



PAPER

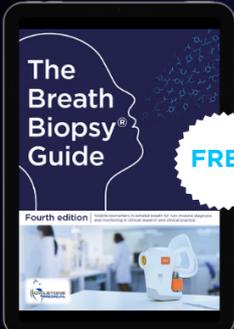
A small sample bearing fault diagnosis method based on novel Zernike moment feature attention convolutional neural network

To cite this article: Yunji Zhao and Jun Xu 2024 *Meas. Sci. Technol.* **35** 066208

View the [article online](#) for updates and enhancements.

You may also like

- [Development of a 3D cell printed construct considering angiogenesis for liver tissue engineering](#)
Jin Woo Lee, Yeong-Jin Choi, Woon-Jae Yong et al.
- [Paper-supported co-culture system for dynamic investigations of the lung-tropic migration of breast cancer cells](#)
Dongguo Lin, Xiao Chen, Zhun Lin et al.
- [Cost and greenhouse gas emission tradeoffs of alternative uses of lignin for second generation ethanol](#)
Ghasideh Pourhashem, Paul R Adler, Andrew J McAloon et al.



FREE **The Breath Biopsy® Guide**
Fourth edition

[DOWNLOAD THE FREE E-BOOK](#)

BREATH BIOPSY



A small sample bearing fault diagnosis method based on novel Zernike moment feature attention convolutional neural network

Yunji Zhao^{1,2,*}  and Jun Xu² 

¹ School of Electrical Engineering and Automation, Henan Polytechnic university, Jiaozuo 454003, People's Republic of China

² Henan International Joint Laboratory of Direct Drive and Control of Intelligent Equipment, Jiaozuo 454003, People's Republic of China

E-mail: ayjz@hpu.edu.cn

Received 11 January 2024, revised 7 March 2024

Accepted for publication 14 March 2024

Published 26 March 2024



CrossMark

Abstract

Bearings are one of the core components of rotating machine machinery. Monitoring their health status can ensure the safe and stable operation of rotating machine equipment. The limited nature of bearing fault samples makes it difficult to meet the demand for sufficient samples based on deep learning methods. Therefore, how to solve the problem of small-samples is the key to achieving intelligent fault diagnosis. In bearing failures based on vibration signals, the complex operating environment causes the vibration signals to inevitably mix with noise. The mixing of fault signature features and noise intensifies the strong spatial coupling of different types of fault features. In addition, diagnosing bearing failures under different loads is challenging because of the complex working conditions of bearings. Given the above problems, a small sample-bearing fault diagnosis method based on a high and low-frequency layered algorithm (HLFLA) and a novel Zernike moment feature attention convolutional neural network (ZMFA-CNN) is proposed. First, the proposed HLFLA converts one-dimensional time series signals into two-dimensional signals distributed rectangularly according to different frequency bands, and is used to simplify network feature screening, reduce the impact of noise, and retain adjacent signal constraint information. In addition, a new ZMFA-CNN is proposed to further extract multi-order moment features and attention weights, and can significantly improve the model generalization ability without increasing model parameters. At the same time, it is combined with FilterResponseNorm2d and thresholded linear unit to further improve model performance. Finally, sufficient experiments verified that the algorithm proposed in this paper can solve the above problems and has excellent transfer generalization ability and noise robustness. In addition, the experimental results of applying the algorithm proposed in this article to gas turbine main bearing fault diagnosis prove the reliability of the algorithm proposed in this article.

Keywords: bearing, fault diagnosis, HLFLA, ZMFA-CNN

* Author to whom any correspondence should be addressed.

1. Introduction

As the central component of rotating machinery, the operating status of bearings directly affects the performance and safety of the entire mechanical system [1–3]. During long-term high-speed operation, bearings are easily affected by various factors, such as load, temperature, grease, etc which can easily lead to bearing surface damage, wear, or cracks [1, 4]. If not discovered and dealt with in time, it will lead to loss of bearing function and even cause failure and accidents of the entire mechanical system. Therefore, in engineering practice, effective detection and diagnosis of bearing faults is an important means to improve the reliability of mechanical equipment and reduce maintenance costs.

In recent years, with the rapid development of artificial intelligence, sensor technology, information processing, and other technologies, fault diagnosis-related technologies have made great progress, and diagnostic methods have become increasingly diversified. Bearing fault diagnosis methods can usually be divided into traditional signal processing-based methods; mechanism modeling-based methods; and data-driven intelligent fault diagnosis methods. Methods based on signal processing, such as frequency domain analysis [5], time domain analysis [6], Wavelet transform (WT) [7, 8], and mode decomposition (MD) [9, 10]. This type of method mainly achieves fault diagnosis by processing information and extracting features from the system's measurement signals. There are also methods based on mechanism modeling: unscented Kalman filter (UKF) [11], Particle filter [12], Sliding mode observer [13], etc. These methods use state observers or filters to estimate the state variables of the system, ultimately enabling fault diagnosis. As well as intelligent fault diagnosis methods based on machine learning that has emerged with the development of artificial intelligence: support vector machines (SVMs) [14, 15], K-nearest neighbor algorithm (KNN) [16], artificial neural network (ANN) [17], deep learning and so on. Each of these methods has its own advantages and disadvantages. Compared with the first two types of methods, intelligent diagnosis methods do not require prior knowledge and can directly establish fault diagnosis models through data drive, so they have received widespread attention from researchers.

Among intelligent diagnosis technologies based on machine learning, SVM, ANN, and KNN are limited by their simple structure and difficulty in extracting high-order features, and their fault diagnosis performance under complex tasks is poor. Therefore, deep learning-based models are widely used in fault diagnosis due to their powerful feature extraction capabilities. Such as Convolutional neural networks (CNNs) [18], recurrent neural networks (RNNs) [19], and transformer [20]. These methods are widely used, and after continuous improvement, they have achieved excellent results in bearing fault diagnosis under tasks such as noise and variable working conditions. For example, [21] proposed MANANR and designed two multi-scale noise reduction modules, which can adaptively eliminate noise in multi-scale convolution features; [22] proposed a method that combines RNN and GRU, its excellent anti-noise ability has been verified

through experiments; [23] proposed Wavelet-SANet, which uses the self-attention mechanism and Transformer module to suppress noise in the wavelet domain and time-domain scattering impulse noise respectively. Li *et al* [24] improved the IICN and Inception network and finally realized the fault diagnosis of bearings with variable speed [25] Proposed MECNN, which uses Efficient channel attention to enhance features and combines CNN and Mode normalization to effectively solve the bearing variable load problem; [26] proposed a method based on CWT and T-ResNet, using transfer learning to effectively solve the problem of variable speed and load. These methods have made great contributions to the development of bearing fault diagnosis. However, for the above two problems, bearing fault diagnosis based on small samples is still a challenge worthy of study.

As we all know, as neural networks become more complex, the feature extraction capabilities of models become stronger and stronger. The various CNNs and Transformers that have emerged have varying degrees of anti-noise capabilities. With methods such as meta-learning and transfer learning, the problem of changing working conditions has also been properly solved. However, although these methods have powerful feature extraction capabilities, the number of model parameters inevitably increases. Parameter growth may not have an impact on tasks with sufficient training samples. In the field of fault diagnosis, it is not easy to obtain sufficient and effective fault samples for bearing fault diagnosis. Therefore, How to use a small number of training samples to solve the problems of noise and complex working conditions in fault diagnosis has huge research value. To this end, [27] proposed DCA-BiGRU, which can effectively realize small sample fault diagnosis and has excellent generalization ability and noise immunity [28]. Proposed C-ECAFormer, which solves the small sample problem by reducing the number of model parameters and has heavy noise robustness [29]. Proposed DRHRML to reduce noise through data reconstruction and meta-learning and achieve small sample fault diagnosis under various working conditions.

Given the impact of the above problems on bearing fault diagnosis, to solve the above problems more effectively and further improve the performance of bearing fault diagnosis. This paper proposes a small sample bearing fault diagnosis method based on the High-low frequency layering algorithm and the new Zernike moment feature attention CNN (ZMFA-CNN).

First, the high and low-frequency layered algorithm (HLFLA) is proposed. This algorithm uses DFT to convert time-domain signals into frequency-domain signals to increase the distinction between fault features and noise, making it easier for the network to extract fault features and suppress noise. Then the matrix spiral filling algorithm is used to convert the one-dimensional signal into a two-dimensional signal. The signal processed by the algorithm appears in the form of a rectangular distribution with high frequency in the center and low frequency on the edge, which facilitates the network to allocate weights to different frequency bands. In addition, this algorithm can retain the correlation of adjacent position information and avoid the loss of features. Second,

the ZMFA-CNN is proposed to obtain Zernike multi-order moment features of different frequency bands online and assign weights to each frequency band, thereby extracting different levels of features and obtaining effective features.

The main contributions of the paper are as follows:

- (i) For small-sample bearing fault diagnosis with heavy noise and multiple working conditions, a new fault diagnosis method based on HLFLA and ZMFA-CNN is proposed, and the effects of different loss functions, optimizers, normalization methods, and activation functions are discussed. In addition, this method has higher testing accuracy.
- (ii) A new network model is proposed, which improves the Zernike moment feature acquisition algorithm and embeds it into the neural network. The improved algorithm has better results than the traditional algorithm, and compared with the traditional offline method of obtaining Zernike moment features, the computational efficiency is greatly improved. The proposed ZMFA-NN can extract rich fault information with a small number of model parameters, can effectively avoid small sample overfitting, and has superior generalization ability.
- (iii) A method to convert one-dimensional signals into two-dimensional signals is proposed. The proposed HLFLA is an improvement on the existing matrix spiral filling algorithm (MSFA), and its purpose is to solve the defects of Zernike moment features in fault diagnosis. In addition, the problem of correlation loss of adjacent information existing in the traditional one-dimensional signal to two-dimensional signal conversion method can be avoided, and the accuracy of small sample bearing fault diagnosis can be effectively improved.

The rest of this paper is organized as follows: section 2 introduces the theory of high and low-frequency hierarchical algorithms and novel ZMFA-CNNs. Section 3 introduces in detail the proposed bearing fault diagnosis method based on the high and low-frequency hierarchical algorithm and the new ZMFA-CNN. In section 4, we conduct a variety of experiments to comprehensively analyze the performance of the algorithm. Finally, section 5 provides a summary and prospects for future research.

