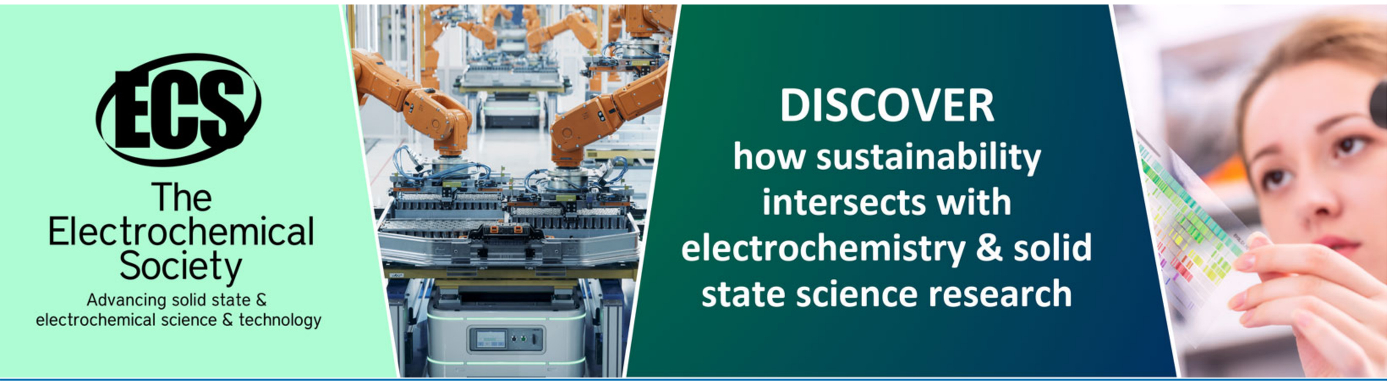
Measurement Science and Technology

100 year anniversary
Measurement Science and Technology

PAPER

A self-adaptive DRSN-GPReLU for bearing fault diagnosis under variable working conditions

To cite this article: Zhijin Zhang *et al* 2022 *Meas. Sci. Technol.* **33** 124005



IOP Publishing

Measurement Science and Technology

Meas. Sci. Technol. **33** (2022) 124005 (16pp)

https://doi.org/10.1088/1361-6501/ac86e3

A self-adaptive DRSN-GPReLU for bearing fault diagnosis under variable working conditions

Zhijin Zhang[1], Chunlei Zhang[1], Xin Zhang[1], Lei Chen[2], Huaitao Shi[3] and He Li[1,*]

[1] School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, People's Republic of China
[2] Midea Group, Foshan 528311, People's Republic of China
[3] School of Mechanical Engineering, Shenyang Jianzhu University, Shenyang 110168, People's Republic of China

E-mail: hli@mail.neu.edu.cn

Received 27 April 2022, revised 12 July 2022
Accepted for publication 4 August 2022
Published 16 September 2022

Abstract

Recently, deep learning has been widely used for intelligent fault diagnosis of rolling bearings due to its no-mankind feature extraction capability. The majority of intelligent diagnosis methods are based on the assumption that the data collected is from constant working conditions. However, rolling bearings often operate under variable working conditions in the real diagnosis scenario, which reduces the generalization capability of the diagnosis model. To solve this problem, a self-adaptive deep residual shrinkage network with a global parametric rectifier linear unit (DRSN-GPReLU) is proposed in this paper. First, the DRSN is used as the basic architecture to improve the anti-noise ability of the proposed method. Then, a novel activation function—the GPReLU—is developed, which can achieve better intra-class compactness for vibration signals, and the inter-class samples are better mapped into remote areas. Finally, a sub-network based on the attention mechanism is designed to automatically infer the slope of the GPReLU. Various experimental results demonstrate that the DRSN-GPReLU can realize better performance compared with traditional methods under variable working conditions, and has better robustness under noise interference.

Keywords: deep residual shrinkage networks, rectifier linear units, attention mechanism, rolling bearing fault diagnosis, variable working conditions

(Some figures may appear in colour only in the online journal)

1. Introduction

Rolling bearings are widely used in aerospace, transportation, manufacturing, and other fields [1]. However, because rolling bearings are frequently utilized in complex and harsh environments, failures are unavoidable and will result in equipment downtime and significant financial losses. Therefore, it is essential to accurately diagnose the faults of rolling bearings [2]. However, the majority of diagnosis methods are based on the assumption that the data collected is from constant working conditions [3], which makes it difficult to adapt to actual application scenarios. Rolling bearings often operate in complicated conditions with varying speeds and loads, which presents a challenge to traditional fault diagnosis methods. Consequently, it is significant to develop a novel method which can achieve accurate fault diagnosis under variable working conditions.

The method based on signal processing is an effective method in the field of fault diagnosis [4–6]. For example,

* Author to whom any correspondence should be addressed.

Xu *et al* [5] proposed a novel empirical scanning spectrum kurtosis for fault diagnosis, which can accurately identify bearing faults. Moreover, researchers have also tried to propose various diagnosis methods based on signal processing to address the challenges of variable working conditions [7, 8]. For example, Chen *et al* [8] proposed a novel approach based on polynomial chirplet transform under variable speed conditions. Generally, signal processing-based fault diagnosis methods usually require technicians to locate fault frequencies based on expert experience. However, this is often difficult due to the impact of complex real-world scenarios. The data-driven fault diagnosis approach provides another effective tool that does not require locating the frequency of faults [9–11]. For example, Kaya *et al* [9] proposed a new 1D-Local binary pattern (LBP)-based feature extraction method, which can be combined with various machine learning methods to accomplish fault diagnosis of bearings. However, the data-driven fault diagnosis method relies on the quality of the fault extraction feature set, i.e. a good extraction feature set can represent the fault feature information more comprehensively. Therefore, how to better extract fault features has been a challenge in the field of data-driven fault diagnosis.

As an alternative, deep learning technology has become mainstream in the field of fault diagnosis, which has nomankind feature extraction capability [12–14]. Among them, the convolutional neural network (CNN) and its variants have been widely applied in various classification tasks due to their excellent pattern recognition performance. The Resnet is a variant of the CNN [15], and its idea of residual error can effectively avoid the phenomenon of gradient vanishing, which is beneficial to the training of the network. Therefore, the Resnet has gained popularity in the field of fault diagnosis [16–18]. For instance, Zhang *et al* [16] introduced the residual learning idea of the Resnet and constructed a new deep learning structure which can achieve good fault diagnosis accuracy for rolling bearings. However, in real-world applications with strong background noise, the diagnosis ability of the above method will inevitably be reduced. Therefore, Zhao *et al* [19] proposed a variant of the Resnet, called the deep residual shrinkage network (DRSN). The DRSN employs a fundamental Resnet structure and perfectly combines the attention mechanism with the signal processing knowledge of the wavelet denoising. The DRSN has become one of the best deep learning architectures in the field of fault diagnosis [20–23]. For instance, Liu *et al* [21] combined a transfer learning method with the DRSN under harsh interference environments. Zhang *et al* [22] modified the shrinkage function of the DRSN to significantly improve the fault diagnosis accuracy of the original DRSN under strong background noise. The above research shows that the DRSN demonstrates better performance than the CNN and Resnet, and therefore this paper uses the DRSN as the basis architecture for further improvement.

Meanwhile, variable working conditions will lead to the degradation of the fault diagnosis ability of traditional deep learning models; therefore, several academics have tried to suggest some methods to solve this problem [24–26]. For instance, Han *et al* [24] improved the effect of the domain alignment of the transfer learning, to ensure accurate diagnosis under the influence of variable working conditions. Zhao *et al* [26] combined the CNN with the batch normalization (BN) and exponential moving average technology, which can eliminate feature distribution differences under variable working conditions. However, the aforementioned studies all fail to take into account a crucial issue, whereby the development of a correct mapping relationship between the class and the original input data depends directly on the choice of activation functions. The fixed activation functions used by all traditional deep learning methods result in identical nonlinear transformation for each vibration signal. These fixed activation functions include the sigmoid, rectifier linear units (ReLUs) [27], and the related variants of ReLUs [28, 29]. However, the fault diagnosis of rolling bearing variables is complicated by the variable working conditions. To be specific, intra-class faults may result in differences in the pulses and waveforms of vibration signals due to different working conditions. Similarly, the fault characteristics of inter-class faults may be the same due to different working conditions. Therefore, it is difficult for traditional activation functions to establish the correct correspondence between the original input data and its corresponding categories, which can easily lead to misclassification. Therefore, to enhance the diagnostic ability of the model under variable working conditions, numerous academics improved the traditional activation function [30–32]. Shao *et al* [30] applied the wavelet function in the auto-encoder to perform nonlinear transformations on vibration signals, which replaces the activation function and accurately identifies bearing conditions. Zhao *et al* [32] developed an adaptive parametric ReLU (APReLU), which can achieve an adaptive nonlinear transformation for each vibration signal; the APReLU can also achieve better diagnosis performance under variable working conditions compared with traditional activation functions. However, the APReLU in [32] only takes into account the nonlinear transformation in the negative region of the feature space, and obviously ignores the nonlinear transformation in the positive region, so that the features are insufficiently extracted.

