Research Article

An Enhanced Gated Recurrent Unit-Based Adaptive Fault Diagnosis of Rotating Machinery

Zhen Li,[1] **Saleem Riaz** [ID],[2] **Muhammad Waqas,**[3] **and Munira Batool** [ID][4]

[1]*Department of Automotive Engineering, Sichuan Vocational and Technical College Communications, Chengdu City, Sichuan Province 611130, China*
[2]*Sool of Automation, Northwestern Polytechnical University, Shaanxi, Xi'an, China*
[3]*School of Electrical Engineering, Beijing University of Technology, No. 100 Ping Le Yuan, Beijing 100124, China*
[4]*Department of Electrical Engineering, University of Engineering and Technology, Taxila, Pakistan*

Correspondence should be addressed to Munira Batool; munira.batool@uettaxila.edu.pk

As the most basic component of rotating machinery, rolling bearing frequently works in harsh environments and complex working conditions, and its health status affects seriously the working efficiency. The health statuses of rolling bearing can not only reduce equipment maintenance costs but also contribute to reducing major accidents. Based on this, an adaptive diagnosis method that combines deep gated recurrent unit (DGRU) with wavelet packet decomposition (WPD) and extreme learning machine (ELM) is proposed for rolling bearing. Firstly, WPD is utilized to eliminate the noise of data. Secondly, DGRU is designed to extract the representative features of denoised data. Finally, ELM is utilized to output the diagnosis results. Massive results prove that the superiority and robustness of our approach outperform existing popular methods. Additionally, the proposed method can also achieve powerful antinoise ability.

1. Introduction

The health state of rotating parts directly affects the operation reliability of the whole mechanical system [1–3]. Once the rotating parts fail, it will cause serious accidents. Machinery and equipment are widely used in various industrial scenarios and electrified transmission systems, and sometimes, this equipment may run under unfavorable conditions, such as high temperature, high humidity, and high load environment, which will eventually lead to equipment failure and cause high maintenance of high maintenance cost, serious property loss, and safety hazards. The faults of mechanical equipment can usually be attributed to different types of faults, including driving inverter faults, stator faults, rotor faults, and bearing failures. According to statistics, bearing faults are the most common types of faults, and the incidence of failure reaches 30% to 40% [4–6]. Since bearing is the most vulnerable parts of mechanical equipment, the diagnosis of accurate bearing faults has been a study of

engineers and scientists in the past few decades. Therefore, an effective rotating machinery condition monitoring and fault identification system are established to ensure the safe operation of equipment and personnel safety. As the most basic component, bearings frequently work in harsh environments and complex working conditions, and its health status affects seriously the working efficiency [7–10]. The health statuses of rolling bearing can not only reduce equipment maintenance costs but also contribute to reducing major accidents [11, 12].

Fault diagnosis methods based on deep learning are booming. This method is based on data-driven methods and integrates feature learning and intelligent recognition. Compared with traditional methods, it gets signal preprocessing and expert knowledge, especially when analyzing massive monitoring data. Bearing fault diagnosis has long been a hot topic of research [13–15]. Deep learning methods have made lots of achievements on the advent of avoiding extracting manually features [16–19]. However, most of

these models can exhibit excellent performance under the premise that the data have the same data distribution. Unfortunately, it is difficult or even unrealistic to meet the premise when considering complicated operating conditions, the degradation of equipment performance [20–23]. The diagnosis performance of most deep learning models will be greatly reduced when the premise is not accessible. Some researchers try to use fine-tuning algorithm or retraining model strategy to tackle the above issue, but a few labeled target data still need to be provided. Collecting labeled data requires lots of expenses or is even impossible in actual scenarios. Hence, it is very necessary to explore some promising methods that can apply the knowledge from relevant areas to solve problems. The generative adversarial network (GAN) was innovatively designed by Goodfellow et al., which utilizes the adversary between generators and discriminators for generating data with the same distribution as the raw. However, the adversarial mechanism renders the model challenging to be in equilibrium. Hence, many scholars have offered research solutions for further improving the GAN model. Radford et al. [24] proposed deep convolutional generative adversarial networks (DCGANs) fusing CNN with GAN, which avoids the GAN model to converge the learned data distribution to the same one.

Unlike DAE, DBN, CNN, and GAN, RNN is still in its infancy in diagnosis field. The main reason is conventional RNN that has an unignored problem-gradient vanishing [25]. Gated recurrent unit (GRU) can solve this problem [26]. GRU, as the newest variant of RNN, has achieved huge success in fault diagnosis issues [27, 28]. Thus, in this paper, GRU-based network is developed to effectively solve problems. However, the vibration signals are always contaminated by the noise that heavily influence the diagnosis performance of network [29]. Thus, wavelet packet decomposition (WPD) that has been recognized as an effective vibration signals denoising method is used for eliminating the noise of vibration signals [30–34].

An adaptive diagnosis method that combines deep gated recurrent unit (DGRU) with wavelet packet decomposition (WPD) and extreme learning machine (ELM) is proposed for rolling bearing. Firstly, WPD is utilized to eliminate the noise of data. Secondly, DGRU is designed to extract the representative features of denoised data. Finally, ELM is utilized to output the diagnosis results. Massive results prove that the superiority and robustness of our method outperform existing popular methods. Additionally, the proposed method can also achieve powerful antinoise ability.

The specific arrangements of this paper are as follows: Section 2 describes basic theoretical knowledge. A concrete introduction of our method is given in Section 3. Section 4 analyzes the effectiveness. Conclusions are generalized in Section 5.

2. The Brief Theory of Gated Recurrent Unit

Similar to long short-term memory neural network (LSTM), gated recurrent unit (GRU) is also a method proposed to solve the problem, but it is simpler than LSTM [25, 26]. GRU uses an update gate and a reset gate. These two gates together determine the output of GRU [35]. The specific structure is shown in Figure 1.

$$
\begin{aligned}
z_t &= \sigma\left(E_z X_t + F_z Y_{t-1}\right), \\
r_t &= \sigma\left(E_r X_t + F_r Y_{t-1}\right), \\
H_t &= \tanh\left(E x_t + F\left(r_t Y_{t-1}\right)\right), \\
Y_t &= \left(1 - z_t\right) Y_{t-1} + z_t H_t,
\end{aligned}
\tag{1}
$$

where σ and $\tanh$ denote the sigmoid and tangent activation functions. E_z, E_r, E, F_z, F_z, and F are the weight matrices and element-wise multiplications. y_t is an activation at time t, and h_t means a candidate activation.