2. Theoretical backgrounds

2.1. High-low frequency layering algorithm

For bearing fault diagnosis in small samples and complex tasks, effective preprocessing of data can improve the accuracy of fault diagnosis to a certain extent. To effectively tap the performance of ZMFA-CNN and retain the ability of 2D convolution to extract spatial features, this paper proposes HLFLA, a preprocessing method for converting one-dimensional data to two-dimensional data. This algorithm can facilitate the network to extract fault characteristics and suppress noise, as well as retain the constraint information of adjacent data.

Compared with traditional matrix transformation, HLFLA can retain the correlation of adjacent position information and avoid the loss of features. Compared with existing 1D to 2D data preprocessing methods, such as Grammi angle field (GAFS) [30], Markov transition field (MTF) [31], Mel-frequency cepstrum coefficient (MFCC) [32]. HLFLA and ZMFA-CNN have better synergy effects, and experiments have proved that HLFLA has a better fault diagnosis effect. The HLFLA preprocessing process is as follows. Figure 1 shows the overall process of HLFLA and the comparison effect with matrix transformation.

2.1.1. Pre-emphasis, dispersed Fourier transform. Pass the original vibration signal through a high-pass filter to enhance the high-frequency part of the signal. Then DFT is performed to convert the time domain fault signal and noise into frequency domain signals according to frequency distribution, which facilitates the network to extract fault characteristics and suppress noise. The formulas for pre-emphasis and DFT are as follows:

$$g(t) = x(t) - \alpha x(t-1) \quad (1)$$

$$S_i(k) = \sum_{n=1}^N s_i(n) e^{-j2\pi kn} \quad 1 \leq k \leq K \quad (2)$$

2.1.2. Data dimension transformation based on matrix spiral filling algorithm. The process of the MSFA is roughly shown in the lower part of figure 1. After DFT, the signal appears as a frequency distribution from low to high. At this time, the high-frequency part is placed in the center of the matrix according to the spiral structure, and the low-frequency part is placed in the edge area of the matrix. After filling, the matrix is composed of multiple layers of frequency bands, which facilitates ZMFA-CNN to assign weights to different frequency bands and can retain the correlation of adjacent position information to avoid the loss of features. MSFA significantly improves fault diagnosis performance by combining pre-emphasis and DFT. And the optimized algorithm runs at a very impressive speed, with almost no increase in time burden.

2.2. A novel ZMFA-CNN

Fault diagnosis based on small samples and complex tasks has strict requirements on the parameter amount of the neural network model. When the number of model parameters is too large, a small number of samples may lead to overfitting; however, when the number of model parameters is too small, the model may perform poorly in the face of complex tasks such as noise and changing working conditions. To this end, this article proposes ZMFA-CNN, which has a small number of parameters but strong model generalization capabilities and can effectively realize bearing fault diagnosis based on small samples and complex tasks. The structure of the network is as follows.

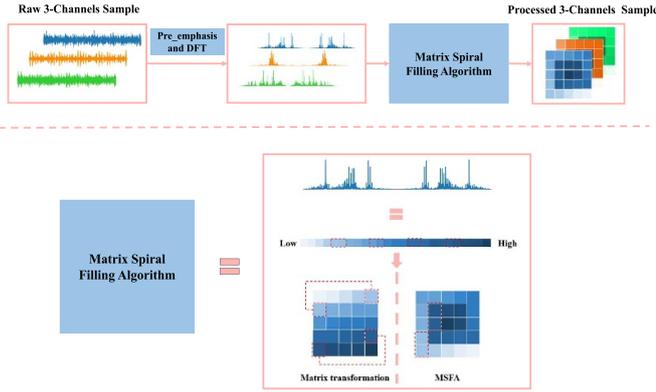


Figure 1. The implementation process of HLFLA, the upper part is the overall process, the lower part is the MSFA schematic diagram and the comparison with the traditional matrix transformation.

2.2.1. Zernike moment feature enhancement. To solve the problem of changing working conditions, this paper introduces the Zernike moment [33] into the field of bearing fault diagnosis for the first time. As shown in figure 2, take the two fault categories F1 and F2 in the CWRU data set as an example. The original fault sample signals of F1 and F2 have obvious shape differences after being converted into images. Under different loads, the shapes of the original fault sample 2D signals of the same fault category are roughly the same. The difference lies in the rotation and translational phenomena between the images. Zernike moments have rotation invariance and translation invariance, which means that they will not be affected by the rotation and translation of the image, and will not cause a significant decrease in fault diagnosis accuracy due to changes in load. The Zernike moment can effectively extract the edge and shape information of the image, which is important for distinguishing different fault types. In addition, in figure 2, the fault signal mapped by the Zernike moment feature presents multiple circular-shaped signal distributions. This is because Zernike moments can use different orders and times to represent different details of the image, which can improve the resolution of the image and help improve the discrimination of different faults.

However, there are advantages and disadvantages. The original method of extracting Zernike moment features is not effective in fault diagnosis, and it is slow and difficult to meet the needs of fault diagnosis. The following article will describe the extraction process of new Zernike moment features and the method to solve the problems. This article maps the frequency domain features of the input network into the Zernike moment space, which can effectively increase the richness and robustness of the features. The general way of Zernike moment feature mapping is shown in figure 3, the details are as follows:

- (i) Create a grid with a shape of (H, W). Each coordinate point in the grid corresponds to the input feature matrix one-to-one, denoted as $f(x, y)$. Then normalize the grid coordinates to the polar coordinates of the unit circle, Filter out the points within the circle, and record them as $\partial(r, \theta)$.

		Load	0 HP	1 HP	2 HP	3 HP
F0	Raw signals					
	HLFLA					
	Zernike					
F1	Raw signals					
	HLFLA					
	Zernike					

Figure 2. Renderings of the original data and processed data of two types of faults on the CWRU data set.

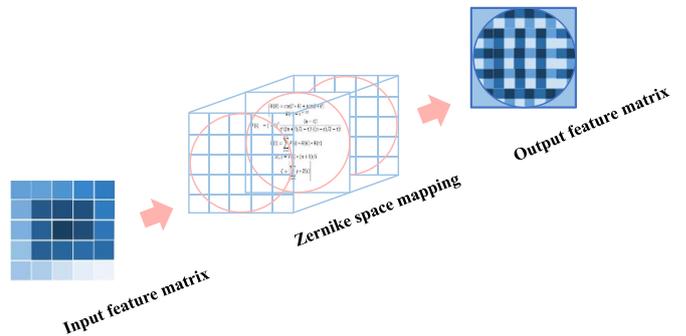


Figure 3. The general process of Zernike moment feature mapping.

$$\begin{cases} x^2 + y^2 \leq 1 \\ f(x, y) \end{cases} = \partial(r, \theta) \quad (3)$$

where x, y, r, θ represent the coordinates of the point in the rectangular coordinate system and the polar coordinate system respectively.

- (ii) Given the Zernike polynomial order and angle degree n, l , and select valid (n, l) pairs and their corresponding differences k according to the condition $\begin{cases} n - l \geq 0 \\ (n - l) \% 2 = 0 \end{cases}$. Then for each valid (n, l) pair, calculate their corresponding angular function value $\Theta(\theta)$ and radial function value $R(r)$. The calculation formula is as follows:

$$\begin{cases} \Theta(\theta) = \cos(l * \theta) + j \sin(l * \theta) \\ R(r) = r^{n-2l} \end{cases} \quad (4)$$

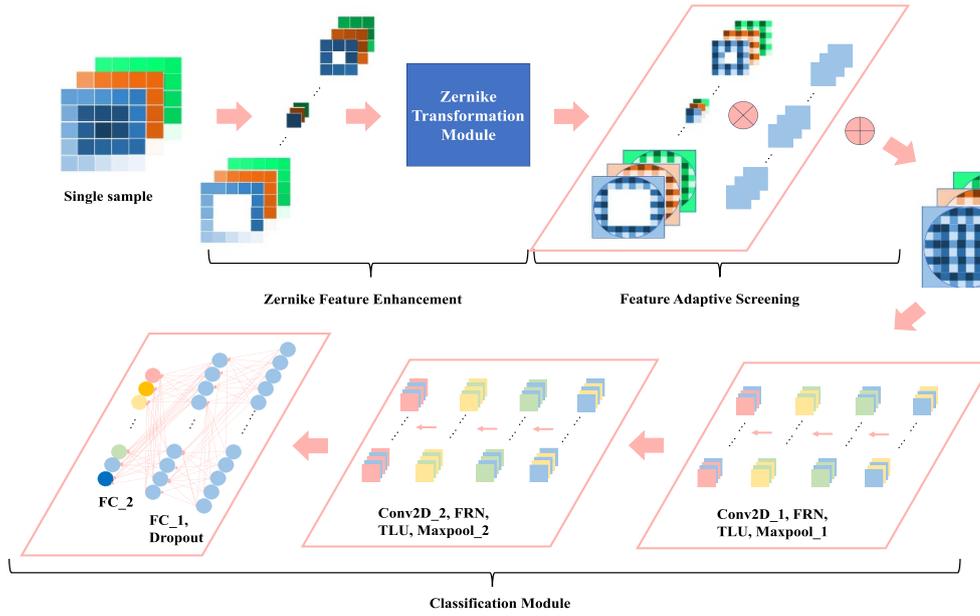


Figure 4. The overall structure of ZMFA-CNN. The ZMFA-CNN is divided into three modules, namely the Zernike feature enhancement module, feature adaptive screening module, classification module.

where j represents the imaginary part of the complex number. $t \in [0, k]$

- (iii) Then calculate the corresponding coefficient values $F(i)$, basis function values $V(i)$, and normalized Zernike polynomial values $Z(i)$ for each (n, l) pair.

$$\begin{aligned}
 F(i) &= (-1)^t \frac{(n-t)!}{t!((n+l)/2-t)!((n-t)/2-t)!} \\
 V(i) &= \sum_{t=0}^{k(i)} F(i) * \Theta(\theta) * R(r) \\
 Z(i) &= \bar{V}(i) * (n+1) / S
 \end{aligned} \tag{5}$$

where i represents the sequence number of each valid (n, l) pair, $\bar{V}(i)$ represents the complex conjugate of $V(i)$.