The research mentioned above shows how important it is to develop a novel activatation function for the deep-learning-based fault diagnosis method. To this end, a novel activation function called the global parametric ReLU (GPReLU) is developed, which takes into account the nonlinear transformation of the global characteristics, including the nonlinear transformation in the negative and the positive regions. Next, a brand-new deep network architecture called the DRSN-GPReLU is proposed, which can achieve adaptive nonlinear transformation for each vibration signal, allowing better projection of intra-class samples into the same area, and better projection of inter-class samples into the distant area. The contributions of this article are as follows:

(a) Firstly, a self-adaptive DRSN-GPReLU structure based on the DRSN is designed. The proposed model is more robust under noise conditions.
(b) Next, a GPReLU is developed to improve the mapping of intra-class vibration signals into the same area and the

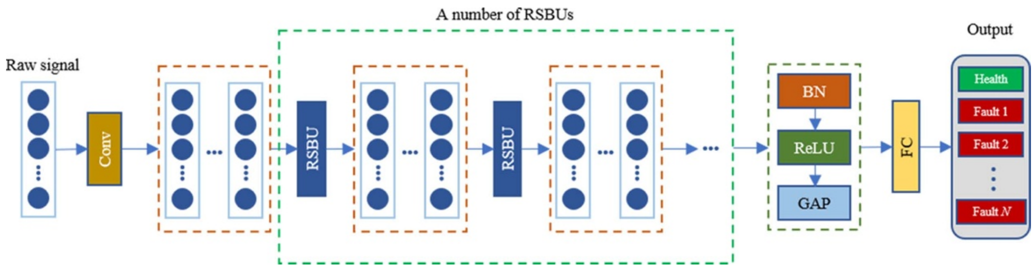


Figure 1. The DRSN's structure.

mapping of inter-class vibration signals into the faraway area.

(c) Finally, to determine the slope of the GPReLU automatically, a novel sub-network based on an attention mechanism is designed.

The rest of this paper is arranged as follows. In section 2, the theory of the self-adaptive DRSN-GPReLU model is introduced. Section 3 details the experimental results and related comparative analysis. Section 4 gives the conclusions.

2. Theory of the self-adaptive DRSN-GPReLU

2.1. Fundamentals of the DRSN

The DRSN is a special variant of the ResNet, which uses a basic Resnet structure, and perfectly integrates the attention mechanism and the signal processing knowledge of the wavelet denoising. It can be seen from figure 1 that the DRSN consists of an original vibration signal input layer, a convolutional layer, some residual shrinkage building units (RSBUs) (RSBUs are the most important core components), a canonical processing module (BN), a nonlinear transformation layer (ReLU), anda global average pooling (GAP) layer. More details about the DRSN can be found in [19].

As depicted in figure 2(a), the RSBU includes two batch normalization layers, two nonlinear transformation layers, two convolutional layers, a threshold module, and an identity shortcut. The BN layer can normalize features and reduce internal covariant shift. The threshold module adopts the soft thresholding function, which can remove the noise, and the threshold is automatically inferred based on an attention mechanism in the process of training, which is realized by a sub-network of the inferred threshold. The components of this sub-network are depicted in figure 2(b), which mainly include the absolute module, the ReLU, the sigmoid function, and so on. The first two layers can guarantee that the threshold is a positive value, and the sigmoid function layer can guarantee that the threshold cannot be too large, so that the inferred thresholds meet the threshold rule.

2.2. The developed GPReLU activation function

As shown in figure 3(a), ReLUs can more effectively prevent the gradient vanishing, which is conducive to neural network training compared to sigmoid and tanh. An ReLU uses the function:

$$y = \max(x, 0) \tag{1}$$

where x and y are the input features and output features, respectively.

To enhance the performance of ReLUs in the deep learning architecture, excellent variants such as Leaky ReLU (LReLUs) and Parametric ReLU (PReLUs) have been derived. As shown in figure 3(b), LReLUs have a non-zero slope value (e.g. 0.3) in the negative region compared with ReLUs. An LReLU uses the function:

$$y = 0.3 \cdot \min(x, 0) + \max(x, 0). \tag{2}$$

Similarly, PReLUs are improved on the basis of LReLUs. As shown in figure 3(c), PReLUs ensure that the slope value in the negative region is not a fixed value, but is automatically inferred by the gradient descent method in neural network training. A PReLU uses the function:

$$y = \alpha \cdot \min(x, 0) + \max(x, 0) \tag{3}$$

where α is the slope value of the negative region.

However, these variants that include ReLUs have two limitations when applied in the fault diagnosis of rolling bearings. Firstly, these activation functions perform the same nonlinear transformation on each sample, which leads to poor diagnostic results under variable working conditions. Secondly, these activation functions only take into account the nonlinear transformation in the negative region of the feature space, and obviously ignore the nonlinear transformation in the positive region, so that the features are insufficiently extracted.

Therefore, to address the above issues, a new activation function called the GPReLUs is developed, which takes into account the nonlinear transformation of the global characteristics, including the nonlinear transformation in the negative and positive regions. A GPReLU uses the function:

$$y = \alpha \cdot \min(x, 0) + \beta \cdot \max(x, 0) \tag{4}$$

where α and β are the coefficients of the positive and negative regions, respectively.

In addition, a novel slope parameter inferred adaptive sub-network that is based on the attention mechanism is designed. The coefficient values of the GPReLU can be adaptively adjusted by the sub-network in response to different vibration signal

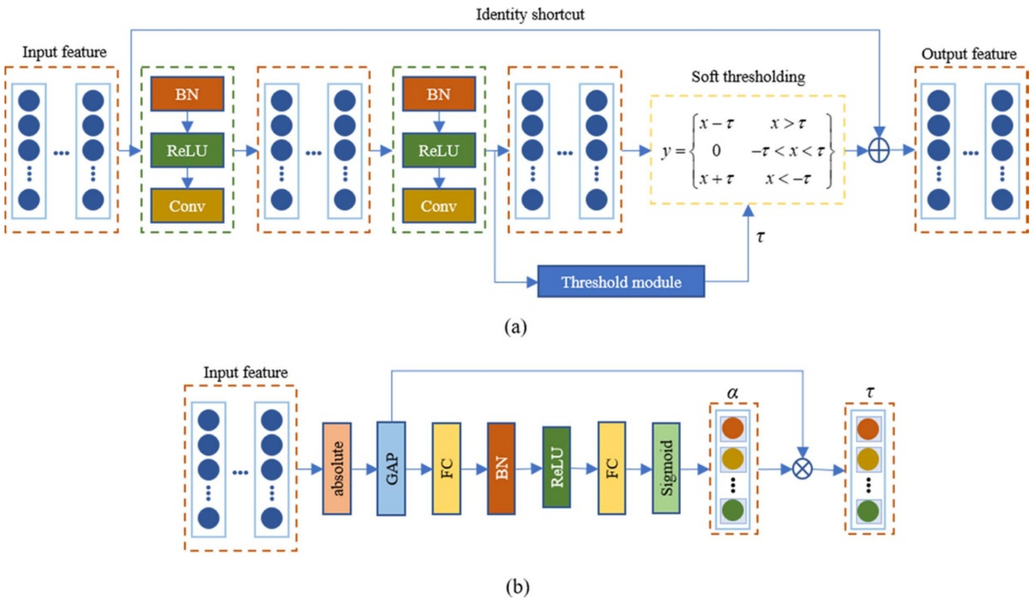


Figure 2. (a) The RSBU's structure, and (b) the threshold module.

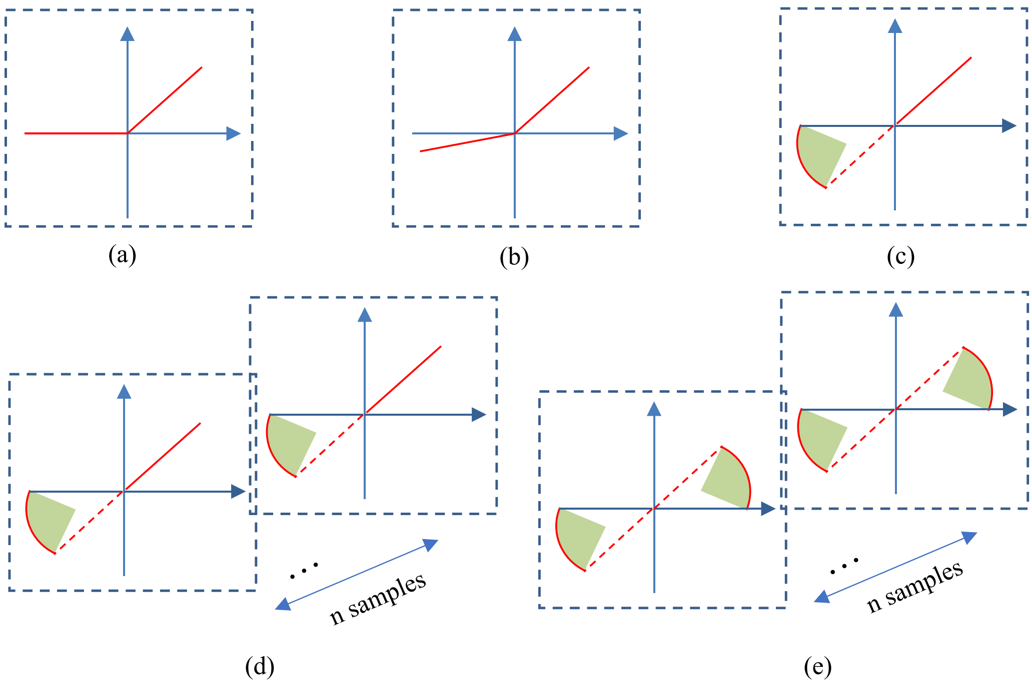


Figure 3. (a) ReLUs, (b) LReLUs, (c) PReLUs, (d) attention PReLUs, and (e) attention GPReLUs.

samples, so that the self-adaptive nonlinear transformation is performed on different samples. As shown in figure 3(d), attention PReLUs in [32] are only limited to the dynamic adaptive nonlinear transformation in the negative region and ignore the positive region. As shown in figure 3(e), the attention GPReLUs proposed in this paper consider the adaptive nonlinear transformation of global features, compared with attention PReLUs. Therefore, attention GPReLUs are better equipped to adapt to variable working conditions and extract features more thoroughly. Additionally, more information about this sub-network will be included in the following.