3. The Proposed Method

Rotating machinery is applied to many fields. Rolling bearing is a necessary component to ensure the normal operation of rotating machinery. It has a direct impact on the accuracy and reliability of rotating machinery equipment. Therefore, rolling bearing faults are one of the most common reasons for rotating mechanical failures. Due to the long-term operation of rotating machinery under harsh and complex conditions, it is inevitable that faults will occur. Therefore, the state of machinery must be monitored in time to diagnose faults as soon as possible. One of the four key tasks is to find out whether the rotation of the machine is abnormal or not, and to predict the severity of the rotation of the machine. Due to the higher requirements for high performance, safety, and reliability, fault diagnosis of rotating machinery becomes not only more and more important but also more and more difficult. Therefore, in recent decades, rotating machinery fault diagnosis has received more and more attention and considerable development. This paper develops a new rotating machinery fault diagnosis method that combines a deep gated recurrent unit (DGRU) with wavelet packet decomposition (WPD) and extreme learning machine (ELM) to identify locomotive bearing fault conditions.

3.1. Wavelet Packet Decomposition Denosing. WPD is generally used to deal with nonstationary signals. It can analyze both time domain and frequency domain, and analyze the characteristics of signals locally. Wavelet transform decomposition mainly focuses on low-frequency signals and cannot decompose high-frequency signals containing a large amount of detailed information, such as rolling bearing vibration signals, remote sensing images, seismic signals, and biomedical signals. WPD is based on the idea of multiresolution analysis, that is, the signal can be decomposed and reconstructed in different frequency bands under the wavelet basis, which is suitable for dealing with discontinuous and nonstationary signals. WPD makes up for the shortcomings of wavelet transform. It can solve the signal energy from different decomposition scales. The multilevel division of frequency band can decompose not only low-frequency signals but also high-frequency signals, making the division of signals more precise. The signal decomposition process reflects the relationship between the
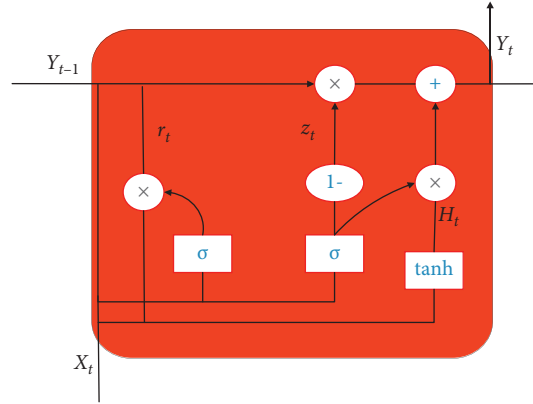
FIGURE 1: Structure of GRU.

wide-band signal and the fine band signal. The nonstationary vibration signal can approach the fault characteristic frequency of the system through WPD to obtain the instantaneous signal containing stationary signal components. The decomposition algorithm principle is to calculate the average and the difference between the first number and the average, and the system fault can be detected by analyzing the energy distribution in different frequency bands; WPD has neither redundancy nor omission. The vibration signals often contain noise that greatly influences the diagnosis accuracy. Thus, it is essential to eliminate the noise firstly. WPD is viewed as an effective method for vibration signal denoising [31]. WPD splits into two branches, low and high frequencies at all decomposition processes [36]. The three-layer binary trees of WPD are shown in Figure 2. The steps of signals denoising using WPD is illustrated in Table 1.

3.2. Deep Gated Recurrent Unit Construction. The operating conditions of mechanical equipment are changing, and the data label information of the training set data under most of the working conditions is unknown, and it is difficult to effectively train intelligent identification models. However, the training process of the above methods uses a small amount of labeling data, so it cannot be used to solve the problem of health status recognition under the case of unknown label information. Rotating machinery plays a vital role in the application of coal industry. Due to the long-term operation of rotating machinery under harsh and complex conditions, it is inevitable that faults will occur. Therefore, the state of machinery must be monitored in time to diagnose faults as soon as possible. One of the four key tasks is to find out whether the rotation of the machine is abnormal or not, and to predict the severity of the rotation of the machine. Due to the higher requirements for high performance, safety, and reliability, fault diagnosis of rotating machinery becomes not only more and more important but also more and more difficult. Therefore, in recent decades, rotating machinery fault diagnosis has received more and more attention and considerable development.

The health state of rotating parts such as bearings and gears directly affect the operation reliability of the whole mechanical system. Once the rotating parts fail, it will cause serious safety accidents and huge economic losses. Therefore, the establishment of an effective rotating machinery condition monitoring and fault identification system is of great significance to ensure the safe operation of equipment and personnel safety. Signal processing technology is an important subject of rotating machinery fault diagnosis, which has been widely used in various industrial fields. In addition, due to more and more attention, artificial intelligence technology has also been applied to rotating machinery fault diagnosis. Based on this, the illustration of DGRU is presented in Figure 3. $X(s)$ means the denoised data processed by WPD. $Y(s)$ and $Z(s)$ represent the extracted first-layer and second-layer features.

$$\sigma(t) = \frac{1}{1 + e^{-t}}. \tag{2}$$

The loss function is cross-entropy loss function that estimates the difference between the predicted label and actual label.

$$L(x, y) = -\sum_{i=1}^{N} y_i \log_2 x_i, \tag{3}$$

where x_i denotes the actual label and y_i is the predicted label.

3.3. Extreme Learning Machine Classification. ELM is the result of improving the algorithm [37]. The structure is shown in Figure 4. x and t are the input vectors and output labels respectively, (W, b) are the weights and bias of input layer and hidden layer, β is the value of the implication layer and the output layer. The difference from the BP neural network trained by gradient descent is that the weights generated during the training of the limit learning machine, and there is no need to adjust after generation. The specific calculation formulas are shown in equations (4), (5), and (6), the hidden layer vector h_i ($i = 1, 2, \ldots, n$), where n is the number of input samples. T is the matrix composed of sample label vector, and H^+ is the generalized inverse matrix of H. Based on the principle of the least square method, the whole process does not need feedback iterative adjustment.
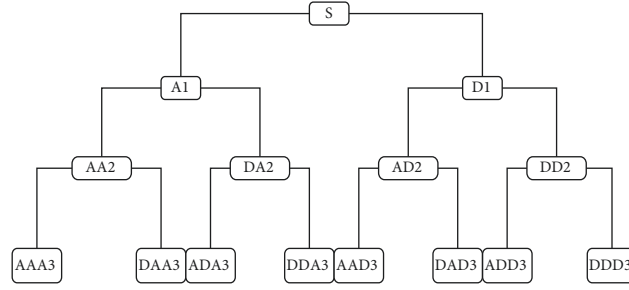
Figure 2: Three-layer binary trees WPD.