- (iv) Finally, the broadcast mechanism is used to simultaneously map each feature matrix $\gamma \in R^{H \times W}$ in the input sample $\Gamma \in R^{B \times C \times H \times W}$ to the Zernike space to obtain the output feature map $\vartheta \in R^{B \times C \times H \times W}$.

$$\zeta = \left| \sum_{i=0}^k \gamma * Z(i) \right| \tag{6}$$

where γ, ζ represent the input and output feature matrices respectively, $\gamma, \zeta \in R^{H \times W}$

The Zernike moment is widely used in the image field. In the image field, the features of a picture are often concentrated in the center area of the picture. Therefore, when extracting the Zernike moment feature, only the unit circle area with the center point of the picture is extracted, and the edge part is given up. However, bearing fault signals in the time domain is complex and disorderly, and it is difficult to distinguish effective fault information. Therefore, the original extraction method is not suitable, which leads to poor results. Therefore, the HLFLA proposed in this article is to solve this

problem. The purpose of proposing HLFLA is to transfer the signal to the frequency domain. Since previous research has confirmed that effective fault information of bearings often exists in high-frequency signals, MSFA is used to transfer the highest-frequency signal to the frequency domain. The signal is set as the center point of the matrix, and the characteristic matrix is mapped into the Zernike moment space according to different frequency bands.

Given the problem that the calculation speed of Zernike moment is very slow, embedding it directly into the neural network will seriously affect the training speed of the network. To this end, this article reconstructs the algorithm. The above formula is a new Zernike moment transformation algorithm. The improved algorithm is combined with the broadcast mechanism to realize the simultaneous calculation of multiple batches and multiple channels, which greatly improves the training speed of the network.

2.2.2. Multi-band feature attention screening (MBAS).

After the Zernike moment feature conversion, the feature map contains rich feature information. To further screen out effective information, based on the spiral distribution of frequency characteristics obtained by MSFA, this article divides the feature map into eight round frequency bands according to different frequencies, as shown in figure 4. Then a weight is assigned to each frequency band, and effective features are screened out through network training. Compared with assigning weights to the entire feature map, this method can effectively avoid overfitting.

2.2.3. Classification module.

The feature maps that have been filtered are used to extract deep features through two layers of convolution pooling, and finally, the fault diagnosis results are obtained through two fully connected layers.

The network parameters are shown in table 1. In addition, FilterResponseNorm2d (FRN) [34] and thresholded linear unit (TLU) [34] were introduced. Small samples generally require a small batch size, because the variance of small samples is large. If a large batch size is used, the model may fall into a local optimum, and a small batch size facilitates jumping out of the local optimum. Compared with batch normalization (BN), which is more suitable for large batch sizes, after experimental verification, FRN for small batch sizes is more suitable for the fault diagnosis method in this article. FRN normalizes each channel of each sample without considering the batch dimension or grouping dimension. This can avoid the impact of batch size or number of groups on the normalization effect, while also preserving the correlation between convolution kernels. Coupled with the TLU activation function to eliminate negative values and enhance nonlinear expression capabilities, experiments show that FRN+TLU has a significant effect on improving fault diagnosis accuracy. The FRN and TLU formulas are as follows:

$$y_{i,j,k} = \frac{x_{i,j,k}}{\sqrt{\frac{1}{H \times W} \sum_{h,w} x_{i,h,w,k}^2 + \varepsilon}} \times \gamma_k + \beta_k$$

$$Z_{i,j,k} = \max(y_{i,j,k}, \tau_k) \quad (7)$$

where $x_{i,j,k}$ is the value of the (j, k) th position of the i th channel of the input feature. H and W are the height and width of the input feature map. ε is a small positive quantity or a learnable parameter, and γ_k and β_k are learnable parameters used to adjust the scale and offset of the normalized features. τ_k is a learnable threshold parameter. $y_{i,j,k}$ is the output of the FRN layer, which normalizes the input features by two norms. $Z_{i,j,k}$ is the output of the TLU layer, which thresholds the output of the FRN layer.

2.3. Model training method

2.3.1. Label smoothing regularization (LSR). Introduce LSR [35] into bearing fault diagnosis to alleviate overfitting. Its principle is to use a softened label instead of the one-hot encoded label when calculating the cross-entropy loss function, that is, retaining a certain degree of confidence in the correct category and assigning some smaller probabilities to other categories. This can reduce the model's over-reliance on correct labels and improve the model's generalization ability and robustness. The LSR formula is as follows:

$$L = -\sum_{i=1}^N (q_i \log p_i + (1 - q_i) \log(1 - p_i)) \quad (8)$$

where N is the total number of categories, p_i is the probability of the i th category predicted by the model, and q_i is the softened label.

2.3.2. Ranger optimizer. Introduce the Ranger optimizer [36] to improve the convergence speed and generalization ability of the model. The essence of the Ranger optimizer is

the combination of Radam [37] and LookAhead [38], and it inherits the advantages of these two optimizers. It can dynamically adjust the adaptive momentum to avoid instability problems caused by excessive gradient variance in the early stages of training. It can also alternate between fast exploration and slow utilization of two-parameter spaces. This reduces sensitivity to hyperparameters while speeding up convergence to a smoother minimum.

3. The proposed method based on HLFLA and ZMFA-CNN

This section will introduce the overall process of the diagnosis method. The overall process structure of the proposed small sample bearing fault diagnosis method based on HLFLA and ZMFA-CNN is shown in figure 5. The specific parameter settings of ZMFA-CNN are shown in table 1.

- (i) Randomly collect bearing vibration signals through sensors and intercept them into fixed-size fault samples. HLFLA is used to process all samples, convert the samples from one dimension to two dimensions, and then divide them into training and testing as input to ZMFA-CNN. HLFLA is used to reconstruct the distribution of sample frequency domain signals, which facilitates network extraction and suppresses noise while retaining the constraint information of sample adjacent data.
- (ii) Input the training samples into the network. First, the Zernike matrix conversion module will map the frequency domain features into the Zernike matrix space, effectively increasing the richness and robustness of the features. Secondly, weights are set for different frequency bands, and fault iconic features are screened out from the rich Zernike moment features to suppress noise. Finally, a classification network composed of convolution, pooling, and fully connected layers is used to further extract features and obtain offline diagnosis results. This paper replaces the new normalization method and activation function, using FRN and TLU to meet the needs of small sample fault diagnosis. It was verified through experiments that FRN and TLU are most suitable for this algorithm.
- (iii) During the training process, LSR is used to optimize the cross-entropy loss function to avoid overfitting. The new Ranger optimizer is used to further improve the convergence speed and generalization ability of the model. Subsequent experiments verified the effects of different optimizers. Then, each trained model is tested using the validation set until the optimal model parameters are obtained.
- (iv) Although the use of public data sets can illustrate the performance of the algorithm to a certain extent, it is still not convincing. Therefore, this article conducts online testing on gas turbine-bearing data to simulate industrial

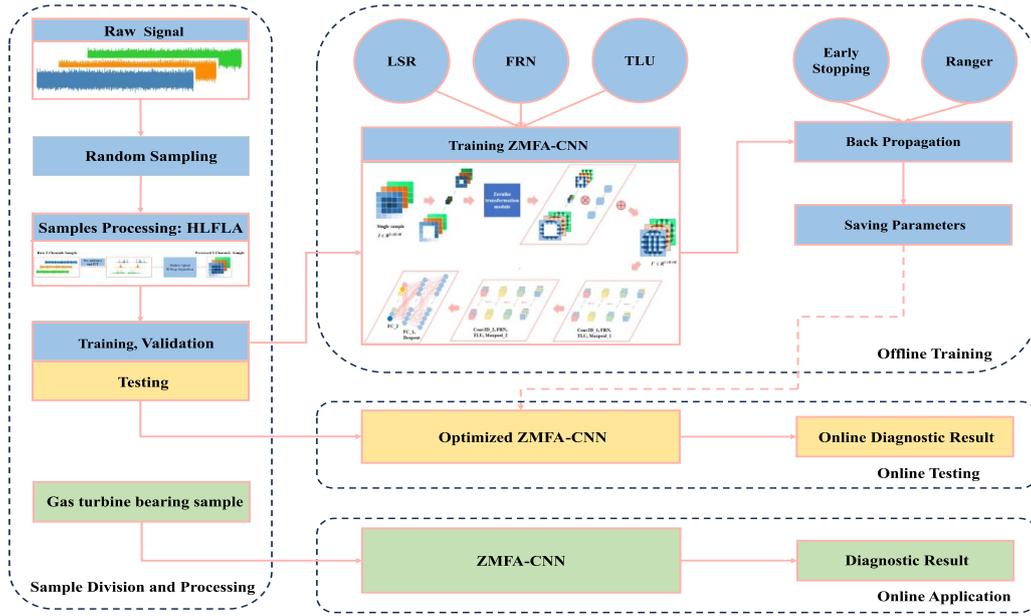


Figure 5. The overall process of bearing fault diagnosis method based on HLFLA and ZMFA-CNN.

Table 1. The structures of ZMFA-CNN.

Type	Kernel / Stride	Unit	TLU	FRN	Input	Output	Parameter
Zernike	/	/	/	/	[B,3,64,64]	[B,3,64,64]	/
MBAS	/	/	/	/	[B,3,64,64]	[B,3,64,64]	24
Conv2d_1	5 / 1	/	Yes	Yes	[B,3,64,64]	[B,16,60,60]	1216
Maxpool_1	2 / 2	/	/	/	[B,16,60,60]	[B,16,30,30]	/
Conv2d_2	5 / 1	/	Yes	Yes	[B,16,30,30]	[B,32,26,26]	12 832
Maxpool_2	2 / 2	/	/	/	[B,32,26,26]	[B,32,13,13]	/
FC_1	/	256	/	/	[B,5408]	[B,256]	1384 704
FC_2	/	10	/	/	[B,256]	[B,10]	2570
							Total: 1401 346

applications. This method proves that the algorithm in this paper can be used as a practical application.

