

A Bearing Fault Diagnosis Model based on Minimum Average Composite Entropy and Parallel Attention Mechanism Convolutional Neural Network

Zhen Zhang^{1,2*} , Shixi Yang^{3,2}, Jun He^{3,2}, Wanchun Zhou², Yanxu Liu²

¹Henan University of Technology, 450000, Zhengzhou, China, zhangzhen1986217@163.com

²Zhejiang University - Zhongyuan Institute, 450001, Zhengzhou, China

³School of Mechanical Engineering, Zhejiang University, 201109, Hangzhou, China

Abstract: This work addresses the issues of low diagnostic accuracy and weak generalization in rotating machinery bearing fault diagnosis, especially under complex noise conditions. In this paper, a novel bearing fault diagnosis method is proposed. This method, known as MACE + PFACNN, combines the minimum average composite entropy (MACE) with a parallel fusion attention convolutional neural network (PFACNN). In MACE, the minimum average composite entropy, which is composed of the Renyi entropy and the sample entropy, is used as a fitness function to guide the dung beetle optimization algorithm for fault feature extraction. Then, the extracted signal features are converted into angle and field and angular difference fields by Gramian angle field transformation. Finally, a PFACNN is used for fault diagnosis. Experimental data and bench tests show that the proposed model achieves a classification accuracy of 99.93 %. Compared with the baseline model, the noise resistance under complex noise conditions has improved by more than 15 %, and the generalization ability has increased by 3.68 %.

Keywords: Renyi entropy, fusion attention mechanism, parallel convolutional neural network, rolling bearings, Gramian angle field

1. INTRODUCTION

Bearings are the "joints" of rotating machinery, which often have to operate under extreme conditions such as high temperatures, high pressure, high speed, high intensity, and variable loads. Failure of these components can severely compromise the safety of the equipment during operation. Therefore, the study of bearing fault diagnosis technology is crucial to minimize economic and life losses [1], [2].

Current bearing fault diagnosis technologies can be mainly categorized into two types: methods based on signal processing and those based on artificial intelligence. Currently, scientific researchers often study one type of fault diagnosis method in isolation. However, it is important to recognize that these two approaches are not independent but rather form a unified whole. It is a challenge to effectively improve the accuracy of fault diagnosis by focusing on only one type of method. For example, artificial intelligence (AI) techniques can help with the parameter optimization of fault diagnosis methods and the classification of faults within signal processing. Conversely, signal processing methods can improve the data quality of AI techniques and thus increase the accuracy and generalization capability of AI-based fault diagnosis methods.

The fault signal feature extraction method represented by variational mode decomposition (VMD) can effectively solve the modal aliasing phenomenon in empirical mode decomposition (EMD) and has attracted the attention of most researchers. However, in VMD, the penalty parameters α and the decomposition layer k need to be specified in advance, which leads to human factors affecting the decomposition and extraction effect of VMD [3]. The use of optimization algorithms to select key parameters for VMD has become mainstream in current scientific research. The selection of the fitness function is crucial for the proper optimization of these parameters. Luo selected a series of time-domain and frequency-domain indices as the fitness function to improve the whale optimization algorithm to achieve feature extraction of the wind turbine gearbox fault signal [4]. Chang extracted the fault signal features using the center frequency and Pearson correlation coefficient of the modal function as the fitness function of the particle swarm optimization (PSO) algorithm.[5]. Liu extracted the features of the oil pulse fluid pressure signal, achieving good results [6]. Entropy, which is a measure of the uncertainty of a random variable, has attracted interest in the optimization of VMD parameters and the selection of effective modes. Song used the generalized

refined composite multiscale dispersion entropy as the fitness function for the grasshopper optimization algorithm to achieve fault signal extraction for sliding bearings [7]. Wang used the energy loss coefficient and information entropy as the fitness function to optimize the key parameters of VMD and achieved good results [8]. However, the use of single entropy as an optimization indicator for VMD parameters and as a method to select effective modal components has its limitations. For example, sample entropy evaluates the complexity of time series by quantifying the probability of generating new patterns within the signal. Information entropy, on the other hand, is commonly used to measure the information content of a system. The different centroids of the different entropies can lead to different optimization results for the VMD parameters, resulting in significant variations in the model's fault diagnosis capability [9], [10].