Table 1: The steps of signals denoising using WPD.

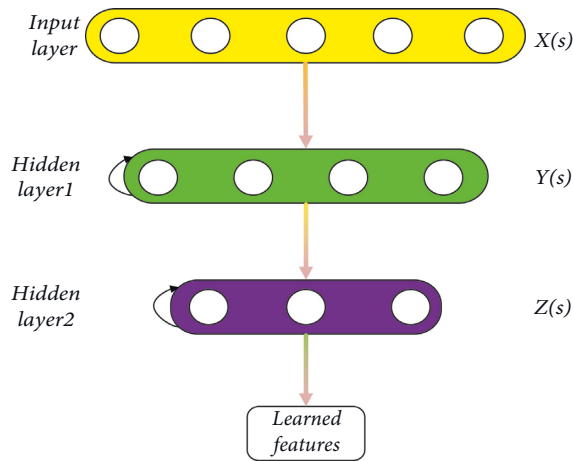| Steps	Detailed description
A. Signal decomposition	Select the wavelet function to decompose the noisy signals. Therefore, the wavelet packet coefficients at each level are obtained.
B. Thresholding	A threshold rule is used to decompose the coefficients to eliminate most of noisy coefficients
C. Signal reconstruction	The inverse WPD of each scale is performed by using the obtained approximate coefficients and the detailed coefficients of denoising.

Figure 3: Structure of DGRU.

Figure 4: The basic framework.

$$H = \begin{bmatrix} F(\beta_1, \gamma_1, y_1) & \cdots & F(\beta_K, \gamma_K, y_1) \\ \vdots & \vdots & \vdots \\ F(\beta_1, \gamma_1, y_M) & \cdots & F(\beta_K, \gamma_K, y_M) \end{bmatrix}_{M \times K}, \quad (4)$$

$$\beta = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_K \end{bmatrix}_{K \times N}, \text{ and } T = \begin{bmatrix} t_1 \\ \vdots \\ t_M \end{bmatrix}_{M \times N}. \quad (5)$$

Known from Reference [38], β can be represented by

$$\beta = H^+ T, \quad (6)$$

where H^+ is the Moore–Penrose generalized inverse of H.

3.4. General Steps. The health state of rotating parts directly affects the operation reliability of the whole mechanical system. Once the rotating parts fail, it will cause serious accidents. Therefore, the establishment of an effective rotating machinery condition monitoring and fault identification system is to ensure the safe operation of equipment and personnel safety. An adaptive diagnosis method that combines DGRU with WPD and ELM is proposed. The updating process of our method is described in Figure 5 with the following three steps.

(i) Step 1: measure data from rotating machinery

(ii) Step 2: eliminate the noise of vibration signals by using WPD

(iii) Step 3: divide the denoised signals into trained and tested samples

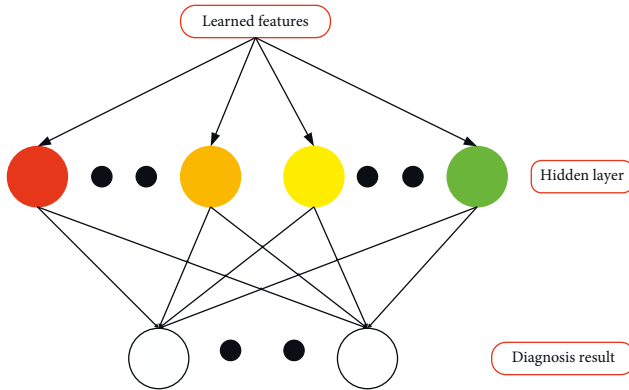(iv) Step 4: DGRU with ELM is constructed to diagnose railway locomotive bearing faults
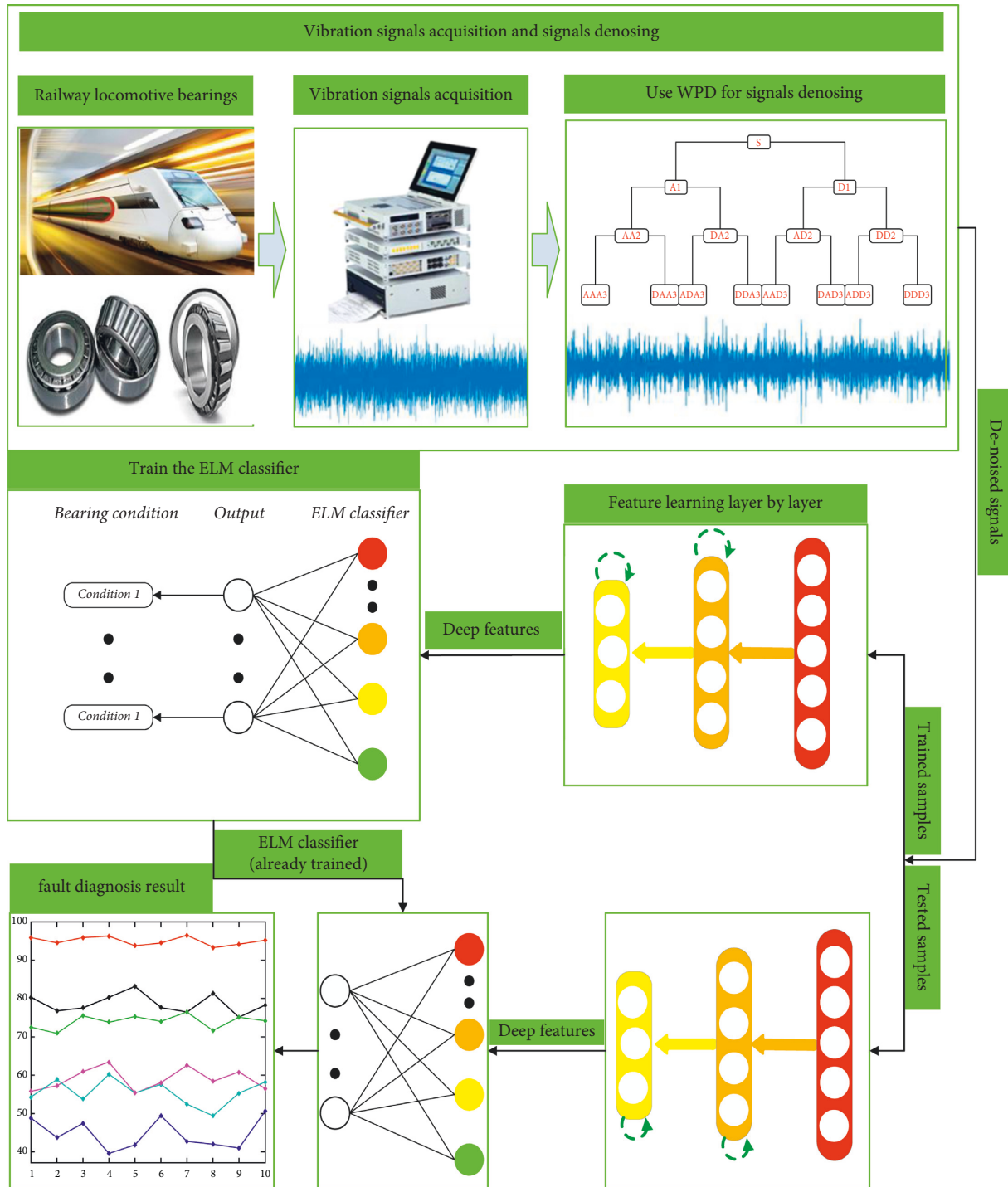
FIGURE 5: The framework of our approach.