4. Experiment validation

To verify the effectiveness of the proposed small-sample bearing fault diagnosis method, this paper conducts experiments on two public data sets. Through a large number of experiments, verify the small sample learning ability and anti-noise ability of different algorithms, as well as whether they have superior generalization ability under different working conditions. Then, To further verify the reliability of the algorithm in this paper, the bearing failure data collected during the operation of the gas turbine was used to test whether it is reliable in industrial applications. All experiments were conducted under the same set of random seeds, and the experimental settings are shown in table 2.

The experimental equipment required for this study is a laptop with Intel(R) Core(TM) I7-12 700 H processor, 32.0 GB memory and NVIDIA GeForce GTX 3060 graphics card

Table 2. Description of experimental parameters.

Settings	Value
Batch size	16
Epochs	50
Optimizer	Ranger
Learning rate	0.01
Loss function	Cross-entropy
Early Stopping (patience)	10

with 6GB video memory. Powered by Windows 11 system, PyCharm is used as the integrated development environment (IDE) and Python is used as the programming language.

4.1. Model evaluation and metrics method

In this article, the fault diagnosis performance is represented by a unified accuracy rate. Each experiment is obtained five times and the average value is obtained. *Acc* is defined as in equation(9)

Table 3. Partition of CWRU data set.

Data	Load	Location	FD (mm)	Label	α
A / B / C / D	0 / 1 / 2 / 3	IR	0.1778 / 0.5334	0 / 1	0.1 ~ 0.5
		B		2 / 3	
		OR_3		4 / 5	
		OR_6		6 / 7	
		OR_12		8 / 9	

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Acc = \frac{1}{5} \sum_{i=1}^5 Accuracy_i \quad (9)$$

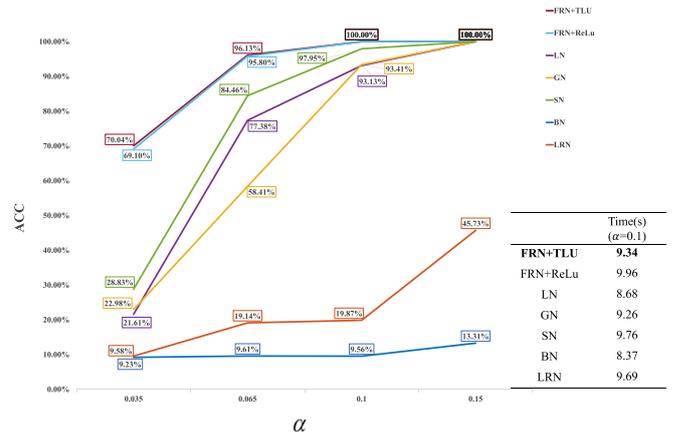
4.2. Case 1: data from CWRU

4.2.1. Description and division of CWRU data. The CWRU bearing data set [39] is provided by Case Western Reserve University. The bearings used are SKF bearings that support the motor. The bearing damage is caused by single-point damage caused by EDM. The single-point damage fault is divided into the inner ring (IR), outer ring (OR), and ball (B). The OR damage is located at 3 o'clock, 6 o'clock, and 12 o'clock. The sampling frequency is 12 Khz, and the load is 0 ~ 3 HP. The Drive end acceleration (DE) data, Fan end acceleration (FE) data, and Base acceleration (BA) data are collected through three vibration sensors.

This article divides four data sets A, B, C, and D according to different loads. Each data set includes ten categories of data and 100 samples in each category. Each sample consists of three channel data of DE, FE, and BA, and the sample size is 3×4096 . A parameter α is set as the proportion of the training set to the entire data set, the verification set proportion occupies 0.1, and the test set proportion is $(1-\alpha) \cdot 0.1$. The specific settings are shown in table 3, where FD represents different damage diameters.

4.2.2. Comparison of normalization layers. To verify FRN can improve the performance of the model under small batch size conditions, multiple normalization methods were used for comparison, including layer normalization (LN) [40], group normalization (GN) [41], switch normalization (SN) [42], BN [43], local response normalization. Data set B is used as the experimental data set. To make the comparison results more obvious, set the batch size to 32. The experimental settings are shown in table 2. The activation function is set to the ReLU function by default. The experimental results are shown in figure 6.

Compared with other methods, FRN can achieve more obvious model performance improvements with minimal time loss, especially in the case of an extremely small number of training samples $\alpha = 0.035$ (35). Moreover, FRN and TLU are complementary, so this article uses TLU as the activation function. The experimental results also show that TLU combined with FRN can further improve model performance and reduce training time.

**Figure 6.** Accuracy comparison and running speed comparison when $\alpha = 0.1$.

4.2.3. Ablation contrast experiment. As shown in figure 7, the ablation comparison test of HLFLA + ZMFA-CNN (M1) was conducted on the four data sets A, B, C, and D. The five methods compared were M2: DFT-R (DFT + Reshape) + ZMFA-CNN, M3: HLFLA + ZMF-CNN (M1 lacking the MBAS module), M4: HLFLA + CNN, M5: CNN, M6: ZMFA-CNN. *Acc* is used as the evaluation index to obtain the accuracy trend line in the figure, batch size = 32. The x-axis shows the proportion of training (α). Compare the accuracy of different methods under different load data sets and different training set proportions α , as well as the running time when $\alpha = 0.1$.

As can be seen from figure 7, as α increases, the test accuracy of all models continues to increase to 100%, indicating the impact of the amount of effective training data on fault diagnosis accuracy. The comparison between M1 and M6, M4 and M5 on four data sets shows that the HLFLA proposed in this article can significantly reduce the feature extraction burden of the model, and it can bring out the performance of Zernike moment feature enhancement module. In addition, through the comparison of M1 and M2, M1 and M3 on four data sets, it is verified that MSFA can retain the correlation of adjacent signals to improve fault diagnosis performance, and the attention screening module can improve the accuracy to a certain extent. All in all, the method proposed in this article has the best performance on the four data sets.

4.2.4. Evaluation under variable load. When the bearing is working, the load on the bearing is not constant. As the load changes, the collected data will also change. If different data

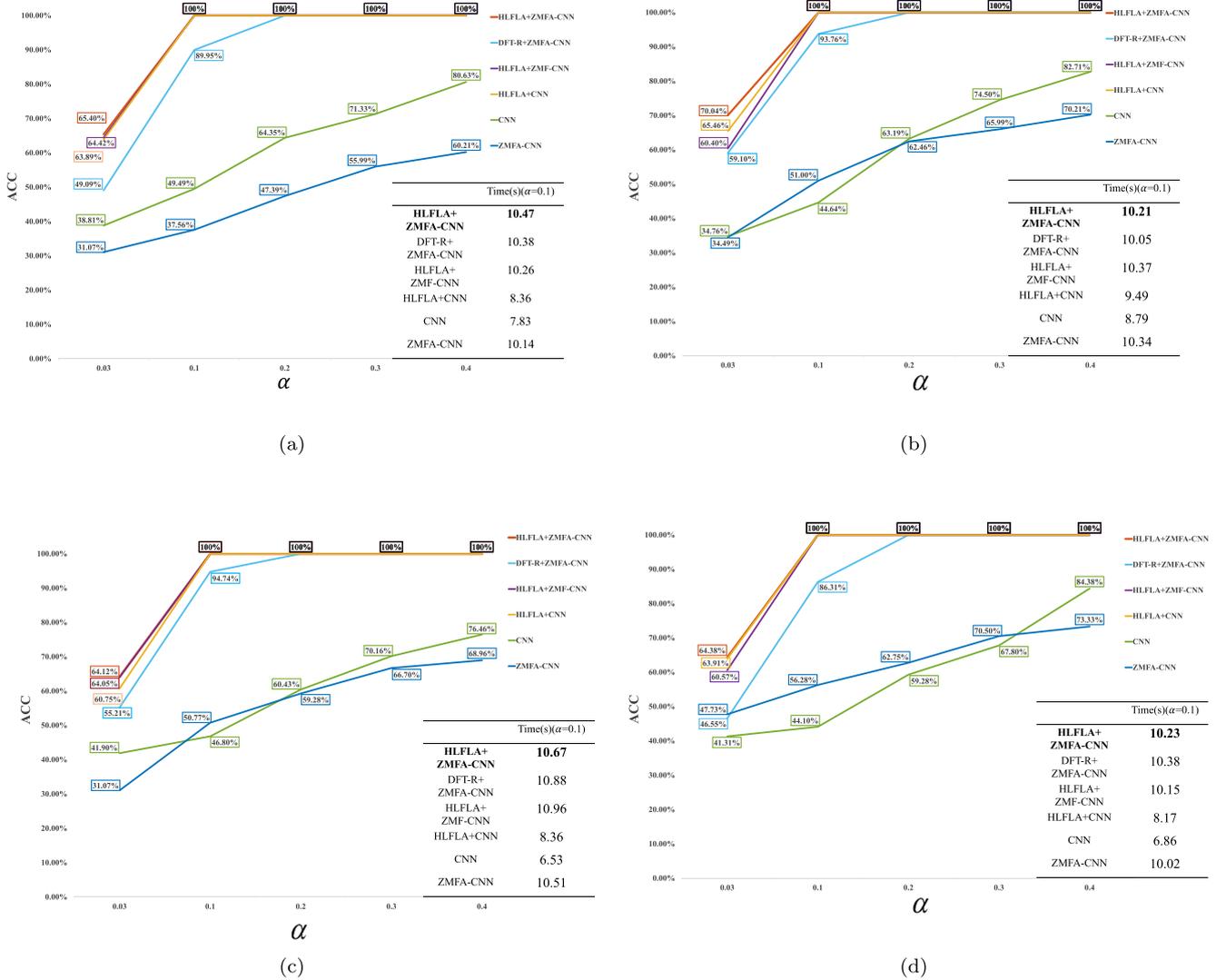


Figure 7. Accuracy comparison, (a) Data set A, (b) Data set B, (c) Data set C, (d) Data set D.

sets are produced for different workloads for training, this will be a tedious and inefficient task. Therefore, fault diagnosis methods with strong generalization capabilities and migration versatility are essential in practical applications.

