitable for mass sample data was developed, which could wholly possess the higher test performance as well as fine stability.

In addition, the flexibility and adaptability of model hyper-parameter selection will become the focus of our future work, and the development of a diagnosis platform for practical engineering applications is also considered.

REFERENCES

[1] T. Sun, G. Yu, M. Gao, L. Zhao, C. Bai, and W. Yang, "Fault diagnosis methods based on machine learning and its applications for wind turbines: A review," *IEEE Access*, vol. 9, pp. 147481–147511, 2021.
[2] J. Yang, G. Xie, Y. Yang, Y. Zhang, and W. Liu, "Deep model integrated with data correlation analysis for multiple intermittent faults diagnosis," *ISA Trans.*, vol. 95, pp. 306–319, Dec. 2019.
[3] J. Yang, G. Xie, and Y. Yang, "A key-factor denoising strategy for quasi periodic non-stationary incipient faults diagnosis," *Measurement*, vol. 197, Jun. 2022, Art. no. 111304.
[4] A. M. Mahmoud, M. M. A. Lashin, F. Alrowais, and H. Karamti, "A dual soft-computing based on genetic algorithm and fuzzy logic defect recognition for gearbox and motors: Attempts toward optimal performance," *IEEE Access*, vol. 10, pp. 73956–73968, 2022.
[5] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, May 2016.
[6] G. Xie, J. Yang, and Y. Yang, "An improved sparse autoencoder and multilevel denoising strategy for diagnosing early multiple intermittent faults," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 2, pp. 869–880, Feb. 2022.