    (v) Step 5: the constructed model is used for learning the trained-samples features and verified by the tested-samples features

    (vi) Step 6: output the diagnosis result.

4. Experimental Verification

4.1. Experimental Data Description. Because rotating machinery under harsh and complex conditions, it is inevitable that faults will occur. Therefore, the state of machinery must be monitored in time to diagnose faults as soon as possible. One of the four key tasks is to find out whether the rotation of the machine is abnormal or not, and to predict the severity of the rotation of the machine. Due to the higher requirements for high performance, safety, and reliability, fault diagnosis of rotating machinery becomes not only more and more important but also more and more difficult.

In this section, a representative simulating high-speed bearings dataset is selected to validate the feasibility of the proposed method. Various indicators are adopted to prove the effectiveness of data augmentation. Simulated fault diagnosis experiments are conducted separately with other approaches to corroborate the superiority of the algorithm in this paper. The simulated high-speed bearing dataset is applied to the laboratory dataset. The dataset is provided by railway locomotive bearing. According to different health conditions and three damage levels, 9 health categories with 12.8 kHz sampling frequency are obtained to form the dataset. The experimental platform presented in Figures 6 and 7 represents the specific faults of rolling bearings.

4.2. Compared with Traditional Methods. To evaluate the effectiveness of our approach for noisy signals, the collected data are added white Gaussian noise (WGN) in this part as shown in Figure 8. The noisy signals are described as

$$V(s) = v(s) + k \times n(s), \tag{7}$$

where $V(s)$ represents the noisy signals, $v(s)$ is the collected vibration signals, $n(s)$ is the WGN, and k is the coefficient. Larger k means heavier noise.

In this part, k is 0.4 and then we obtain the noisy signals. Then, the noisy signals are one input. The denoised signals that are processed by WPD are another input. Figure 9 describes the noisy data and the denoised data of each condition, each condition contains 8,192 data points. There are two important points to be explained: (1) For the proposed method and DGRU with SVM classifier, the only input is the noisy signals. (2) SVM and ANN both have two inputs, the noisy signals and the denoised signals.

To demonstrate the superiority of our, six methods are considered as the comparison methods. More details about these methods are provided as follows. As in this article, the optimization algorithm is the Adam algorithm. The learning rate is 0.0002 in all experiments. The relevant parameters of these methods are determined by relevant literature and experiments so that these methods could achieve the best recognition performance for different diagnosis tasks. These results are depicted in Figure 10. The confusion matrix is illustrated in Figure 11. Table 2 shows the results per method in all tasks.

It is obviously observed from Table 3 that the average accuracy of the proposed method is 94.98%, which is obviously higher than the other five methods, which are 78.64%, 55.47%, 73.95%, 44.61%, and 58.85%, respectively. The standard deviation is only 1.10, obviously lower than the other five methods which are 2.35, 3.28, 2.25, 3.96, and 3.22. The results present that: (1) Comparing all the methods, we can clearly observe that DGRU, SVM, and ANN are all sensitive about the noise. (2) Comparing Method 1 with Method 2, it can be known that the denoised signals could make much better diagnosis accuracy. It also proves the necessity and effectiveness of the noisy signals processed by WPD. (3) By comparing Method 1, Method 4, and Method 6, it can be seen that the proposed method has much more accurate and robust performance than SVM and ANN. The
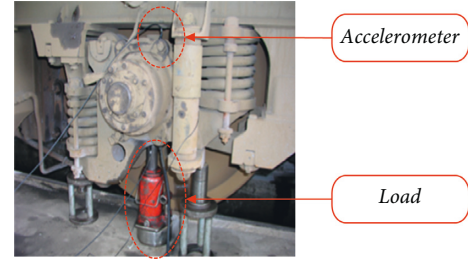

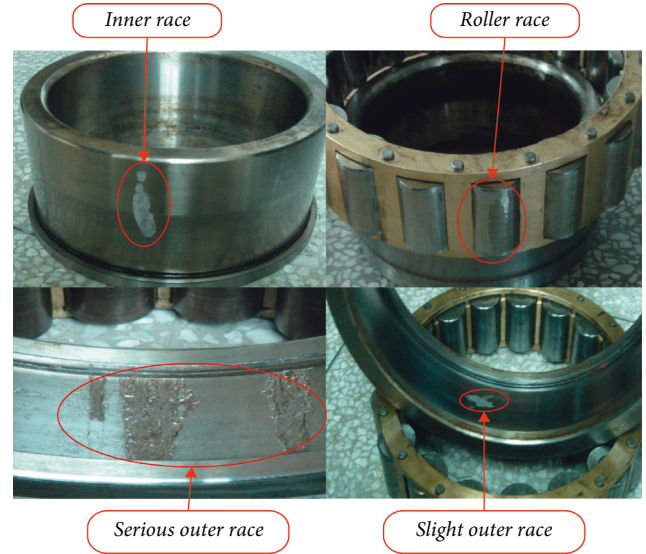FIGURE 6: Experimental locomotive bearing device.

FIGURE 7: Four types of rolling bearing faults.

main reason is that the deep architecture has a more powerful ability to learn functions. Therefore, it can automatically learn more appropriate internal error characteristics from the inputs and provide more reliable conclusions.

4.3. The Antinoise Ability of the Proposed Method. This part is mainly to research the influence of different noisy signals and the antinoise ability of the proposed method. To avoid the chance of result, each condition runs 5 times. The description of each condition and the average accuracy is shown in Table 4. The noisy signals represent the signals contain noise; the denoised signals represent the noisy signals processed by WPD. The concrete diagnosis result of each trail is shown in Figure 12.