2.3. The sub-network of inferred slopes

As indicated in figure 4, a novel sub-network architecture is designed, which can automatically infer the parameters α and β, respectively. The automatic derivation process of the two parameters is consistent. Firstly, the feature information of the negative and positive regions can be extracted by the min(x,0) module and max(x,0) module, respectively. Then, the extracted feature information is transformed into one-dimensional vectors through the GAP layer, which can improve the training speed and solve the problem of shift variation. Then, the

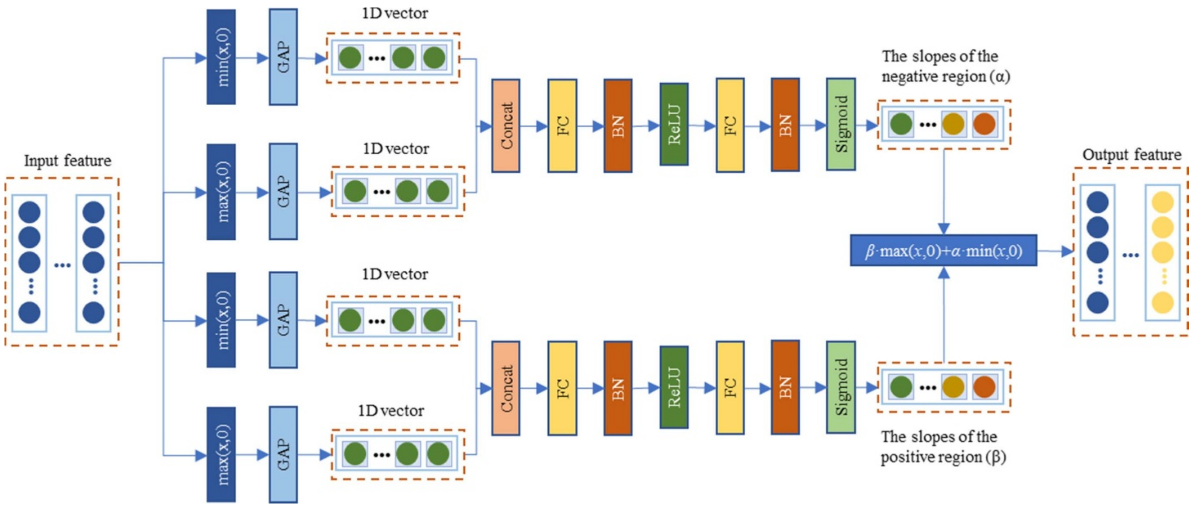


Figure 4. The network design of the inferred slopes.

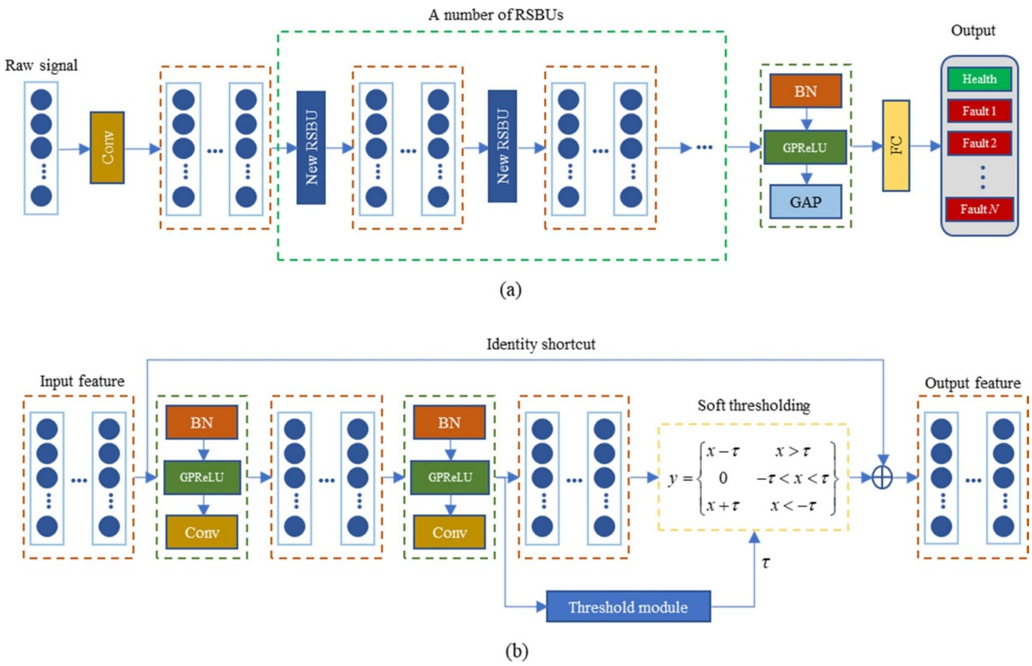


Figure 5. (a) The architecture of the self-adaptive DRSN-GPReLU, and (b) the composition of the new RSBU.

global information is fused by the Concat layer. Next, feature normalization is completed by the BN module. Finally, the value range of parameters α and β is limited to (0, 1) by the sigmoid module. Overall, the sub-network not only applies the self-adaptive nonlinear transformation to different vibration signal samples, but also considers the global signal features, so that the self-adaptive GPReLU effectively overcomes the limitations of the traditional activation function.

2.4. The architecture of self-adaptive DRSN-GPReLU

In this section, the overall structure of the self-adaptive DRSN-GPReLU is given in figure 5(a). The network structure of the self-adaptive DRSN-GPReLU is similar to the DRSN, and there are two main differences. Firstly, the last ReLU activation function in the overall architecture is replaced by the attention GPReLU. Secondly, the ReLU in the original RSBU is replaced with attention GPReLU to generate a new RSBU, as shown in figure 5(b), so that the new RSBU has self-adaptive nonlinear transformation that takes into account global features.

3. Experimental validation and results

In this section, the self-adaptive DRSN-GPReLU proposed in this paper is used for bearing fault diagnosis under variable working conditions, and is compared with traditional activation functions.

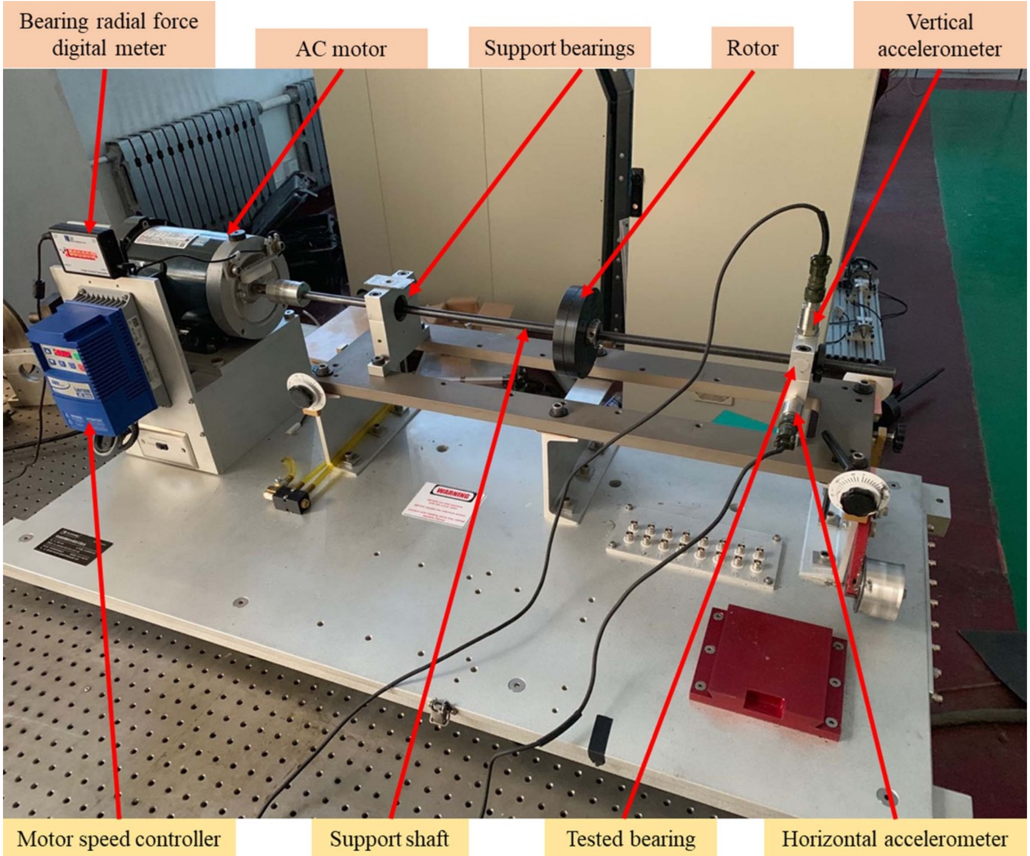


Figure 6. The rolling bearing comprehensive fault test stand.