[7] L. Chetot, M. Egan, and J.-M. Gorce, "Joint identification and channel estimation for fault detection in industrial IoT with correlated sensors," *IEEE Access*, vol. 9, pp. 116692–116701, 2021.
[8] J. Singh, A. K. Darpe, and S. P. Singh, "Bearing damage assessment using Jensen–Rényi divergence based on EEMD," *Mech. Syst. Signal Process.*, vol. 87, pp. 307–339, Mar. 2017.
[9] Z. Su, B. Tang, Z. Liu, and Y. Qin, "Multi-fault diagnosis for rotating machinery based on orthogonal supervised linear local tangent space alignment and least square support vector machine," *Neurocomputing*, vol. 157, pp. 208–222, Jun. 2015.
[10] M. Kedadouche, M. Thomas, and A. Tahan, "A comparative study between empirical wavelet transforms and empirical mode decomposition methods: Application to bearing defect diagnosis," *Mech. Syst. Signal Process.*, vol. 81, pp. 88–107, Dec. 2016.
[11] F. Jia, Y. Lei, H. Shan, and J. Lin, "Early fault diagnosis of bearings using an improved spectral kurtosis by maximum correlated kurtosis deconvolution," *Sensors*, vol. 15, no. 11, pp. 29363–29377, Nov. 2015.
[12] Y. Lv, R. Yuan, T. Wang, H. Li, and G. Song, "Health degradation monitoring and early fault diagnosis of a rolling bearing based on CEEMDAN and improved MMSE," *Materials*, vol. 11, no. 6, pp. 1009–1030, Jun. 2018.
[13] H. Ren, J. F. Qu, Y. Chai, Q. Tang, and X. Ye, "Deep learning for fault diagnosis: The state of the art and challenge," *Control Decis.*, vol. 32, no. 8, pp. 1345–1358, Aug. 2017.
[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
[15] J. Yang, Y. Yang, and G. Xie, "Diagnosis of incipient fault based on sliding-scale resampling strategy and improved deep autoencoder," *IEEE Sensors J.*, vol. 20, no. 15, pp. 8336–8348, Aug. 2020.
[16] S. Yan, H. Shao, Y. Xiao, B. Liu, and J. Wan, "Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noises," *Robot. Comput.-Integr. Manuf.*, vol. 79, Feb. 2023, Art. no. 102441.
[17] A. Qamar, I. H. Al-Kharsan, Z. Uddin, and A. Alkhayyat, "Grounding grid fault diagnosis with emphasis on substation electromagnetic interference," *IEEE Access*, vol. 10, pp. 15217–15226, 2022.
[18] Z. He, H. Shao, P. Wang, J. Lin, J. Cheng, and Y. Yang, "Deep transfer multi-wavelet auto-encoder for intelligent fault diagnosis of gearbox with few target training samples," *Knowl.-Based Syst.*, vol. 191, Mar. 2020, Art. no. 105313.
[19] W. Xu, Y. Wan, T. Zuo, and X. Sha, "Transfer learning based data feature transfer for fault diagnosis," *IEEE Access*, vol. 8, pp. 76120–76129, 2020.
[20] J. Sun, C. Yan, and J. Wen, "Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 1, pp. 185–195, Jan. 2018.
[21] J. Yang, G. Xie, Y. Yang, X. Li, L. Mu, S. Takahashi, and H. Mochizuki, "An improved deep network for intelligent diagnosis of machinery faults with similar features," *IEEJ Trans. Electr. Electron. Eng.*, vol. 14, no. 12, pp. 1851–1864, Dec. 2019.
[22] H. Shao, M. Xia, J. Wan, and C. W. de Silva, "Modified stacked autoencoder using adaptive Morlet wavelet for intelligent fault diagnosis of rotating machinery," *IEEE/ASME Trans. Mechatronics*, vol. 27, no. 1, pp. 24–33, Feb. 2022.
[23] 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.
[24] J. Yang, G. Xie, and Y. Yang, "An improved ensemble fusion autoencoder model for fault diagnosis from imbalanced and incomplete data," *Control Eng. Pract.*, vol. 98, May 2020, Art. no. 104358.
[25] H. Karamti, M. M. A. Lashin, F. Alrowais, and A. M. Mahmoud, "Accompanying deep framework for faults in motor and gearbox with disproportion vibrational samples," *Neural Comput. Appl.*, vol. 12, pp. 1–18, Dec. 2022.
[26] X. Guo, C. Shen, and L. Chen, "Deep fault recognizer: An integrated model to denoise and extract features for fault diagnosis in rotating machinery," *Appl. Sci.*, vol. 7, no. 1, p. 41, Jan. 2017.
[27] F. Jia, Y. G. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 303–315, May 2016.
[28] M. Sohaib, C. H. Kim, and J. M. Kim, "A hybrid feature model and deep-learning-based bearing fault diagnosis," *Sensors*, vol. 17, pp. 2876–2891, Dec. 2017.
[29] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu, "Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery," *IEEE Access*, vol. 5, pp. 15066–15079, 2017.
[30] P. Flandrin, G. Rilling, and P. Gonçalves, "Empirical mode decomposition as a filter bank," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 112–114, Feb. 2004.
[31] S. J. Loutridis, "Damage detection in gear systems using empirical mode decomposition," *Eng. Struct.*, vol. 26, no. 12, pp. 1833–1841, Oct. 2004.
[32] S. J. Loutridis, "Instantaneous energy density as a feature for gear fault detection," *Mech. Syst. Signal Process.*, vol. 20, no. 5, pp. 1239–1253, Jul. 2006.
[33] C. Lu, Z. Wang, and B. Zhou, "Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification," *Adv. Eng. Informat.*, vol. 32, pp. 139–151, Apr. 2017.
[34] S. Tyagi and S. K. Panigrahi, "An SVM-ANN hybrid classifier for diagnosis of gear fault," *Appl. Artif. Intell.*, vol. 31, no. 3, pp. 209–231, May 2017.
[35] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, "Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification," in *Proc. Eur. Conf. Prognostics Health Manage. Soc.*, 2016, pp. 1–18.

JING YANG received the B.S. and M.S. degrees from Central South University (CSU), China, in 2008 and 2011, respectively, and the D.E. degree from the Xi'an University of Technology (XUT), China, in 2021. She is currently a Professor with the School of Mechatronics and Automotive Engineering, Tianshui Normal University, China. Her research interests include machine learning, pattern recognition, and data based fault diagnosis.

GUO XIE (Member, IEEE) received the B.S. and M.S. degrees from the Xi'an University of Technology (XUT), China, in 2005 and 2008, respectively, and the D.E. degree from Nihon University, Tokyo, Japan, in 2013. He is currently a Professor with the School of Automation and Information Engineering, XUT. His research interests include safety and reliability of railway systems, optimal control, and stochastic control. He is a member of the CAA and CCF. He received the Japanese Government Scholarship from the Japanese Ministry of Education, Culture, Sports, Science and Technology (Monbukagakusho).

YANXI YANG received the B.S., M.S., and Ph.D. degrees from the Xi'an University of Technology (XUT), China. He is currently a Professor with the School of Automation and Information Engineering, XUT. His research interests include complex system control, intelligent diagnosis, digital image processing, machine learning, intelligent robot, pattern recognition, and computer vision.