The average accuracy and standard deviation of each condition are shown in Table 4. It can be known that for the DGRU with ELM classifier, the denoised signals could lead to much better accuracy and robust performance than the noisy signals. With more noise of the vibration signals, the diagnosis accuracy becomes lower and more and more unstable. No matter how powerful the denoising method is, the denoised signals could not be better than the normal vibration signals. It also can be found that with the increase of noise, the result of WPD dealing with noisy signals is also getting worse. However,
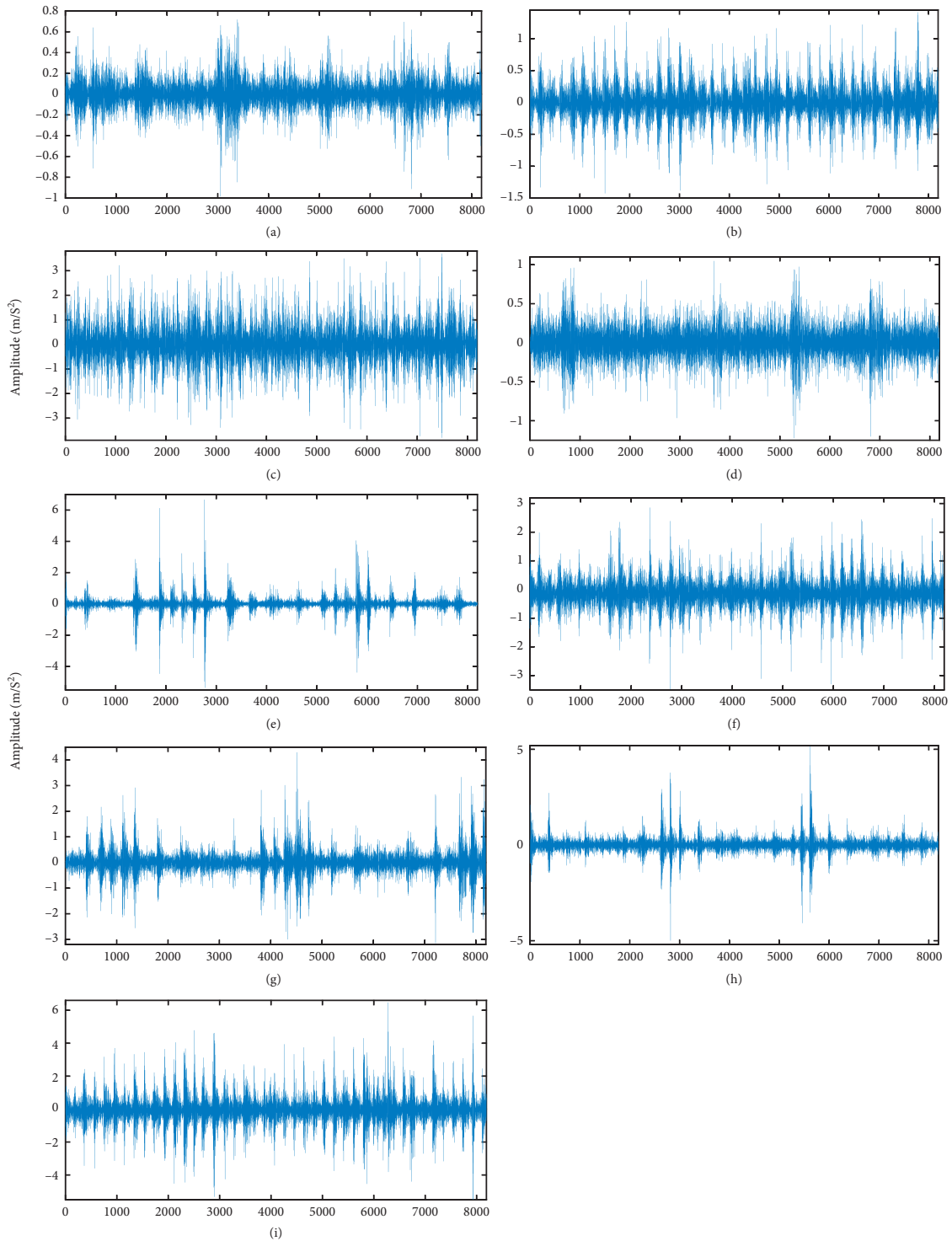
FIGURE 8: The collected vibration signals of the nine conditions.

the diagnosis result is also higher than 90%. According to the above-mentioned and diagnosis results, it can be confirmed that the proposed method has a powerful antinoise ability.