3.1. Case 1: rolling bearing comprehensive fault test stand

3.1.1. Dataset collection. The comprehensive fault test stand of a rolling bearing was used to collect vibration signals in our experiments. As indicated in figure 6, the test bench includes a bearing radial force digital meter, an alternating current (AC) motor, a support bearing, a rotor, a motor speed controller, a support shaft, two accelerometers, and the tested bearing. The speed is provided by an AC motor. The test bench can simulate various types of failures of rolling bearings, and adjusts the speed through a motor speed controller to simulate rolling bearing failures under different working conditions. In addition, the vibration signals collected in this experiment are collected by two accelerometers, including a vertical accelerometer and a horizontal accelerometer.

As presented in table 1, four conditions of rolling bearings are simulated in this experiment, including a normal state and three fault states. The specific rolling bearing fault settings are shown in figure 7. In addition, four different speeds are set in each bearing condition, including 20, 30, 40, and 50 Hz. The sampling frequency of the vibration signals is 1000 Hz. In this experiment, 50 signal samples are collected for each condition, and each sample contains 1024 vibration points. Therefore, the total number of vibration signal samples is 800.

Table 1. Experimental data of rolling bearings.

Number	Description	Label
1	Normal (N) under a speed of 20, 30, 40, and 50 Hz	N
2	Inner ring fault (IF) under a speed of 20, 30, 40, and 50 Hz	IF1
3	Outer ring fault (OF) under a speed of 20, 30, 40, and 50 Hz	OF1
4	Ball fault (BF) under a speed of 20, 30, 40, and 50 Hz	BF1

3.1.2. Hyperparameter setup for self-adaptive GPReLU-DRSN. In table 2, all the hyperparameters used in this experiment are listed. It is worth mentioning that this article mainly proposes a new method rather than the optimization of parameters. To obtain as good a diagnostic accuracy as possible, the hyperparameter of this article refers to the parameter setting in [19].

3.1.3. Performance comparison. In the experiment, as presented in table 3, the diagnostic performance of the GPReLU is significantly better than other traditional activation functions under variable working conditions, which relies

Figure 7. A schematic diagram of rolling bearing failure.

Table 2. The hyperparameters of the experiment.

Number	Parameters	Value
1	Learning rate	0.001
2	Momentum ratio	0.9
3	The coefficient of L2 regularization	0.0001
4	Mini-batch	16
5	The size of convolutional kernels	(3,3)
6	The number of convolutional kernels in the first RSBU	4
7	The number of convolutional kernels in the second RSBU	4
8	The number of convolutional kernels in the third RSBU	8
9	The number of convolutional kernels in the fourth RSBU	8
10	The number of convolutional kernels in the fifth RSBU	16
11	The number of convolutional kernels in the sixth RSBU	16

Table 3. Diagnostic accuracy results of different methods.

SNR (dB)	−4	−2	0	2	4	6
DRSN-Sigmoid (%)	84.0 ± 2.7	86.1 ± 2.5	87.1 ± 1.7	90.1 ± 2.4	92.0 ± 1.3	94.6 ± 1.5
DRSN-ReLU (%)	85.9 ± 0.8	86.6 ± 1.2	87.9 ± 1.4	89.9 ± 2.4	93.4 ± 2.8	95.9 ± 2.9
DRSN-LReLU (%)	86.7 ± 1.3	87.1 ± 2.4	89.1 ± 1.7	91.3 ± 2.8	93.5 ± 1.0	95.6 ± 1.1
DRSN-PReLU (%)	84.2 ± 1.6	87.8 ± 3.5	90.1 ± 2.0	92.1 ± 1.9	93.0 ± 2.4	95.2 ± 1.3
DRSN-APReLU (%)	85.6 ± 1.8	89.7 ± 0.7	92.7 ± 1.6	93.2 ± 2.8	94.8 ± 2.3	97.1 ± 2.1
DRSN-GPReLU (%)	***95.1 ± 3.1***	***96.9 ± 2.0***	***97.9 ± 1.3***	***98.8 ± 1.0***	***99.0 ± 1.6***	***99.3 ± 0.5***

on its strong ability of the self-adaptive nonlinear transformation. To be specific, the fault diagnosis accuracy of the GPReLU yields improvements of 8.9%, 7.9%, 7.2%, 7.4%, and 5.6%, compared to the Sigmoid, the ReLU, the LReLU, the PReLU, and the APReLU under different noise intensities. Moreover, it can be found through experimental results that the DRSN-GPReLU still has a high diagnostic accuracy under a strong noise environment, which indicates that the DRSN-GPReLU demonstrates better anti-noise performance. It needs to be emphasized that the network structure and hyperparameter settings are unified, but the activation function is replaced, to validate the superior performance of the new activation function developed in this paper, compared with other traditional activation functions. In addition, each set of experiments was run 10 times, to ensure that the experimental results were not random.

Then, to more clearly identify the actual diagnosis results of each category, the confusion matrices of the different methods are indicated in figure 8. From the confusion matrix in figure 8, our proposed DRSN-GPReLU achieves the best diagnostic performance compared to other methods. The accuracy of the IF category is 95%, 78%, 47%, 57%, 50%, and 60%, which can further illustrate the superiority of our method.

The accuracies and loss curves of the DRSN-GPReLU and other methods are depicted in figure 9, respectively, when the SNR $= -4$ dB. It can be observed that the average accuracy of the DRSN-GPReLU is higher than the other methods, and the average loss of our proposed method is lower than the other methods. This is because the adaptive nonlinear transformation of the DRSN-GPReLU can improve diagnostic accuracy under variable working conditions. Furthermore, it is worth mentioning that the superiority of the DRSN-GPReLU

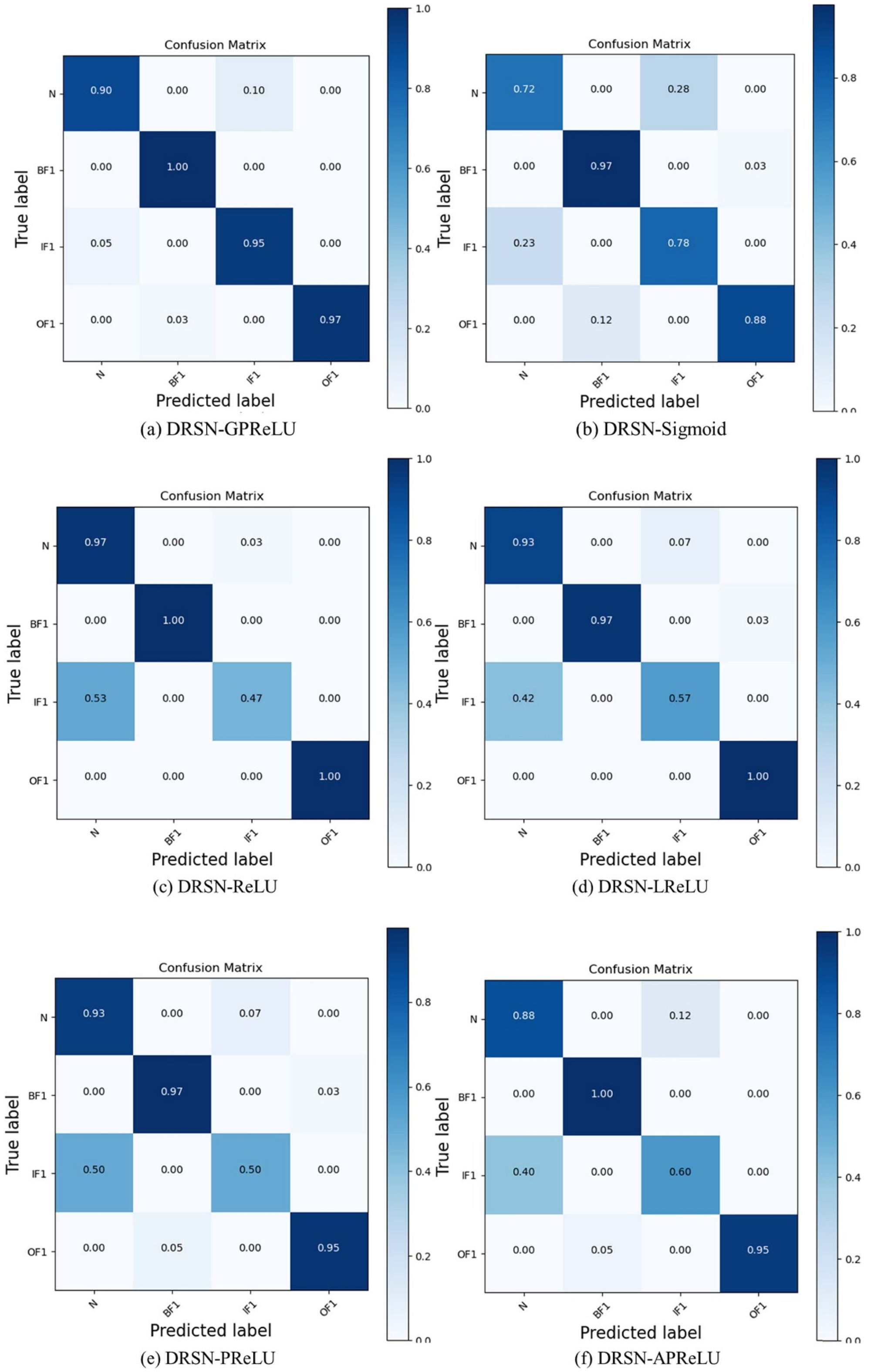
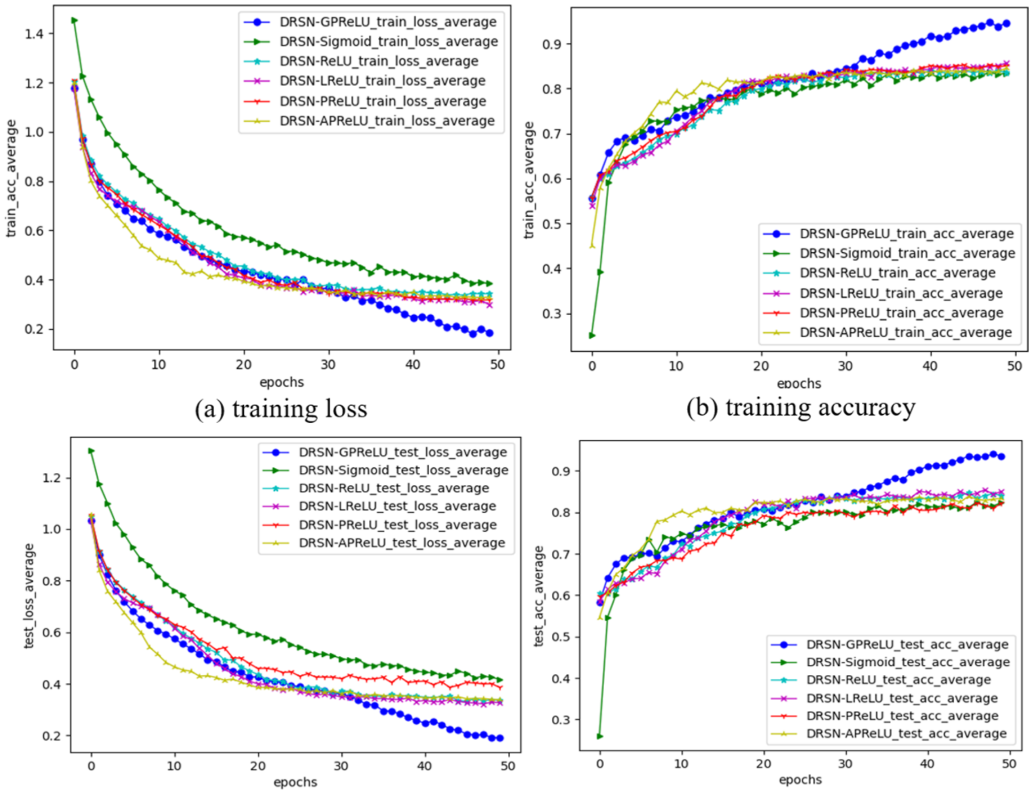