To evaluate the migration ability of HLFLA+ZMFA-CNN, the data set B was used as the training set, and online tests were conducted on A, C, and D respectively. The load and data distribution are different between data sets. Under this condition, the versatility under small samples ($\alpha \leq 0.5$) is discussed. The experimental results are shown in table 4. From the table, it can be found that the Zernike moment feature enhancement module combined with the attention filtering module has excellent transfer generalization capabilities. Comparing the comparison results with NIN [44] and Convnext [45] and the actual training situation, inappropriate or complex models are extremely prone to overfitting under small sample conditions, and it is difficult to exert the performance of the model. In addition, for the two data sets with loads 2 and 3 (C, D), $Acc > 97\%$, showing good generalization ability for three-channel data.

However, the migration capability for data set A (load = 0) is reduced, and subsequent optimization is still required. Overall, it is pleasing to be able to achieve such results for a variable load task with a small sample based on three-channel data.

4.2.5. Optimizers comparison. To explore the most suitable adaptive optimization gradient algorithm and improve the versatility of the proposed fault diagnosis method as much as possible, the accuracy of various optimizers for variable load tasks was compared. Each of these optimizers uses an optimal initial optimization rate. The optimizers are Ranger (0.01), Radam (0.001), Adadelata (0.1) [46], Adam (0.001) [47], RMSprop (0.001) [48], AdaBound (0.001) [49], SGD (0.01) [50]. The comparison results are shown in figure 8. It can be clearly found that the effects of SGD and AdaBound are poor. The remaining optimizers have achieved an accuracy of more than 90%. However, after obtaining the Acc of different test sets, this article chose the Ranger optimizer.

Table 4. Acc under different training ratios and loads.

Method (model params size)	α	Acc(%)			
		B \rightarrow A	B \rightarrow B	B \rightarrow C	B \rightarrow D
HLFLA+ ZMFA-CNN (83.94 M)	0.1	99.67	100	99.89	96.88
	0.2	97.63	100	100	100
	0.3	94.91	100	97.82	98.11
	0.4	98.14	100	98.99	99.49
	0.5	98.59	100	100	97.78
HLFLA + CNN (83.93 M)	0.1	70.42	100	71.98	79.53
	0.2	72.62	100	68.62	73.15
	0.3	70.81	100	71.59	80.87
	0.4	69.33	100	77.90	83.67
	0.5	73.25	100	74.75	83.08
HLFLA + ConvNext (654.75 M)	0.1	83.67	95.64	83.22	81.91
	0.2	88.23	96.18	86.08	89.36
	0.3	88.12	98.51	87.16	90.32
	0.4	84.80	99.93	84.03	85.77
	0.5	90.29	100	85.04	90.92
ConvNext (654.75 M)	0.1	72.93	73.87	71.38	71.42
	0.2	78.08	84.49	71.41	72.82
	0.3	76.35	87.10	73.57	72.96
	0.4	78.33	95.37	84.47	84.87
	0.5	79.92	98.42	74.71	76.33
NIN (87.27 M)	0.1	69.68	85.33	79.45	79.21
	0.2	76.69	93.67	76.60	76.28
	0.3	76.32	98.33	80.34	81.80
	0.4	79.36	99.20	82.33	79.57
	0.5	85.38	100	84.79	83.88

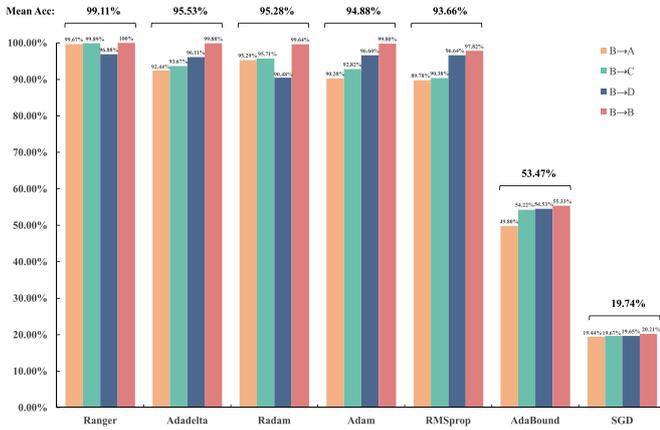


Figure 8. Accuracy comparison of optimizers.

4.2.6. Noise robustness experiment. In reality, signals mostly contain noise. Therefore, the study will analyze the anti-noise robustness under different signal-to-noise ratios (SNR), which is defined as in equation (10),

$$\begin{cases} P_{\text{signal}} = \frac{1}{N} \sum_{i=1}^N x_i^2 \\ SNR_{dB} = 10 \lg \left(\frac{P_{\text{signal}}}{P_{\text{noise}}} \right) \end{cases} \quad (10)$$

where P_{signal} represents the signal power and P_{noise} represents the noise power.

Gaussian white noise with $SNR = -6 \sim 6$ dB is added to the training data set for training. Previous studies have confirmed that the higher values of SNR and α , the higher accuracy of the test. For this purpose, the B data set of $\alpha = 0.1$ is used for performance testing. The accuracy and running time (data processing + training) under different SNR are shown in table 5.

In addition to testing the anti-noise ability of HLFLA + ZMFA-CNN, several other data preprocessing methods were also selected for comparison, namely GAFS, MTF, and MFCC. Experimental results show that this method has excellent anti-noise ability and can still achieve more than 95% Acc in a very noisy environment. Compared with other data preprocessing methods, this method is more suitable for small sample fault diagnosis and has stronger noise robustness.

4.3. Case 2: data from SUFD

4.3.1. Description and division of SUFD data. The SUFD dataset [51] is provided by Southeast University, which contains gearbox data and bearing data. This article selects bearing data for experiments. The data includes four fault types and one normal data. The faults are divided into Balls, IRs, ORs, and composite faults of inner and ORs. There are five

Table 5. Accuracy and running time at different SNR on CWRU data set.

Method	Acc(%) Time(s)													
	SNR = -6		SNR = -4		SNR = -2		SNR = 0		SNR = 2		SNR = 4		SNR = 6	
HLFLA + ZMFA-CNN	97.50	16.09	98.33	16.42	98.98	16.44	99.60	16.08	99.93	16.80	100	16.22	100	16.94
DFT-R + ZMFA-CNN	89.73	17.29	88.85	17.31	89.42	18.07	88.37	17.61	95.20	17.81	97.30	17.47	97.78	17.58
MFCC + ZMFA-CNN	83.73	13.29	84.11	13.95	86.65	14.11	87.05	14.16	89.90	14.22	94.30	14.14	95.00	14.38
MFCC + CNN	74.75	8.59	73.53	8.64	78.32	8.91	81.50	8.89	88.18	8.91	91.32	9.0	95.17	9.06
MTF + ZMFA-CNN	79.88	144.35	80.64	140.11	86.30	147.62	86.89	145.77	86.19	137.56	88.23	139.44	87.9	143.26
MTF + CNN	79.42	147.23	78.95	143.16	79.48	148.73	79.44	155.78	81.37	143.95	83.63	145.32	86.52	149.81
GAFS + ZMFA-CNN	69.42	16.10	73.45	16.65	78.98	17.38	79.88	17.90	79.90	17.30	80.13	17.45	81.92	17.54
GAFS + CNN	69.60	9.44	75.62	9.43	79.68	9.46	79.55	9.48	79.80	9.49	80.28	9.30	80.63	9.98

Table 6. Partition of SUFD data set.

Data	Location	Working condition	Label	α
A	B		0 / 1	0.1 ~ 0.5
	IR & OR	Speed (1200 rpm) & Load (0 Nm)	2 / 3	
	N	/	4 / 5	
	IR	Speed (1800 rpm) & Load (7.32 Nm)	6 / 7	
	OR		8 / 9	

data types in total. As shown in table 6, according to two different working conditions, the data is divided into ten categories for fault diagnosis. In addition, the sampling frequency is 5120 Hz, and the vibration data of the three axes of X, Y, and Z are selected as the three-channel data form.

This paper combines the two working condition data sets and records them as A data set. The dataset includes ten categories of data, with 100 samples in each category. Each sample consists of X, Y, and Z channel data, and the sample size is 3×4096 . And like the CWRU data set, α is also used as the proportion of the training set to the entire data set.

4.3.2. Ablation contrast experiment. To further verify the effectiveness of each module of this algorithm, ablation experiment was conducted on the SUFD dataset. The ablation contrast methods are the same as that in the CWRU dataset. The accuracy of different methods under different training ratios and the running time at $\alpha = 0.1$ are shown in figure 9.

It can be seen from the figure that with the increase of effective training samples, the fault diagnosis accuracy is significantly improved. By comparing the results, it can be seen that the proposed HLFLA can significantly reduce the feature extraction burden of the model, and ZMFA-CNN has excellent feature extraction capabilities and generalization capabilities.

4.3.3. Noise robustness experiments. To verify the anti-noise ability of this method on different data, the following experiments were conducted. White noise is used as noise, and the SNR setting is the same as that of the CWRU data set. The comparison methods add a comparison of two neural networks NIN and ConvNext based on the CWRU data anti-noise experiment. The same experiment is set to the accuracy and time of fault diagnosis in the case of A data set and $\alpha = 0.1$. The experimental results are shown in table 7.

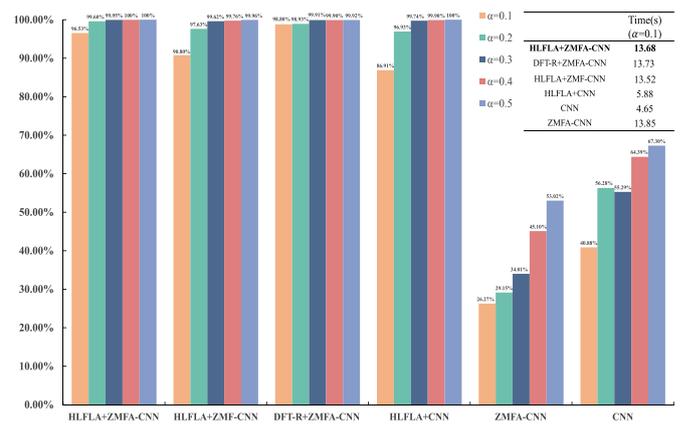


Figure 9. The results of ablation contrast experiment.