The classifiers with the identical parameters are employed for the same purpose. Ablation experiments are required for the CNN classifier to ensure that it is resistant to engineering noise interference. The settings and results of the ablation
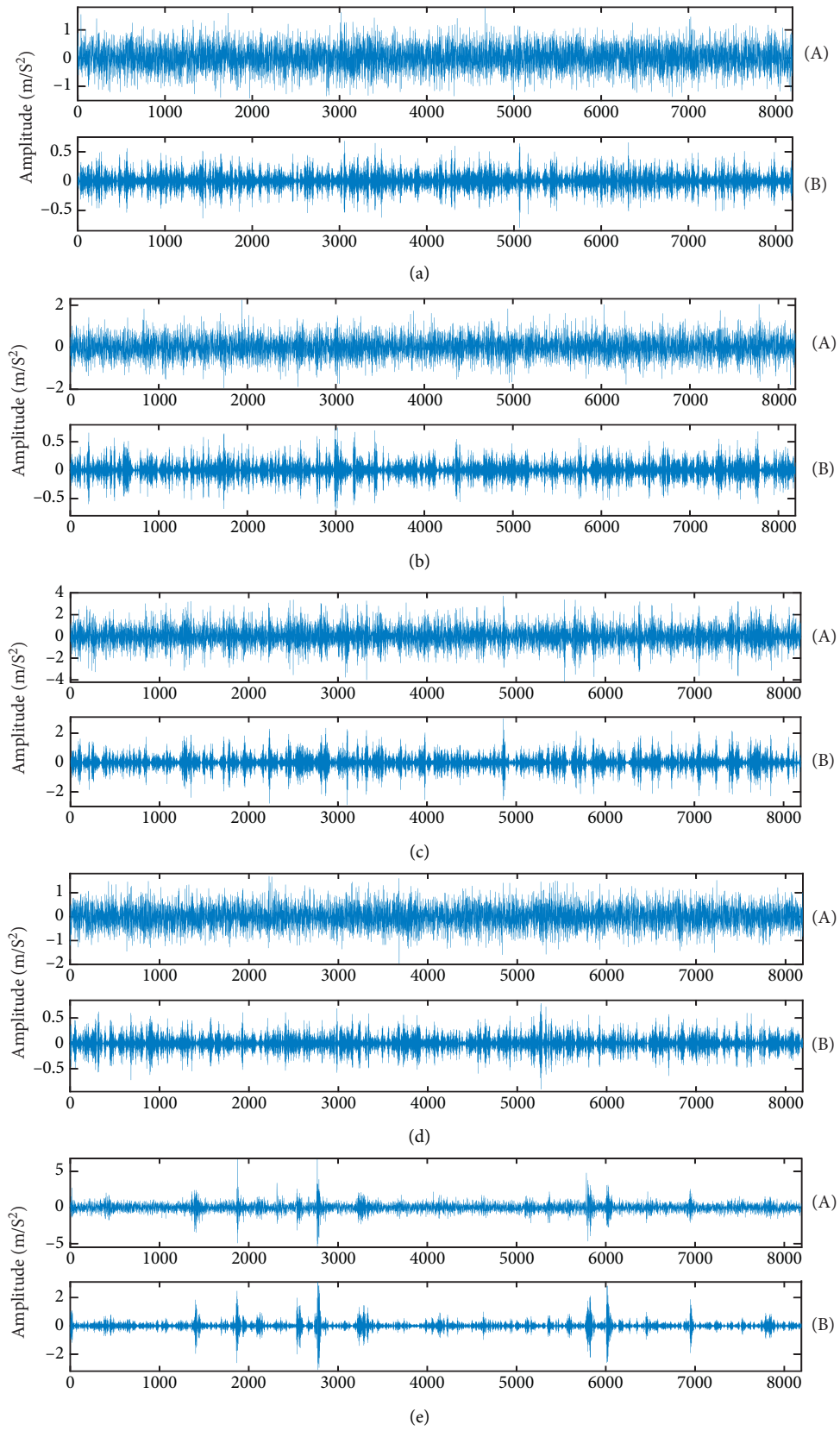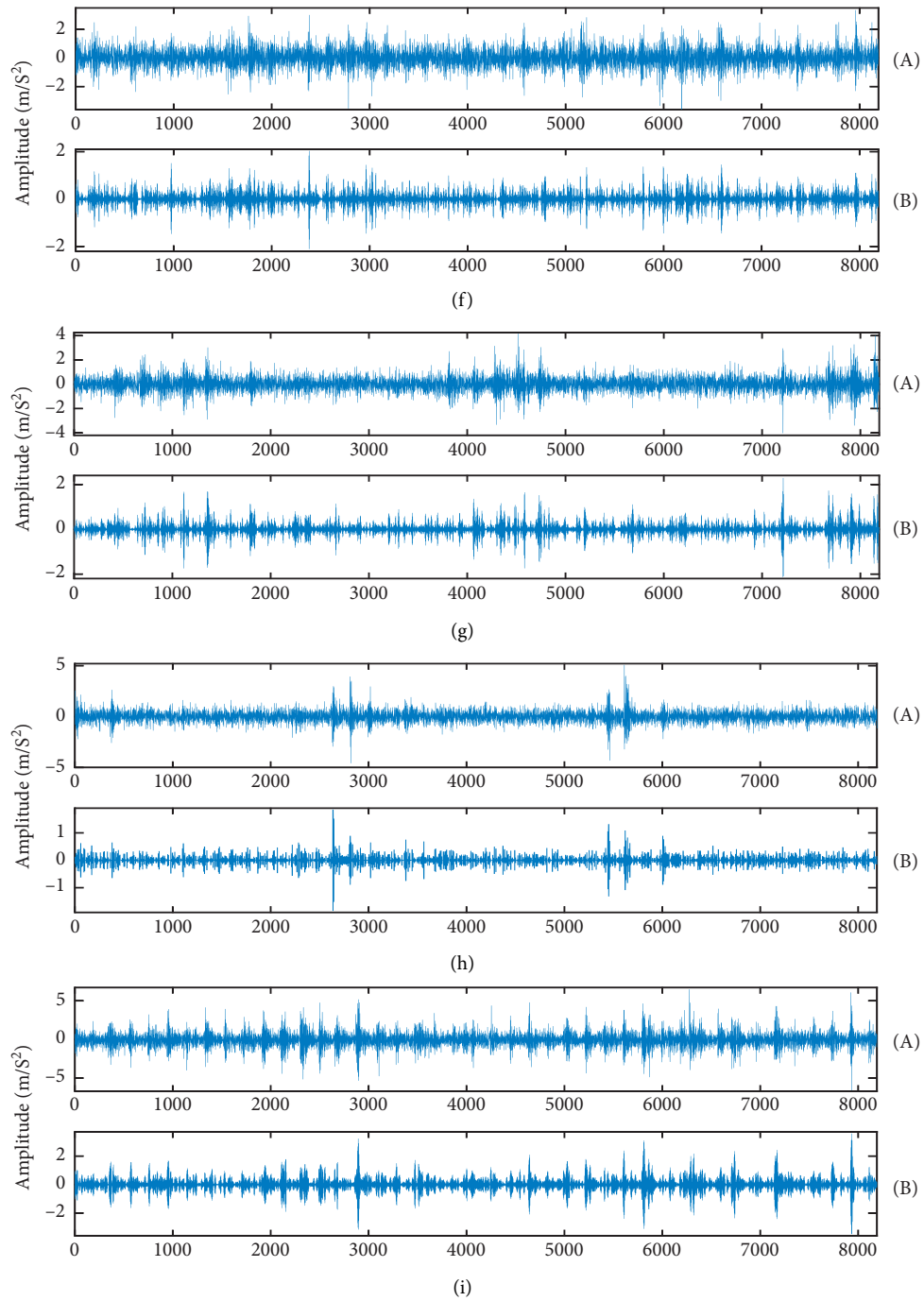
FIGURE 9: Continued.