Figure 8. Confusion matrices of the (a) DRSN-GPReLU and (b)–(f) other methods.

starts to be significant, when the epochs are greater than about 30. This phenomenon shows that our adaptive method cannot oversaturate as easily as other functions after 30 epochs, but can also effectively adjust itself according to the signal.

To visually demonstrate the superiority of the developed GPReLU under variable working conditions from the perspective of spatial feature distribution, the t-distributed stochastic neighbor embedding (t-SNE) method [33] is used



(a) training loss (b) training accuracy

(c) test loss (d) test accuracy

Figure 9. The loss and accuracy curves: (a) training loss, (b) training accuracy, (c) test loss, and (d) test accuracy.

to visualize the high-dimensional features of the six methods. As shown in figure 10, all features are represented by coordinate points mapped to a two-dimensional space. As indicated in figure 10(a), it can be observed that the sample features become more separable in the DRSN-GPReLU under variable working conditions, when the SNR $= -4$ dB. In contrast, the N and IF classes highly overlapped in the other methods, from sample points in figures 10(b)–(f). The main reason for this may be that the vibration signal characteristics under normal conditions are similar to the vibration signal characteristics under the inner ring failure conditions due to the change in the speed in this experiment. Firstly, the traditional activation function can only rely on a fixed nonlinear transformation. Secondly, the nonlinear transformation in the positive region of the feature space is usually ignored, which makes it difficult to accurately diagnose faults.

3.1.4. An analysis of slope values. As shown in table 4, the slopes of various activation functions are examined using eight different vibration signal samples as examples. It can be seen that LReLU and PReLU conduct identical nonlinear transformation for each sample because their slope values are fixed. The main difference is that the slope value of LReLU is manually selected, while PReLU's slope value is automatically inferred during training. In contrast, the APReLU and the developed GPReLU have different slope values for each sample, which means that they can perform different nonlinear transformation for each vibration signal, and these slope values are all obtained based on an attention mechanism. The difference between them is that the GPReLU proposed in this paper has two inferred slopes at the same time and considers the global feature information, while the APReLU only considers the negative region and ignores the positive region.

3.2. Case 2: the public bearing fault dataset

3.2.1. Dataset description. In this section, the bearings dataset of Case Western Reserve University (CWRU) is adopted. The CRWU's test bench is indicated in figure 11. The vibration signals of a driver end bearing are used in this paper, which are collected at 1797, 1772, 1750, and 1730 RPM, respectively. The sampling frequency is 12 000 Hz. More details about the experimental data are shown in table 5. Each type of fault is also set with three fault sizes, including 7, 14, and 21 mils. In this experiment, 100 samples for each class of states are used, and each sample contains 1024 vibration signals. The proportion of the training set is 80%.

3.2.2. Performance comparison. In the experiment, the architecture and hyperparameters are the same as in case 1. Due to the larger number of training data, the batch size in this experiment is 32, and the samples are subjected to

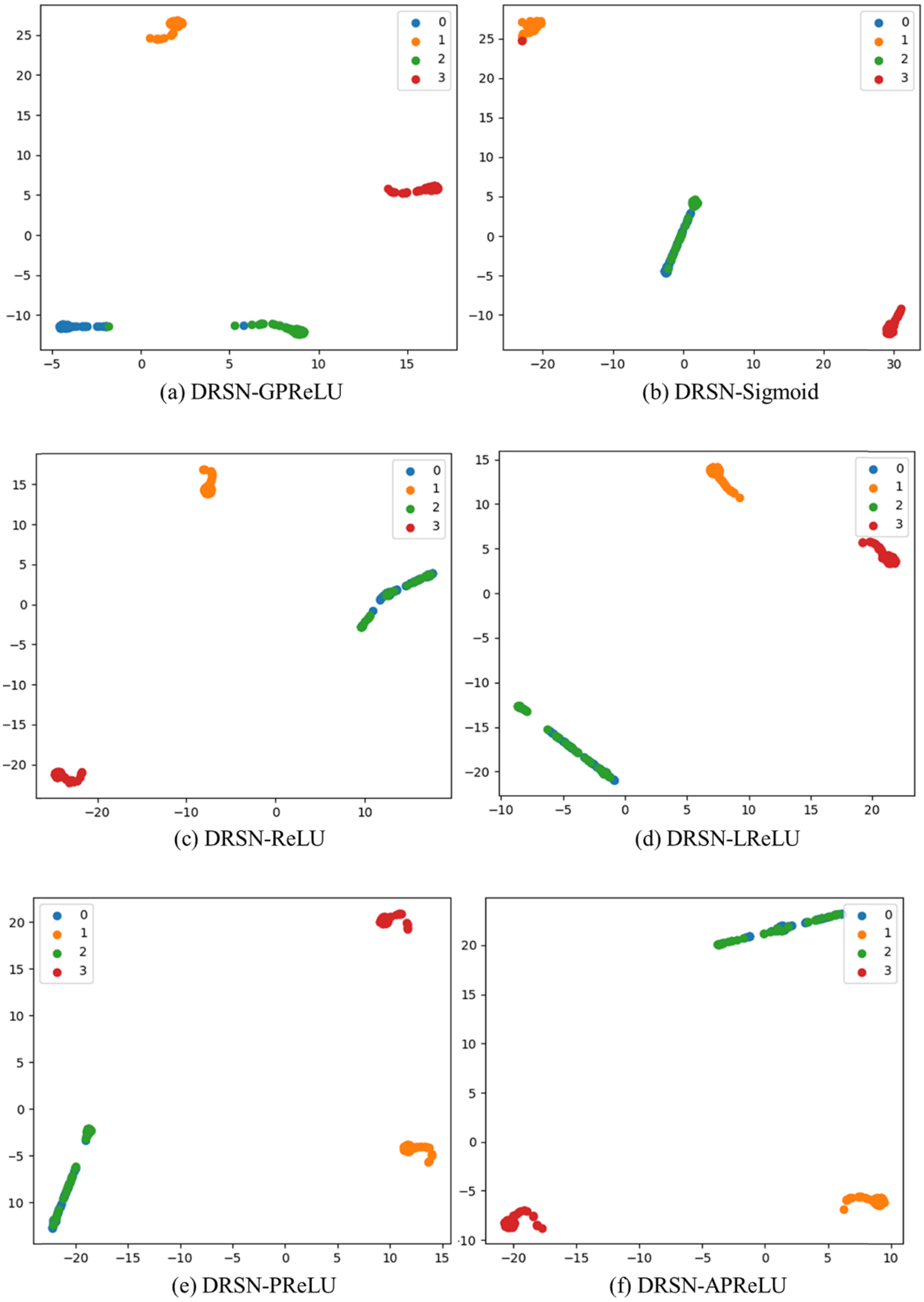


Figure 10. Visualization of feature spaces of the (a) DRSN-GPReLU and (b)–(f) other methods.