Experimental results show that the method in this paper has excellent anti-noise ability and can still achieve more than 97% Acc in a heavily noisy environment. Compared with other data preprocessing methods, this method is more suitable for small sample fault diagnosis and has stronger noise robustness.

4.4. Random noise robustness experiments

To simulate the random occurrence of noise in the actual environment, we make modifications based on Gaussian white noise. The original Gaussian white noise is applied to each position of the sample. But this paper randomize the position where the noise appears, design a parameter to control the total length of the noise superimposed on the sample signal. In addition, the intensity of the noise is set to SNR = -6, and verify the anti-noise effect of the proposed method and other methods on two data sets.

As shown in table 8, 100% in the table means that noise is superimposed on every position of the signal, and 20% means

Table 7. Accuracy and running time at different SNR on SUFD data set.

Method	Acc(%) Time(s)													
	SNR = -6		SNR = -4		SNR = -2		SNR = 0		SNR = 2		SNR = 4		SNR = 6	
HLFLA + ZMFA-CNN	98.25	13.92	97.70	13.53	98.20	14.43	98.03	14.17	98.30	14.25	99.15	14.24	99.38	14.39
HLFLA + NIN	74.21	7.59	76.56	7.91	80.58	7.86	83.43	8.01	86.29	7.67	89.55	8.26	90.64	7.99
HLFLA + ConvNext	81.83	53.93	83.54	54.77	85.88	55.32	89.12	55.87	93.35	56.34	93.35	56.34	95.95	56.81
DFT-R + ZMFA-CNN	95.90	13.89	93.60	14.13	94.78	13.51	93.32	13.97	94.90	13.78	94.03	13.93	93.55	13.97
DFT-R + NIN	72.57	7.64	72.35	7.66	72.58	7.65	79.20	7.67	84.99	7.59	87.25	7.56	91.65	7.71
DFT-R + ConvNext	76.37	54.45	75.52	56.32	79.97	56.89	80.10	56.90	88.40	57.29	91.99	57.32	93.24	57.76
GAFS + ZMFA-CNN	68.80	13.70	71.65	14.35	73.18	13.60	75.85	14.5	77.70	13.55	79.67	14.65	81.25	14.10
GAFS + NIN	62.80	7.87	63.60	8.14	66.05	8.26	65.13	8.06	67.23	7.82	70.02	7.91	71.02	8.39
GAFS + ConvNext	65.82	56.30	67.58	56.63	69.00	56.66	71.33	57.67	73.25	46.01	75.43	47.00	77.50	47.92
MTF + ZMFA-CNN	78.23	152.76	80.07	152.10	79.31	153.42	80.75	152.56	21.19	152.71	83.71	151.22	85.13	151.97
MTF + NIN	70.01	146.03	70.54	145.11	71.33	147.35	72.87	146.78	75.19	147.54	76.22	149.47	77.18	153.25
MTF + ConvNext	75.45	185.76	80.13	185.67	80.97	186.95	82.12	186.86	85.13	186.61	87.30	184.57	90.25	185.07
MFCC + ZMFA-CNN	86.60	12.17	87.63	11.79	88.78	12.46	91.28	11.38	92.88	11.75	93.98	8.38	95.68	12.10
MFCC + NIN	87.68	8.17	87.15	8.56	89.32	4.91	91.90	8.69	94.23	8.66	97.05	8.67	98.58	12.01
MFCC + ConvNext	94.80	48.65	95.83	49.15	96.95	50.99	97.65	51.96	98.47	52.05	98.32	51.98	98.75	53.34

Table 8. Accuracy comparison of different methods under random noise conditions on two data sets. The results of proposed method are shown in bold.

Method		100%	80%	60%	40%	20%
CWRU	HLFLA + ZMFA-CNN	97.50 %	97.45%	98.70%	98.77%	99.83%
	MFCC+ZMFA-CNN	83.73%	83.46%	87.35%	90.46%	92.16%
	MTF + ZMFA-CNN	79.88%	80.78%	82.97%	84.26%	85.44%
	GAFS + ZMFA-CNN	69.42%	75.88%	77.19%	84.11%	90.12%
SUFD	HLFLA + ZMFA-CNN	98.25%	98.53%	98.35%	98.18%	98.55%
	MFCC+ZMFA-CNN	86.60%	87.32%	88.22%	88.57%	90.10%
	MTF + ZMFA-CNN	78.23%	81.14%	83.49%	86.47%	85.65%
	GAFS + ZMFA-CNN	68.80%	79.32%	85.86%	87.55%	88.87%

Table 9. Information of comparison methods. SN represents the sample number and SS represents the sample size.

Comparative articles	Years	Method	SN	SS	α
Zhang <i>et al</i> [22].	2021	FDGRU	9600	(1,1,4096)	0.83
Tian <i>et al</i> [23].	2023	Wavelet-SANet	3000	(1,1,1024)	0.7
Qiao <i>et al</i> [52].	2020	TF-WConvLSTM	8000	(1,1,1200)	0.7
Xu <i>et al</i> [53].	2021	IMS-FACNN	24 000	\	0.8
Weng <i>et al</i> [54].	2023	HTNet	300	(1,1,1024)	0.8
Zhang <i>et al</i> [55].	2018	TICNN	6850	(1,1,2048)	0.96
Yang <i>et al</i> [56].	2021	RWKDCAE	1000	(1,1,1024)	0.7
Zhang <i>et al</i> [27].	2022	DCA-BiGRU	\	(1,1,1024)	0.1
\	\	HLFLA+ZMFA-CNN(Our)	1000	(3,1,4096)	0.1

that noise is superimposed on any 20% area of the signal. Experimental results show that the random occurrence of noise does not affect the anti-noise ability of this method, and even the fault diagnosis performance improves as the noise coverage area decreases.

4.5. Comparison studies of diagnostic method

To verify the superiority of this method, it is compared with existing methods under variable working conditions and noise conditions. And because the CWRU data set is very popular in

the field of fault diagnosis, it is used as an experimental data set. The comparison methods used and their basic conditions are shown in table 9. The accuracy of different methods in variable working conditions tasks and anti-noise tasks are listed in tables 10 and 11 respectively.

The comparison results show that the method in this paper has good performance in both tasks. In the variable load task, the method in this paper can show similar migration and generalization capabilities to existing methods. In the anti-noise task, the best results were achieved among the above methods. Although the compared methods all use single-channel

Table 10. Comparison of transfer generalization ability.

Method	Acc(%)						
	B → A	B → C	B → D	C → B	C → D	D → B	D → C
TICNN	\	99.92	94.16	94.27	96.8	83.82	91.62
IMS-FACNN	\	95.88	91.73	94.57	98.38	84.05	92.52
HTNet	\	99.98	99.65	98.33	99.88	96.08	99.9
DCA-BiGRU	95.3	99.21	93.24	\	\	\	\
HLFLA+ZMFA-CNN	99.67	99.89	96.88	98.66	100	96.2	99.66

Table 11. Comparison of noise immunity.

Method	Acc(%)						
	SNR = -6	SNR = -4	SNR = -2	SNR = 0	SNR = 2	SNR = 4	SNR = 6
FDGRU	\	94.86	97.97	99.19	98.79	99.67	99.51
TF-WConvLSTM	\	<87	<93	<95	<99	<99	<99
TICNN	\	82.05	96.47	98.22	99.27	99.61	99.59
Wavelet-SANet	\	96.67	99.02	99.42	99.87	99.98	99.98
IMS-FACNN	\	79.41	<90	92.92	<98	97.82	<100
DCA-BiGRU	\	90.72	92.11	94.84	96.2	98.32	98.32
RWKDAE	\	\	\	99	99.33	99.67	99.67
HLFLA + ZMFA-CNN	97.5	98.33	98.98	99.6	99.93	100	100

CWRU data, and the number of samples and sample sizes are different, this still reflects the superior performance of the proposed HLFLA + ZMFA-CNN method.

4.6. Industrial application: gas turbine bearing fault diagnosis

4.6.1. Description and division of gas turbine bearing data.

To verify the reliability of the algorithm, practical application is essential. As shown in the upper part of figure 10, this study collects power-end bearing fault signals from a running gas turbine. The sampling frequency is 8 KHz, and the collected data is six-channel data, which are the power end main bearing data of units 1 and 2. Each unit uses three vibration sensors to obtain three-channel data from the X, Y, and Z axes. This article uses the power end-bearing data of unit 1.

The data was collected from an operating gas turbine. After more than two months of operation, the status of the gas turbine bearings ranged from healthy to severely worn. After more than two months of data collection, obtained a large amount of data, which was named the BaiChuan data set. As shown in the lower part of figure 10, faults are divided into four damage levels according to RMS. The specific running time and fault classification are shown in table 12. This paper randomly intercepts 1000 samples from the BaiChuan data set, 200 for each category, and a single sample size of 3 × 4096. 100 are used for training, and the remaining 900 are used for testing, and the parameter settings are consistent with table 2.

4.6.2. Experiment results. As shown in figure 11, 100% accuracy is achieved for gas turbine main bearing fault diagnosis, and the training and testing time are also very considerable. Fault diagnosis on real data proves the reliability of the proposed algorithm to a certain extent.

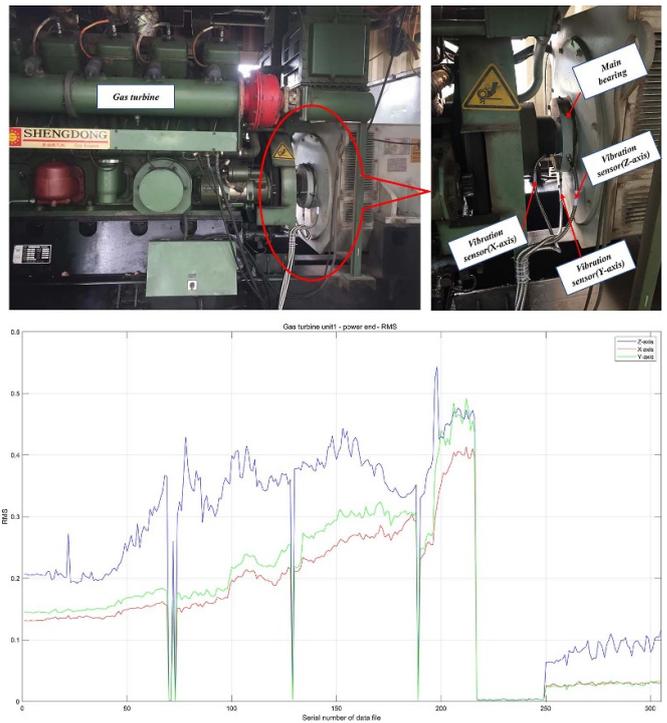


Figure 10. Gas turbine unit and its main bearing operating status.