FIGURE 9: The noisy signals and the denoised signals of the nine conditions. (A) Noisy signals. (B) Denoised signals.
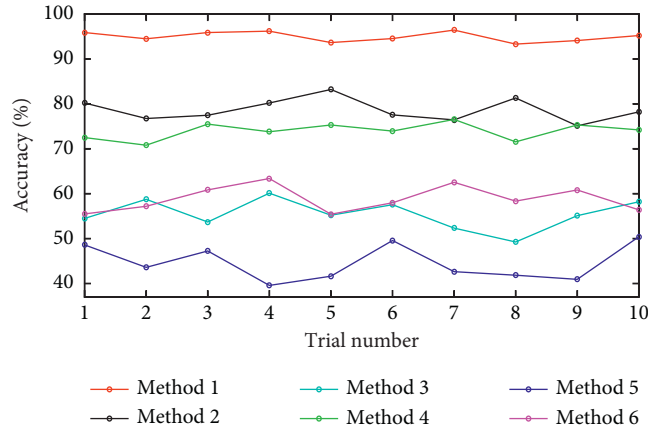
Figure 10: Diagnosis results of the 10 trials.
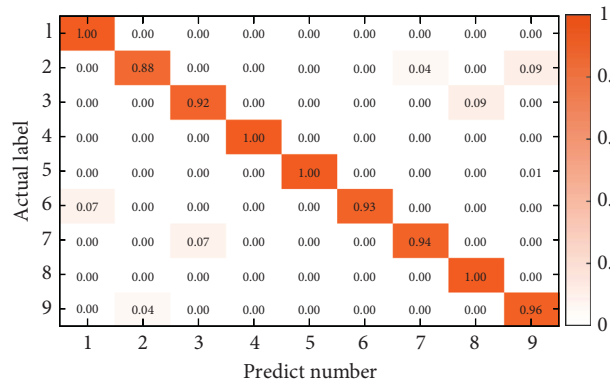
Figure 11: The confusion matrix using the proposed method (the first trial).

Table 2: Comparison of methodologies.

Method	Description	Inputs
1	Suggested technique	Noisy signals
2	DGRU + ELM classifier	Noisy signals
3	SVM	Noisy signals
4	SVM	Denoised signals
5	ANN	Noisy signals
6	ANN	Denoised signals

Table 3: The concrete classification accuracy and standard deviation per method.

Method	Average accuracy (%)	Standard deviation
1	94.98	**1.10**
2	78.64	2.35
3	55.47	3.28
4	73.95	2.25
5	44.61	3.96
6	58.85	3.22

TABLE 4: Average accuracy and standard deviation of each condition.

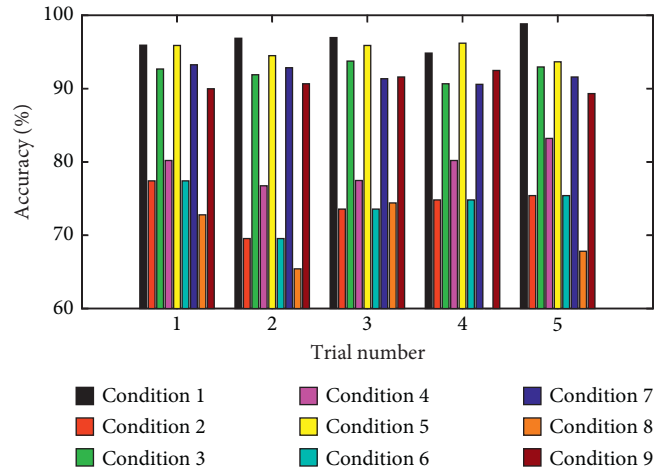| Condition	k	Inputs	Average accuracy (%)	Standard deviation
1	0	Collected signals	96.69	1.47
2	0.2	Noisy signals	82.46	2.45
3		Denoised signals	92.39	1.17
4	0.4	Noisy signals	79.57	2.57
5		Denoised signals	95.03	1.10
6	0.6	Noisy signals	74.15	2.93
7		Denoised signals	91.93	1.10
8	0.8	Noisy signals	67.80	6.31
9		Denoised signals	90.36	1.25

FIGURE 12: Diagnosis accuracy of the 5 trials.

TABLE 5: The framework of CNN for ablation experiment.

Parameter profile	Framework A	Framework B	Framework C	Framework D
Input	1024	1024	1024	1024
Convolution	32, 3, 1, same	32, 5, 1, same	32, 3, 2, same	48, 3, 1, same
Maxpooling	2	2	—	2
Convolution	64, 3, 1, same	64, 5, 1, same	64, 3, 2, same	96, 3, 1, same
Maxpooling	2	2	—	2
Convolution	128, 3, 1, same	128, 5, 1, same	128, 3, 2, same	192, 3, 1, same
Maxpooling	2	2	—	2
Fully connection	1024	1024	1024	1024
Fully connection	9	9	9	9
Keep_prob	0.9	0.9	0.9	0.9
Learning rate	0.1	0.1	0.1	0.1
Accuracy	**95.18% ± 1.64**	94.22% ± 1.84	93.16% ± 1.86	92.52% ± 1.98

experiments are listed in Table 5. CNNs with different frameworks and parameters are adopted to fault diagnose on the dataset to select the best CNN diagnostic model. From Table 5, it is clear that framework A is more robust for better fault diagnosis compared to other structures. Accordingly, an alternative ablation experiment is conducted on the key parameters of CNN with framework A as the basis.

5. Conclusion

In this paper, an adaptive diagnosis method that combines deep gated recurrent unit (DGRU) with wavelet packet decomposition (WPD) and extreme learning machine (ELM) is proposed for rolling bearing. Firstly, WPD is utilized to eliminate the noise of data. Secondly, DGRU is designed to extract the representative features of denoised data. Finally, ELM is utilized to output the diagnosis results. Massive experimental results prove that the superiority and robustness of the proposed method outperform existing popular methods. Additionally, the proposed method can also achieve powerful antinoise ability.

In conclusion, we will further improve our model to deal with the challenge of transfer experimental data knowledge to diagnose practical engineering equipment faults in future research [39].

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

All the authors equally contributed in the manuscript for "conceptualization, validation, formal analysis, investigation, and resources." All authors have read and agreed to the published version of the manuscript.

References

[1] J. Li, Z. Meng, P. Zuozhou, C. Lixiao, and F. Fengjie, "Multi-source feature extraction of rolling bearing compression measurement signal based on independent component analysis," *Measurement*, vol. 172, Article ID 108908, 2021.

[2] H. G. Yang, H. B. Lin, and K. Ding, "Sliding window denoising K-Singular Value Decomposition and its application on rolling bearing impact fault diagnosis," *Sound and Vibration*, vol. 421, pp. 205–219, 2018.

[3] K. Zhao, H. K. Jiang, C. Q. Liu, Y. F. Wang, and K. Zhu, "A new data generation approach with modified Wasserstein auto-encoder for rotating machinery fault diagnosis with limited fault data," *Knowledge-Based Systems*, vol. 238, Article ID 107892, 2022.

[4] W. H. Li, S. H. Zhang, and S. Rakheja, "Feature denoising and nearest–farthest distance preserving projection for machine fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 1, pp. 393–404, 2016.

[5] S. B. Wang, I. Selesnick, G. G. Cai, Y. N. Feng, X. Sui, and X. F. Chen, "Nonconvex sparse regularization and convex optimization for bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 9, pp. 7332–7342, 2018.

[6] Z. Y. Wang, C. Lu, and B. Zhou, "Fault diagnosis for rotary machinery with selective ensemble neural networks," *Mechanical Systems and Signal Processing*, vol. 113, pp. 112–130, 2018.