Table 4. A summary of the slope value analysis.

Sample	LReLU	PReLU	APReLU	GPReLU (α)	GPReLU (β)
1	0.3	0.546	0.166	0.705	0.169
2	0.3	0.546	0.397	0.640	0.232
3	0.3	0.546	0.685	0.272	0.767
4	0.3	0.546	0.333	0.628	0.236
5	0.3	0.546	0.610	0.354	0.662
6	0.3	0.546	0.685	0.272	0.767
7	0.3	0.546	0.503	0.476	0.489
8	0.3	0.546	0.685	0.272	0.767

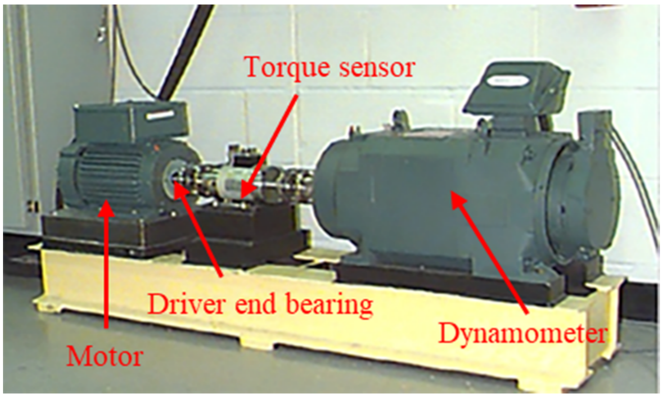


Figure 11. The CRWU's test bench.

Table 5. A detailed description of the dataset.

Rotating speed (r $\min^{-1}$)	Healthy condition	Fault size (inch)	Class label	Number of samples
1797 & 1772 & 1750 & 1730	N	0	0	100 & 100 & 100 & 100
1797 & 1772 & 1750 & 1730	BF1	0.007	1	100 & 100 & 100 & 100
1797 & 1772 & 1750 & 1730	BF2	0.014	2	100 & 100 & 100 & 100
1797 & 1772 & 1750 & 1730	BF3	0.021	3	100 & 100 & 100 & 100
1797 & 1772 & 1750 & 1730	IF1	0.007	4	100 & 100 & 100 & 100
1797 & 1772 & 1750 & 1730	IF2	0.014	5	100 & 100 & 100 & 100
1797 & 1772 & 1750 & 1730	IF3	0.021	6	100 & 100 & 100 & 100
1797 & 1772 & 1750 & 1730	OF1	0.007	7	100 & 100 & 100 & 100
1797 & 1772 & 1750 & 1730	OF2	0.014	8	100 & 100 & 100 & 100
1797 & 1772 & 1750 & 1730	OF3	0.021	9	100 & 100 & 100 & 100

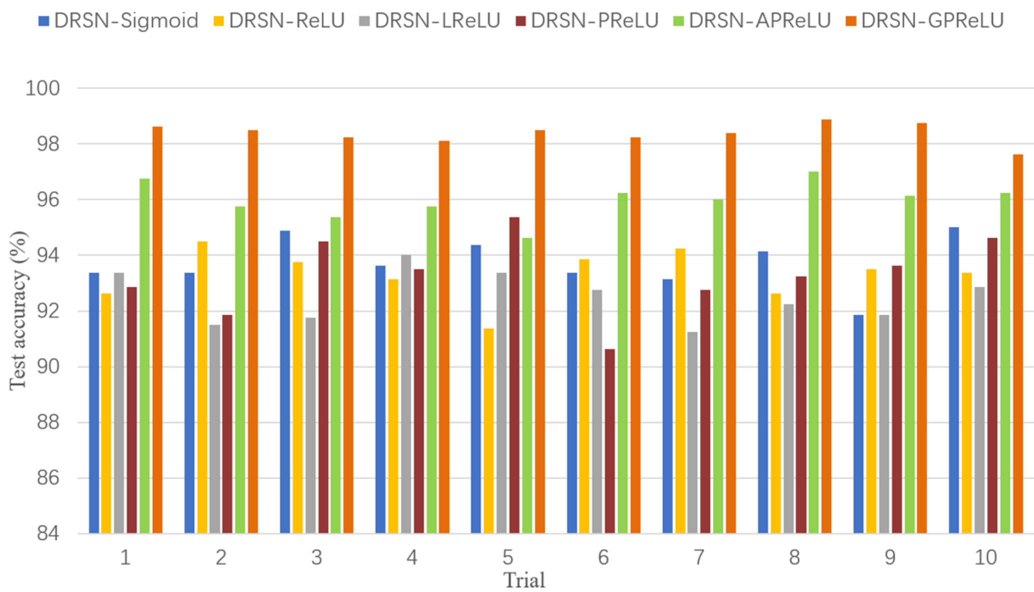
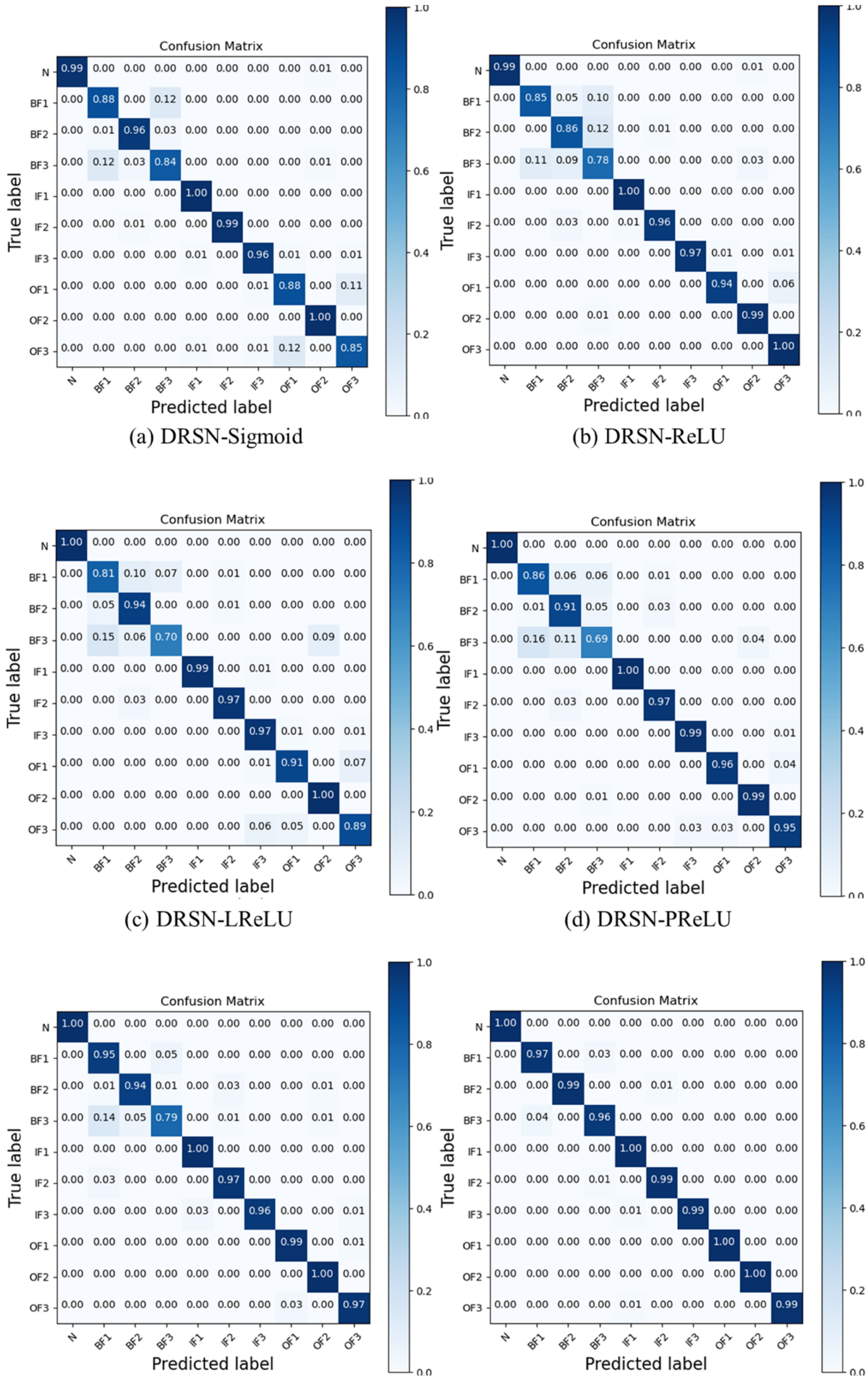


Figure 12. Test accuracy of different methods in ten trials.

white Gaussian noise, the intensity of which is 6 dB. Each set of experiments is performed ten times to guarantee that the results are not random. As indicated in figure 12, the diagnosis performance of the GPReLU is far more superior than other traditional activation functions under variable working conditions, and the average accuracy of fault diagnosis of the DRSN-GPReLU reaches more than 98%. In addition, it is worth mentioning that the DRSN-APReLU has superior fault diagnosis accuracy compared to other methods, with the exception of our method. This further demonstrates that self-adaptive nonlinear transformation can better adapt to the effects of variable working conditions.