4.6.3. Visual analysis. To show the effect of the ZMFA-CNN, the data is visualized. t-distributed stochastic neighbor embedding(t-SNE) [57] is a data visualization tool that can reduce high-dimensional data to 2–3 dimensions and then display it in the form of images. It is a nonlinear dimensionality reduction algorithm that can map high-dimensional data to low-dimensional space while preserving the local structure of the data. As shown in figure 12, t-SNE is used to visualize the original data and the data classified by the network.

Table 12. Partition of BaiChuan data set.

Data set	Fault level	Serial number of data files	Collection date	Label
BaiChuan	Fault-1	0 – 70	8.02 – 8.16	1
	Fault-2	73 – 126	8.17 – 8.27	2
	Fault-3	130 – 187	8.28 – 9.03	3
	Fault-4	189 – 216	9.09 – 9.24	4
	Normal	250 – 300	7.20 – 8.01	0

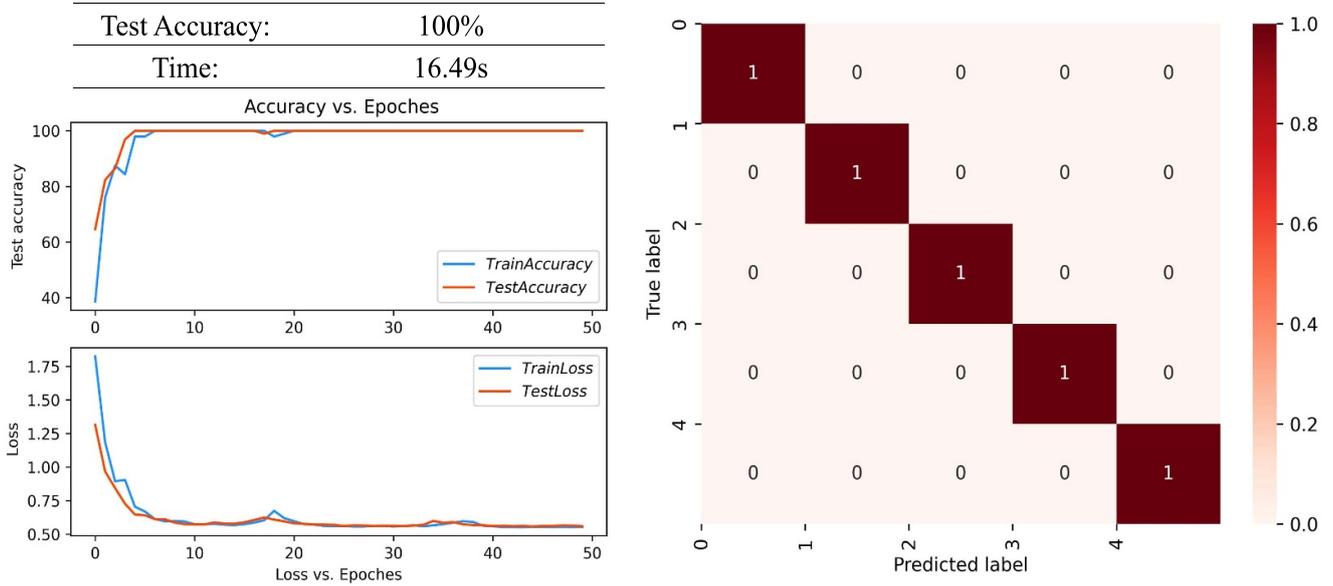


Figure 11. The experimental results of the BaiChuan data set.

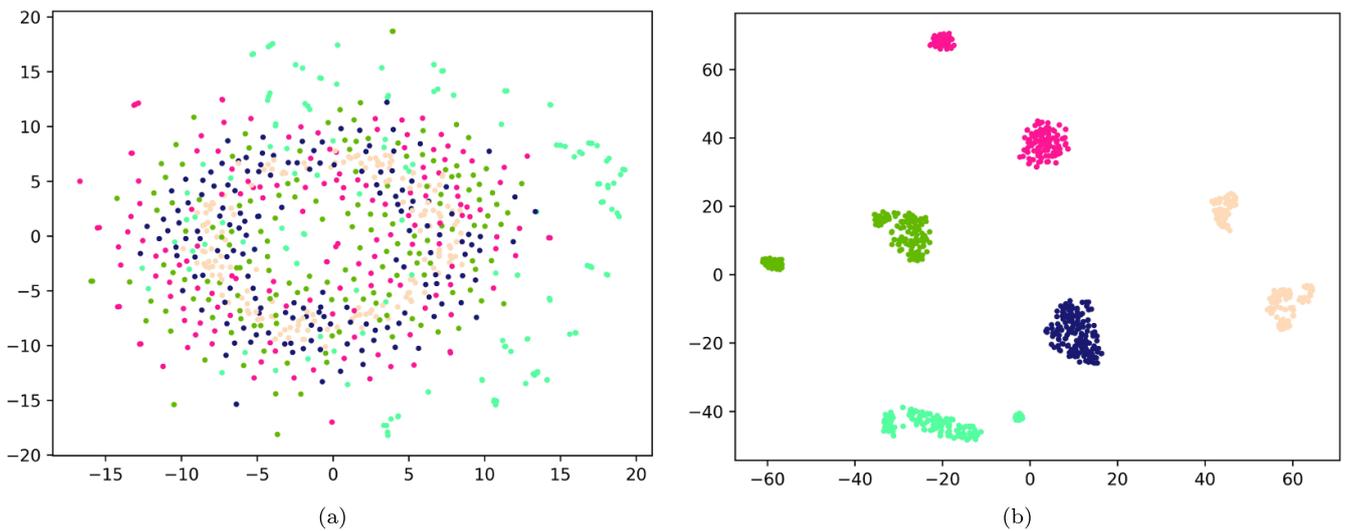


Figure 12. t-SNE visualization. (a) Raw data, (b) data after ZMFA-CNN classification.

The distinction between different categories of data after the ZMFA-CNN classification has been significantly improved.

5. Conclusion

A high-low frequency layering algorithm and a Zernike moment feature attention convolutional network is proposed

to identify the health status of equipment under small samples. HLFLA is used to reduce the difficulty of network feature extraction and noise suppression and to retain the constraint information of adjacent data. ZMFA-CNN has superior generalization capabilities and can achieve high-precision fault diagnosis under different loads. The sensitivity of HLFLA and ZMFA-CNN to the proportion of the training set is discussed through experiments, and various normalization

methods and gradient descent algorithms are explored. New technologies such as FRN, TLU, and Ranger are introduced to further improve generalization capabilities and robustness. Subsequently, the noise robustness of the algorithm was verified on the CWRU and SUFD data sets, indicating that the proposed algorithm has superior noise immunity. Variable load experiments were conducted on the CWRU data set, which proved that the algorithm also has certain migration and generalization capabilities. Finally, it was verified on a real gas turbine bearing data set, proving that the proposed HLFLA + ZMFA-CNN algorithm can achieve bearing fault diagnosis based on small samples.

It is worth noting that the proposed algorithm is difficult to transfer and generalize on multiple data sets, and different working conditions include not only load but also different rotating speeds, different bearing diameters, etc which are challenges worthy of study. In the future, research will be conducted in areas such as meta-learning, transfer learning, or large natural language models to solve more complex small-sample fault diagnosis situations, which are worthy of further research.

Data availability statement

The data cannot be made publicly available upon publication because they are owned by a third party and the terms of use prevent public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgment

This research was funded by the Foundation of the National Natural Science Foundation of China Grant Nos. 61973105, 61573130 and 52177039; The Key Technologies R&D Program of Henan Province of China. (Nos. 242102221034 and 222102220016).

Conflicts of interest

No potential conflicts of interest are reported by the authors.