[7] J. Y. Jiao, M. Zhao, J. Lin, and J. Zhao, "A multivariate encoder information based convolutional neural network for intelligent fault diagnosis of planetary gearboxes," *Knowledge-Based Systems*, vol. 160, pp. 237–250, 2018.

[8] Z. H. An, S. M. Li, J. R. Wang, and X. X. Jiang, "A Novel Bearing Intelligent Fault Diagnosis Framework under Time-Varying Working Conditions Using Recurrent Neural Network," *ISA Transactions*, vol. 100, pp. 155–170, 2019.

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

[10] K. Zhao, H. K. Jiang, K. B. Wang, and Z. Y. Pei, "Joint distribution adaptation network with adversarial learning for rolling bearing fault diagnosis," *Knowledge-Based Systems*, vol. 222, Article ID 106974, 2021.

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

[12] M. D. Sun, H. Wang, P. Liu, S. D. Huang, and P. Fan, "A sparse stacked denoising autoencoder with optimized transfer learning applied to the fault diagnosis of rolling bearings," *Measurement*, vol. 146, pp. 305–314, 2019.

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

[14] Z. H. Wu, H. K. Jiang, K. Zhao, and X. Q. Li, "An adaptive deep transfer learning method for bearing fault diagnosis," *Measurement*, vol. 151, Article ID 107227, 2020.

[15] X. A. Yan and M. P. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, 2018.

[16] J. X. Ma, F. Y. Xu, K. Huang, and R. Huang, "GNAR-GARCH model and its application in feature extraction for rolling bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 93, pp. 175–203, 2017.

[17] K. Zhao, H. K. Jiang, Z. H. Wu, and T. F. Lu, "A novel transfer learning fault diagnosis method based on Manifold Embedded Distribution Alignment with a little labeled data," *Journal of Intelligent Manufacturing*, vol. 33, no. 1, pp. 151–165, 2020.

[18] Y. Zhu, X. D. Wu, P. P. Li, Y. H. Zhang, and X. G. Hu, "Transfer learning with deep manifold regularized auto-encoders," *Neurocomputing*, vol. 369, pp. 145–154, 2019.

[19] Z. F. Kang, B. Yang, Z. S. Li, and P. Wang, "OTLAMC: an online transfer learning algorithm for multi-class classification," *Knowledge-Based Systems*, vol. 176, pp. 133–146, July 2019.

[20] X. Li, X. D. Jia, W. Zhang, H. Ma, and X. Li, "Intelligent Cross-Machine Fault Diagnosis Approach with Deep Auto-Encoder and Domain Adaptation," *Neurocomputing*, vol. 383, pp. 235–247, 2019.

[21] L. Wen, L. Gao, and X. Y. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 1, pp. 136–144, 2019.

[22] L. Guo, Y. G. Lei, S. B. Xing, T. Yan, and N. P. Li, "Deep convolutional transfer learning network: a new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 9, pp. 7316–7325, 2018.

[23] D. Y. Xiao, Y. X. Huang, L. J. Zhao, C. Q. Qin, H. T. Shi, and C. L. Liu, "Domain adaptive motor fault diagnosis using deep transfer learning," *IEEE Access*, vol. 7, Article ID 80937, 2019.

[24] X. X. Wang, H. B. He, and L. S. Li, "Hierarchical deep domain adaptation approach for fault diagnosis of power plant thermal system," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 9, pp. 5139–5148, 2019.

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

[26] T. Han, C. Liu, W. G. Yang, and D. X. Jiang, "Learning transferable features in deep convolutional neural networks for diagnosing unseen machine conditions," *ISA Transactions*, vol. 93, pp. 341–353, October 2019.

[27] J. D. Wang, Y. Q. Chen, S. J. Hao, W. J. Feng, and Z. Q. Shen, "Balanced distribution adaptation for transfer learning," in *Proceedings of the 2017 IEEE International Conference on Data Mining*, pp. 1129–1134, New Orleans, LA, USA, November 2017.

[28] B. D. Shai, B. John, C. Koby, and P. Fernando, "Analysis of representations for domain adaptation," *Advances in Neural Information Processing Systems 19*, in *Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems*, Vancouver, Canada, December 2006.

[29] R. Zhao, D. Z. Wang, R. Q. Yan, K. Z. Mao, F. Shen, and J. J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1539–1548, 2018.

[30] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada, "Maximum classifier discrepancy for unsupervised domain adaptation," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3723–3732, Salt Lake City, AL, USA, June 2018.

[31] H. D. Shao, J. S. Cheng, H. K. Jiang, Y. Yang, and Z. T. Wu, "Enhanced Deep Gated Recurrent Unit and Complex Wavelet Packet Energy Moment Entropy for Early Fault Prognosis of Bearing," *Knowledge.-Based System*, vol. 188, Article ID 105022, 2019.

[32] H. K. Jiang, X. Q. Li, H. Shao, and K. Zhao, "Intelligent fault diagnosis of rolling bearings using an improved deep recurrent neural network," *Measurement. Science. Technology.* vol. 29, no. 6, Article ID 65107, 2018.

[33] K. Zhao, H. K. Jiang, X. Q. Li, and R. X. Wang, "An optimal deep sparse autoencoder with gated recurrent unit for rolling bearing fault diagnosis," *Measurement. Science. Technology.* vol. 31, no. 1, Article ID 15005, 2019.

[34] K. A. Loparo, "Case Western Reserve University Bearing Data Center Website," 2012, http://csegroups.case.edu/bearingdatacenter/home.

[35] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.

[36] M. S. Long, J. M. Wang, G. G. Ding, J. G. Sun, and P. S. Yu, "Transfer Feature Learning with Joint Distribution Adaptation," in *Proceedings of the 2013 IEEE International Conference on Computer Vision*, pp. 2200–2207, Sydney, Australia, December 2013.

[37] Ke Zhao, H. Jiang, and Z. Wu, "Transfer residual convolutional neural network for rotating machine fault diagnosis under different working conditions," in *Proceedings of the 2021 12th International Conference on Mechanical and Aerospace Engineering (ICMAE)*, pp. 477–483, IEEE, Athens, Greece, July 2021.

[38] Y. Ganin, E. Ustinova, H. Ajakan et al., "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016.

[39] W. Zhang, J. Chen, S. Riaz, N. Zheng, and L. Li, "Research on distributed cooperative control strategy of microgrid hybrid energy storage based on adaptive event triggering," *CMES-Computer Modeling in Engineering & Sciences*, vol. 132, no. 2, pp. 585–604, 2022.