(a) DRSN-Sigmoid (b) DRSN-ReLU

(c) DRSN-LReLU (d) DRSN-PReLU

(e) DRSN-APReLU (f) DRSN-GPReLU

Figure 13. Confusion matrices of (a)–(e) other methods and (f) the DRSN-GPReLU.

Next, to more clearly identify the actual diagnosis results of each category, the confusion matrix of the DRSN-Sigmoid, the DRSN-ReLU, the DRSN-LReLU, the DRSN-PReLU, the DRSN-APReLU and the DRSN-GPReLU are represented in figure 13, when the SNR = 6 dB. From the confusion matrix in figure 13, our developed method achieves the best diagnostic performance compared to other methods.

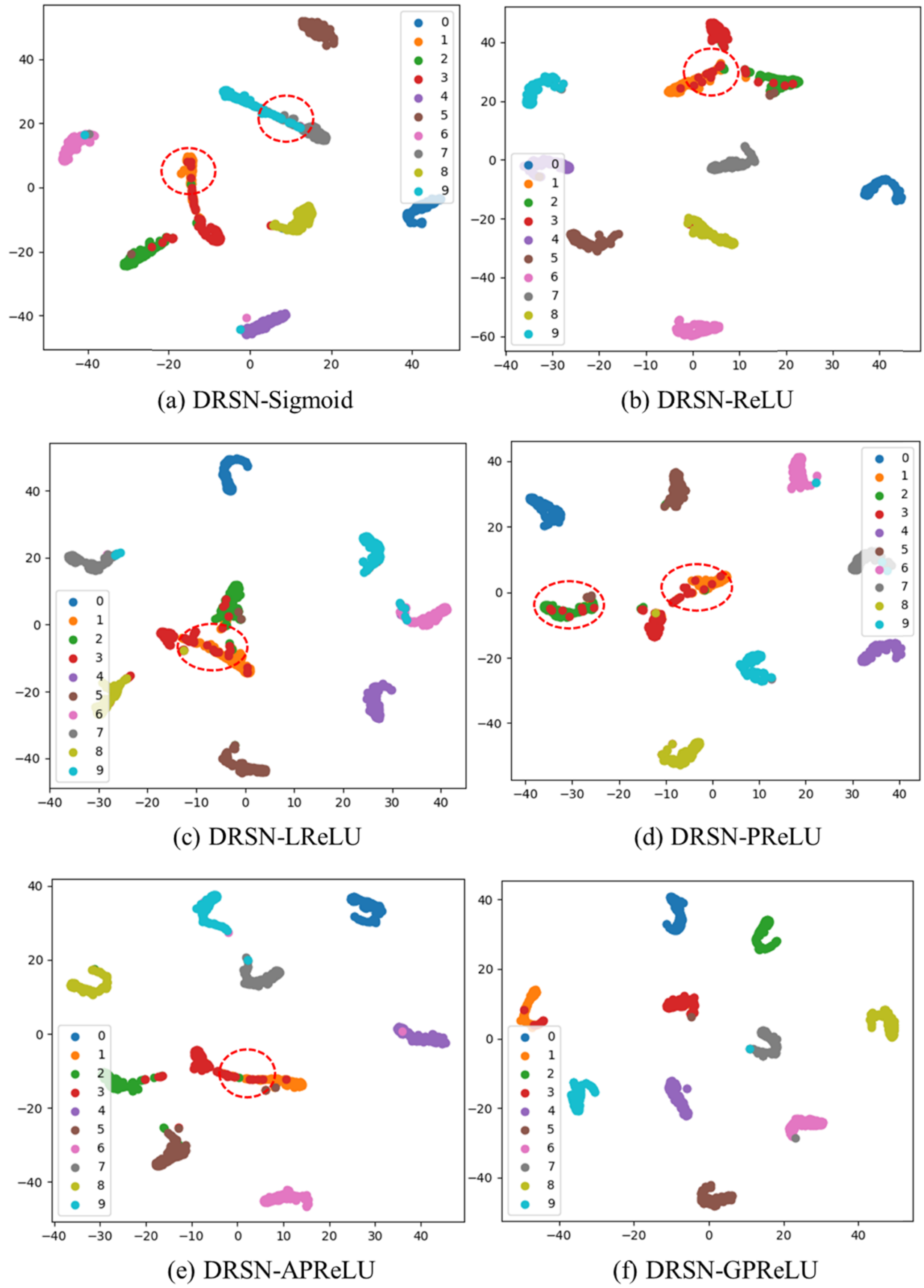


Figure 14. Visualization of feature spaces of (a)–(e) other methods and (f) the DRSN-GPReLU.

Then, the superiority of the DRSN-GPReLU is graphically demonstrated using the t-SNE method from the viewpoint of spatial feature distribution. As indicated in figure 14(f), it can be observed that the sample points after dimensionality reduction become more separable in the DRSN-GPReLU proposed in this paper under variable working conditions. In contrast, as indicated in figures 14(a)–(e), there are still some classes with highly overlapping features in the other methods.

3.3. The comparison with previous studies on bearing fault diagnosis

In this section, the reported studies on bearing fault diagnosis are given and compared with our proposed method in this paper. Table 6 demonstrates that there has not been enough research carried out on bearing fault diagnosis under variable working conditions. In addition, better results have been

Table 6. The reported studies on bearing fault diagnosis.

Methods	Dataset	Fault type	Accuracy (%)
FMM-RF (2017) [34]	• CWRU	• Normal condition (N), inner ring fault (IF), outer ring fault (OF), and ball fault (BF)	• 99.90
SVM + PCA (2018) [35]	• Experimental setup of authors	• N, IF, OF, and BF	• 97.44
CNN (2019) [36]	• CWRU	• N, IF, OF, and BF	• 97.74
SGMM (2019) [37]	• CWRU	• N, IF, OF, and BF	• 100
	• The bearing data of Hunan University	• N, IF, and OF	• 99.83
Signal2Image + LBP (2020) [38]	• Experimental setup of authors	• N, IF, OF, and BF	• 100
		• Fault size	• 100
		• Motor speed	• 95.90
DFT–IDFT autoencoders (2020) [39]	• Experimental setup of authors	• N, IF, OF, BF, Cage	• 99.92
Spark-IRFA (2021) [40]	• CWRU	• N, IF, OF, and BF	• 98.12
RCFOA-ELM (2021) [41]	• CWRU	• N, IF, OF, and BF	• 98.34
		• N, and IF (consider fault size)	• 95.34
GMA-DRSNs (2022) [20]	• CWRU	• N, IF, OF, and BF (consider fault size and noise conditions)	• 100
VMD + FE + IBOA-DBN (2022) [42]	• Experimental setup of authors	• N, IF, OF, and BF	• 98.33
	• CWRU		• 100
MD + SGM + CNN (2022) [43]	• CWRU	• N, IF, OF, and BF	• 97.21
Proposed method	• Experimental setup of authors	• N, IF, OF, and BF (consider variable working conditions and noise conditions)	• 99.3
	• CWRU	• N, IF, OF, and BF (consider variable working conditions, fault size and noise conditions)	• 99.03

achieved in several previous studies, some of which have even reached 100% accuracy. However, they ignore the effect of variable working conditions and noise conditions. In this paper, the DRSN-GPReLU can still achieve more than 99% diagnosis accuracy under variable working conditions and noise conditions, which illustrates the superiority of the proposed method.

4. Conclusions

Variable working conditions inevitably bring a challenge to the effectiveness of traditional intelligent diagnostic methods. Specifically, the characteristics of different vibration signal samples in the same health state may vary as the working conditions change and, similarly, the characteristics of different samples in different health states may become similar as the working conditions change. Therefore, the traditional activation function uses fixed nonlinear transformation for each sample, which reduces the classification ability of the model, i.e. projecting the intra-class signals to neighboring regions and the inter-class signals to distant regions.

To effectively solve the above problem, this paper develops a novel deep neural network architecture called the DRSN-GPReLU. Specifically, firstly, the DRSN is used as the basic architecture to improve the anti-noise ability of this method. Secondly, a new activation function called the GPReLU is developed, which is the main innovation of this article. The advantages of the GPReLU over traditional activation functions are: (a) it can perform adaptive nonlinear transformations on different vibration signal samples, thus enhancing the classification capability of the model; and (b) the nonlinear transformation of the GPReLU takes into account global features, thus allowing features to be extracted more sufficiently. Finally, a novel sub-network is designed based on an attention mechanism, to automatically infer the slope of the GPReLU.

The results of experiments on the two bearing fault diagnosis datasets have shown that the GPReLU proposed in this study outperforms other traditional activation functions under variable working conditions. The primary factor contributing to the improvement is that the attention GPReLU can perform adaptive nonlinear transformation for each vibration signal sample and considers the global signal features, which allows the DRSN-GPReLU to achieve a higher level of diagnosis accuracy under variable working conditions. We highlight that the GPReLU developed in this article can easily be inserted into other deep learning architectures to improve their performance.

In this study, the GPReLU has a more complex network structure than traditional activation functions and, therefore, inevitably increases the computational complexity, which leads to lower execution efficiency in real industrial field applications. In future study, we will further optimize the network structure of the GPReLU and reduce its computational complexity.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://engineering.case.edu/bearingdatacenter/welcome.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (No. 51675091).