ORCID iDs

Yunji Zhao  <https://orcid.org/0000-0001-6024-9105>

Jun Xu  <https://orcid.org/0000-0002-8344-2669>

References

- [1] Duan Y, Cao X, Zhao J and Xu X 2022 Health indicator construction and status assessment of rotating machinery by spatio-temporal fusion of multi-domain mixed features *Measurement* **205** 112170
- [2] Zhu H, He Z, Wei J, Wang J and Zhou H 2021 Bearing fault feature extraction and fault diagnosis method based on feature fusion *Sensors* **21** 2524
- [3] Sharma A, Bhardwaj S and Kankar P K 2019 Fault diagnosis of rolling element bearings using fractional linear prediction and AI techniques *Life Cycle Reliab. Saf. Eng.* **8** 11–19
- [4] Liu H, Liu H, Zhu C and Parker R G 2020 Effects of lubrication on gear performance: a review *Mech. Mach. Theory* **145** 103701
- [5] Zhang K, Xu Y, Liao Z, Song L and Chen P 2021 A novel fast entrogram and its applications in rolling bearing fault diagnosis *Mech. Syst. Signal Process.* **154** 107582
- [6] Xu L, Chatterton S and Pennacchi P 2021 Rolling element bearing diagnosis based on singular value decomposition and composite squared envelope spectrum *Mech. Syst. Signal Process.* **148** 107174
- [7] Wang L, Zhang X, Liu Z and Wang J 2020 Sparsity-based fractional spline wavelet denoising via overlapping group shrinkage with non-convex regularization and convex optimization for bearing fault diagnosis *Meas. Sci. Technol.* **31** 055003
- [8] Liu Y, Dan B, Yi C, Huang T and Zhang F 2023 Self-matching extraction fractional wavelet transform for mechanical equipment fault diagnosis *Meas. Sci. Technol.* **35** 035102
- [9] Chen S and Zheng X 2024 A bearing fault diagnosis method with improved symplectic geometry mode decomposition and feature selection *Meas. Sci. Technol.* **35** 046111
- [10] Shi P, Jia L, Yi S and Han D 2024 Wind turbines fault diagnosis method under variable working conditions based on AMVMD and deep discrimination transfer learning network *Meas. Sci. Technol.* **35** 046120
- [11] Mosalanejad M and Arefi M M 2018 UKF-based soft sensor design for joint estimation of chemical processes with multi-sensor information fusion and infrequent measurements *IET Sci. Meas. Technol.* **12** 755–63
- [12] Alrowaie F, Gopaluni R B and Kwok K E 2012 Fault detection and isolation in stochastic non-linear state-space models using particle filters *Control Eng. Pract.* **20** 1016–32
- [13] Alwi H and Edwards C 2014 Robust fault reconstruction for linear parameter varying systems using sliding mode observers *Int. J. Robust Nonlinear Control* **24** 1947–68
- [14] Wang Z, Yao L, Chen G and Ding J 2021 Modified multiscale weighted permutation entropy and optimized support vector machine method for rolling bearing fault diagnosis with complex signals *ISA Trans.* **114** 470–84
- [15] Hongwei F, Ceyi X, Jiateng M, Xiangang C and Xuhui Z 2023 A novel intelligent diagnosis method of rolling bearing and rotor composite faults based on vibration signal-to-image mapping and CNN-SVM *Meas. Sci. Technol.* **34** 044008
- [16] Kumar H S and Upadhyaya G 2023 Fault diagnosis of rolling element bearing using continuous wavelet transform and K-nearest neighbour *Mater. Today Proc.* **92** 56–60
- [17] Qi B, Zhang L, Liang J and Tong J 2022 Combinatorial techniques for fault diagnosis in nuclear power plants based on Bayesian neural network and simplified Bayesian network-artificial neural network *Front. Energy Res.* **10** 920194
- [18] Tang S, Yuan S and Zhu Y 2020 Convolutional neural network in intelligent fault diagnosis toward rotatory machinery *IEEE Access* **8** 86510–9
- [19] Zhao M, Zhong S, Fu X, Tang B and Pecht M 2020 Deep residual shrinkage networks for fault diagnosis *IEEE Trans. Ind. Inf.* **16** 4681–90
- [20] Bao Z, Du J, Zhang W, Wang J, Qiu T and Cao Y 2021 A transformer model-based approach to bearing fault diagnosis *Data Science* ed J Zeng, P Qin, W Jing, X Song and Z Lu vol 1451 pp 65–79
- [21] Wang X, Zhang H and Du Z 2023 Multiscale noise reduction attention network for aeroengine bearing fault diagnosis *IEEE Trans. Instrum. Meas.* **72** 1–10

- [22] Zhang Y, Zhou T, Huang X, Cao L and Zhou Q 2021 Fault diagnosis of rotating machinery based on recurrent neural networks *Measurement* **171** 108774
- [23] Tian A, Zhang Y, Ma C, Chen H, Sheng W and Zhou S 2023 Noise-robust machinery fault diagnosis based on self-attention mechanism in wavelet domain *Measurement* **207** 112327
- [24] Li X, Kong X, Zhang J, Hu Z and Shi C 2021 A study on fault diagnosis of bearing pitting under different speed condition based on an improved inception capsule network *Measurement* **181** 109656
- [25] Yao P, Wang J, Zhang F, Li W, Lv S, Jiang M and Jia L 2022 Intelligent rolling bearing imbalanced fault diagnosis based on mel-frequency cepstrum coefficient and convolutional neural networks *Measurement* **205** 112143
- [26] Diao N, Wang Z, Ma H and Yang W 2022 Fault diagnosis of rolling bearing under variable working conditions based on CWT and T-resnet *J. Vib. Eng. Technol.* **11** 3747–57
- [27] Zhang X, He C, Lu Y, Chen B, Zhu L and Zhang Li 2022 Fault diagnosis for small samples based on attention mechanism *Measurement* **187** 110242
- [28] Wang J, Shao H, Yan S and Liu B 2023 C-ECAformer: a new lightweight fault diagnosis framework towards heavy noise and small samples *Eng. Appl. Artif. Intell.* **126** 107031
- [29] Su H, Xiang L, Hu A, Xu Y and Yang X 2022 A novel method based on meta-learning for bearing fault diagnosis with small sample learning under different working conditions *Mech. Syst. Signal Process.* **169** 108765
- [30] Li G, Ao J, Hu J, Hu D, Liu Y and Huang Z 2024 Dual-source gramian angular field method and its application on fault diagnosis of drilling pump fluid end *Expert Syst. Appl.* **237** 121521
- [31] Lei C, Xue L, Jiao M, Zhang H and Shi J 2022 Rolling bearing fault diagnosis by Markov transition field and multi-dimension convolutional neural network *Meas. Sci. Technol.* **33** 114009
- [32] Zhao Y, Qin B, Zhou Y and Xu X 2023 Bearing fault diagnosis based on inverted mel-scale frequency cepstral coefficients and deformable convolution networks *Meas. Sci. Technol.* **34** 055404
- [33] Li Z, Qian Y, Wang H, Zhou X, Sheng G and Jiang X 2022 Partial discharge fault diagnosis based on Zernike moment and improved bacterial foraging optimization algorithm *Electr. Power Syst. Res.* **207** 107854
- [34] Singh S and Krishnan S 2020 Filter response normalization layer: eliminating batch dependence in the training of deep neural networks *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)* pp 11234–43
- [35] Szegedy C, Vanhoucke V, Ioffe S, Shlens J and Wojna Z 2016 Rethinking the inception architecture for computer vision *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* pp 2818–26
- [36] Liu Y, Zhang M, Zhong Z, Zeng X and Long X 2021 A comparative study of recently deep learning optimizers *Proc. SPIE* **12156** 121560F
- [37] Liu L, Jiang H, He P, Chen W, Liu X, Gao J and Han J 2019 On the variance of the adaptive learning rate and beyond *CoRR* (arXiv:1908.03265)
- [38] Zhang M, Lucas J, Ba J and Hinton G E 2019 Lookahead optimizer: K steps forward, 1 step back *Advances in Neural Information Processing Systems* vol 32, ed H Wallach, H Larochelle, A Beygelzimer, F d'Alché-Buc, E Fox and R Garnett (Curran Associates, Inc) (available at: https://proceedings.neurips.cc/paper_files/paper/2019/file/90fd4f88f588ae64038134f1eeaa023f-Paper.pdf)
- [39] Smith W A and Randall R B 2015 Rolling element bearing diagnostics using the case western reserve university data: a benchmark study *Mech. Syst. Signal Process.* **64–65** 100–31
- [40] Xu J, Sun X, Zhang Z, Zhao G and Lin J 2019 *Understanding and Improving Layer Normalization* (Curran Associates Inc.) (<https://doi.org/10.5555/3454287.3454681>)
- [41] Wu Y and He K 2020 Group normalization *Int. J. Comput. Vis.* **128** 742–55
- [42] Luo P, Zhang R, Ren J, Peng Z and Li J 2019 Switchable normalization for learning-to-normalize deep representation *IEEE Trans. Pattern Anal. Mach. Intell.* **43** 712–28
- [43] Ioffe S and Szegedy C 2015 Batch normalization: accelerating deep network training by reducing internal covariate shift *Int. Conf. on Machine Learning* vol 15 (JMLR.org) pp 448–56
- [44] Cai J, Qu L, Chen G and Yang J 2014 Application of an improved recurrent neural network in network security situation prediction *Adv. Mater. Res.* **951** 173–6
- [45] Liu Z, Mao H, Wu C-Y, Feichtenhofer C, Darrell T and Xie S 2022 A convnet for the 2020s *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)* pp 11966–76
- [46] Irfan D, Rosnelly R, Wahyuni M, Samudra J and Rangka A 2022 Perbandingan optimasi sgd, adadelta, dan adam dalam klasifikasi hydrangea menggunakan cnn *J. Sci. Soc. Res.* **5** 244
- [47] Kingma D and Ba J 2014 Adam: a method for stochastic optimization *Int. Conf. on Learning Representations* (available at: <https://api.semanticscholar.org/CorpusID:6628106>)
- [48] Xu D, Zhang S, Zhang H and Mandic D P 2021 Convergence of the RMSProp deep learning method with penalty for nonconvex optimization *Neural Netw.* **139** 17–23
- [49] Luo L, Xiong Y, Liu Y and Sun X 2019 Adaptive gradient methods with dynamic bound of learning rate *CoRR* (arXiv:1902.09843)
- [50] Mustapha A, Mohamed L and Ali K 2020 An overview of gradient descent algorithm optimization in machine learning: application in the ophthalmology field *Smart Applications and Data Analysis* ed M Hamlich, L Bellatreche, A Mondal and C Ordonez (Springer) pp 349–59
- [51] Shao S, McAleer S, Yan R and Baldi P 2019 Highly accurate machine fault diagnosis using deep transfer learning *IEEE Trans. Ind. Inf.* **15** 2446–55
- [52] Qiao M, Yan S, Tang X and Xu C 2020 Deep convolutional and lstm recurrent neural networks for rolling bearing fault diagnosis under strong noises and variable loads *IEEE Access* **8** 66257–69
- [53] Xu Z, Li C and Yang Y 2021 Fault diagnosis of rolling bearings using an improved multi-scale convolutional neural network with feature attention mechanism *ISA Trans.* **110** 379–93
- [54] Weng C, Lu B, Gu Q and Zhao X 2023 A novel hierarchical transferable network for rolling bearing fault diagnosis under variable working conditions *Nonlinear Dyn.* **111** 11315–34
- [55] Zhang W, Li C, Peng G, Chen Y and Zhang Z 2018 A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load *Mech. Syst. Signal Process.* **100** 439–53
- [56] Yang D, Karimi H R and Sun K 2021 Residual wide-kernel deep convolutional auto-encoder for intelligent rotating machinery fault diagnosis with limited samples *Neural Netw.* **141** 133–44
- [57] van der Maaten L and Hinton G 2008 Visualizing data using t-SNE *J. Mach. Learn. Res.* **9** 2579–605 (available at: <http://jmlr.org/papers/v9/vandermaaten08a.html>)