Conflict of interest

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

ORCID iDs

Zhijin Zhang https://orcid.org/0000-0001-8317-7335
Huaitao Shi https://orcid.org/0000-0002-2808-5638
He Li https://orcid.org/0000-0002-9778-6606

References

[1] Shi H T, Li Y Y, Bai X T, Wang Z N, Zou D F, Bao Z and Wang Z 2021 Investigation of the orbit-spinning behaviors of the outer ring in a full ceramic ball bearing-steel pedestal system in wide temperature ranges *Mech. Syst. Signal Process.* **149** 107317

[2] Cheng Y W, Lin M X, Wu J, Zhu H P and Shao X Y 2021 Intelligent fault diagnosis of rotating machinery based on continuous wavelet transform-local binary convolutional neural network *Knowl.-Based Syst.* **216** 106796

[3] Zhang K, Wang J Y, Shi H T, Zhang X C and Tang Y H 2021 A fault diagnosis method based on improved convolutional neural network for bearings under variable working conditions *Measurement* **182** 109749

[4] Huang W G, Gao G Q, Li N, Jiang X X and Zhu Z K 2019 Time-frequency squeezing and generalized demodulation combined for variable speed bearing fault diagnosis *IEEE Trans. Instrum. Meas.* **68** 2819–29

[5] Xu Y G, Tian W K, Zhang K and Ma C Y 2019 Application of an enhanced fast kurtogram based on empirical wavelet transform for bearing fault diagnosis *Meas. Sci. Technol.* **30** 035001

[6] Bai X T, Shi H T, Sun J, Wang Z N and Zou D F 2021 Localization of double spalling on the outer rings of paired ball bearings by time-delay sound analysis *Meas. Sci. Technol.* **32** 094002

[7] Jiang X X, Huang Q, Shen C Q, Wang Q, Xu K, Liu J, Shi J J and Zhu Z K 2020 Synchronous chirp mode extraction: a promising tool for fault diagnosis of rolling element bearings under varying speed conditions *Chin. J. Aeronaut.* **35** 348–64

[8] Chen X W and Feng Z P 2019 Time-frequency space vector modulus analysis of motor current for planetary gearbox fault diagnosis under variable speed conditions *Mech. Syst. Signal Process.* **121** 636–54

[9] Kaya Y, Kuncan M, Kaplan K, Minaz M R and Ertunc H M 2020 Classification of bearing vibration speeds under 1D-LBP based on eight local directional filters *Soft Comput.* **24** 12175–86

[10] Kaya Y, Kuncan M, Kaplan K, Minaz M R and Ertunc H M 2021 A new feature extraction approach based on one dimensional gray level co-occurrence matrices for bearing fault classification *J. Exp. Theor. Artif. Intell.* **33** 161–78

[11] Kuncan M 2020 An intelligent approach for bearing fault diagnosis: combination of 1D-LBP and GRA *IEEE Access* **8** 137517–29

[12] Xia Y, Shen C Q, Wang D, Shen Y J, Huang W G and Zhu Z K 2022 Moment matching-based intraclass multisource domain adaptation network for bearing fault diagnosis *Mech. Syst. Signal Process.* **168** 108697

[13] Zhu J, Chen N and Shen C Q 2021 A new multiple source domain adaptation fault diagnosis method between different rotating machines *IEEE Trans. Ind. Inf.* **17** 4788–97

[14] Chen L, Li Q, Shen C Q, Zhu J, Wang D and Xia M 2021 Adversarial domain-invariant generalization: a generic domain-regressive framework for bearing fault diagnosis under unseen conditions *IEEE Trans. Ind. Inf.* **18** 1790–800

[15] He K M, Zhang X Y, Ren S Q and Sun J 2016 Deep residual learning for image recognition *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (*Seattle, WA*, *27–30 June*)

[16] Zhang W, Li X and Ding Q 2019 Deep residual learning-based fault diagnosis method for rotating machinery *ISA Trans.* **95** 295–305

[17] Zhao M H, Kang M, Tang B P and Pecht M 2018 Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes *IEEE Trans. Ind. Electron.* **65** 4290–300

[18] Ma S, Chu F L and Han Q K 2019 Deep residual learning with demodulated time-frequency features for fault diagnosis of planetary gearbox under nonstationary running conditions *Mech. Syst. Signal Process.* **127** 190–201

[19] Zhao M H, Zhong S S, Fu X Y, Tang B P and Pecht M 2020 Deep residual shrinkage networks for fault diagnosis *IEEE Trans. Ind. Inf.* **16** 4681–90

[20] Zhang Z J, Chen L, Zhang C L, Shi H T and Li H 2022 GMA-DRSNs: a novel fault diagnosis method with global multi-attention deep residual shrinkage networks *Measurement* **196** 111203

[21] Liu X Y, He Y G and Wang L 2021 Adaptive transfer learning based on a two-stream densely connected residual shrinkage network for transformer fault diagnosis over vibration signals *Electronics* **10** 2130

[22] Zhang Z J, Li H, Chen L and Han P 2021 Rolling bearing fault diagnosis using improved deep residual shrinkage networks *Shock Vib.* **2021** 9942249

[23] Ma X, Hu H and Shang Y Z 2021 A new method for transformer fault prediction based on multifeature enhancement and refined long short-term memory *IEEE Trans. Instrum. Meas.* **70** 1–11

[24] Han B K, Zhang X, Wang J R, An Z H, Jia S X and Zhang G W 2021 Hybrid distance-guided adversarial network for intelligent fault diagnosis under different working conditions *Measurement* **176** 109197

[25] Liao Y X, Huang R Y, Li J P, Chen Z Y and Li W H 2020 Deep semisupervised domain generalization network for rotary machinery fault diagnosis under variable speed *IEEE Trans. Instrum. Meas.* **69** 8064–75

[26] Zhao B, Zhang X M, Li H and Yang Z B 2020 Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions *Knowl.-Based Syst.* **199** 105971

[27] Nair V and Hinton G E 2010 Rectified linear units improve restricted Boltzmann machines Vinod Nair *Int. Conf. on Machine Learning* (*Haifa, Israel*, *21–24 June*) pp 807–14

[28] Maas A L, Hannun A Y and Ng A Y 2013 Rectifier nonlinearities improve neural network acoustic models *Int. Conf. on Machine Learning* (*Atlanta, Georgia, USA*, *16–21 June*)

[29] He K M, Zhang X Y, Ren S Q and Sun J 2015 Delving deep into rectifiers: surpassing human-level performance on imagenet classification *IEEE Int. Conf. on Computer Vision* (*Santiago, Chile*, *11–18 December*) pp 1026–34

[30] Shao H D, Jiang H K, Li X Q and Wu S P 2018 Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine *Knowl.-Based Syst.* **140** 1–14

[31] Zhao X L, Yao J Y, Deng W X, Ding P, Ding Y F, Jia M P and Liu Z 2022 Intelligent fault diagnosis of gearbox under variable working conditions with adaptive intraclass and interclass convolutional neural network *IEEE Trans. Neural Netw. Learn. Syst.* 1–15

[32] Zhao M H, Zhong S S, Fu X Y, Tang B P, Dong S J and Pecht M 2021 Deep residual networks with adaptively parametric rectifier linear units for fault diagnosis *IEEE Trans. Ind. Electron.* **68** 2587–97

[33] van der Maaten L and Hinton G 2008 Visualizing data using t-SNE *J. Mach. Learn. Res.* **9** 2579–605

[34] Seera M, Wong M L D and Nandi A K 2017 Classification of ball bearing faults using a hybrid intelligent model *Appl. Soft Comput.* **57** 427–35

[35] Gu Y-K, Zhou X-Q, Yu D-P and Shen Y-J 2018 Fault diagnosis method of rolling bearing using principal component analysis and support vector machine *J. Mech. Sci. Technol.* **32** 5079–88

[36] Hoang D-T and Kang H-J 2019 Rolling element bearing fault diagnosis using convolutional neural network and vibration image *Cogn. Syst. Res.* **53** 42–50

[37] Pan H Y, Yang Y, Zheng J D, Li X and Cheng J S 2019 A fault diagnosis approach for roller bearing based on symplectic geometry matrix machine *Mech. Mach. Theory* **140** 31–43

[38] Kaplan K, Kaya Y, Kuncan M, Minaz M R and Ertunc H M 2020 An improved feature extraction method using texture analysis with LBP for bearing fault diagnosis *Appl. Soft Comput.* **87** 106019

[39] He M and He D 2020 A new hybrid deep signal processing approach for bearing fault diagnosis using vibration signals *Neurocomputing* **396** 542–55

[40] Wan L J, Gong K, Zhang G, Yuan X P, Li C Y and Deng X J 2021 An efficient rolling bearing fault diagnosis method based on spark and improved random forest algorithm *IEEE Access* **9** 37866–82

[41] He C, Wu T, Gu R W, Jin Z Y, Ma R J and Qu H Y 2021 Rolling bearing fault diagnosis based on composite multiscale permutation entropy and reverse cognitive fruit fly optimization algorithm—extreme learning machine *Measurement* **173** 108636

[42] Jin Z Z and Sun Y Q 2022 Bearing fault diagnosis based on VMD fuzzy entropy and improved deep belief networks *J. Vib. Eng. Technol.* (https://doi.org/10.1007/s42417-022-00595-9)

[43] Xiong J B, Li C L, Wang C-D, Cen J, Wang Q and Wang S Z 2022 Application of convolutional neural network and data preprocessing by mutual dimensionless and similar gram matrix in fault diagnosis *IEEE Trans. Ind. Inf.* **18** 1061–71