Compared to back propagation (BP), convolutional neural network (CNN) has the advantage of detecting the deep connections of the data, preventing overfitting and having high diagnostic accuracy, which has attracted much attention in the academic community [11]. Chao used a CNN to identify the high-speed axial plunger pump cavitation intensity, which greatly improves the identification accuracy [12]. Bao et al. used CNN to identify time-varying stress data and effectively identified the failure damage of a pipe rack of an offshore platform [13]. Jiang applied CNN to the fault diagnosis of nuclear power plant bearings. Compared with other methods, the fault diagnosis model trained with CNN has better anti-noise and generalization ability [14]. However, the above CNN-based fault diagnosis methods are based on one-dimensional images and grayscale graphs. According to research, CNN for two-dimensional image learning ability is much higher than that for one-dimensional signals, and the grayscale graph cannot clearly represent the relationship between signal amplitude and time, and it is easy to ignore the bearing fault signal in the time domain [15]. The attention mechanism (AM) continuously updates the feature weight according to the loss value obtained by network training, and gives each feature the weight corresponding to the weight, so that the training of the CNN is more effective. Han has used AM for fault diagnosis of transmission machinery, which improves the computational efficiency and diagnostic accuracy of CNN and achieves good results [16]. Xiang used AM to fault diagnose the fan fault, and the trained model can predict the abnormal state of the fan [17]. Xu applied AM to solve the problem of low performance of CNN in complex working environments, and the results show that the proposed model has good reliability in Guangdong bearing fault diagnosis [18]. It should be noted that the focus of AM itself is not the same, and there is obviously some deficiency in analyzing complex time-varying non-stationary fault signals with a single attention mechanism [19].

Based on the above analysis, an innovative feature parameter called minimum average composite entropy (MACE) is introduced in this paper, which can effectively compensate for sporadic noise and fault sensitivity. Using this parameter as a fitness function for the optimization algorithm of the important VMD parameters improves the quality of the input signals. In this paper, an innovative fault diagnosis algorithm based on a parallel fusion attention convolutional

neural network (PFACNN) is proposed. By combining with MACE, this algorithm significantly improves the accuracy and generalization ability of the fault diagnosis model under complex noise working conditions. The research framework is shown in Fig. 1.

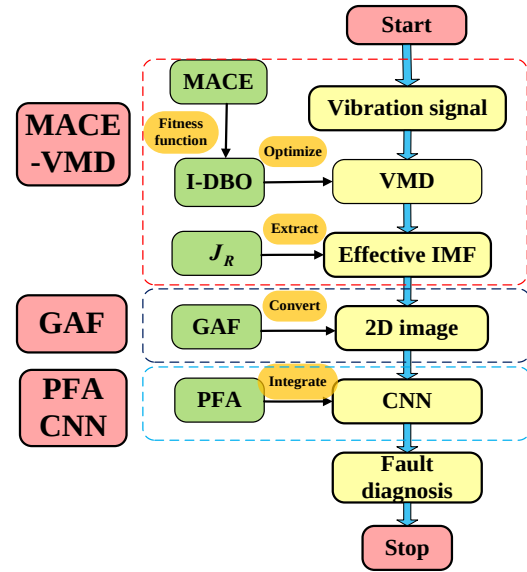


Fig. 1. Research block diagram.

2. SUBJECT & METHODS

A. Signal processing method

Minimum average composite entropy variational mode decomposition (MACE-VMD)

In actual production, fault-bearing signals often contain a large amount of noise and invalid signals. If these cannot be processed correctly, this significantly impairs the diagnostic accuracy of intelligent diagnosis algorithms. VMD can represent input signals as a set of modal functions, which is widely used in signal processing. Reference is made to the decomposition process in the literature [20]. When applying VMD, the mode number k and the penalty factor α must be set artificially. Improper setting of the parameters will significantly affect the decomposition effect of the signal [21]. As an index that can represent the characteristic parameters of the impact signal, Kurtosis (S_k) is often used to select the key parameters of VMD, but its sensitivity to occasional impact noise is too large, which often leads to incorrect selection of target parameters [22]. Skewness (S_r) is limited due to its low fault sensitivity. Therefore, this paper innovatively proposes the Renyi Entropy (R_e), which can better balance the periodic shock signal and the occasional noise signal.

When the bearing fault signal $x = \{x(1), x(2), \dots, x(n)\}$ after VMD is the time series of its k order mode $u_k = \{u_k(1), u_k(2), \dots, u_k(n)\}$, and R_e of its k order mode can be defined as:

$$R_e(k) = \frac{(1/N) \sum_{i=1}^N |u_k(i)|^3}{[(1/N) \sum_{i=1}^N |u_k(i)|]^3} \quad (1)$$

where N is the total number of samples. Using the rolling bearing inner ring fault model, investigate the sensitivity of R_e , S_r and S_k to defects and their stability to sporadic noise. Fault model of the rolling bearing inner ring:

$$x(t) = \sum_j^M a_j e^{-g\gamma(t)} \cos \left[\omega_0 \sqrt{1-g^2} (\gamma(t) - \tau_j) \right] \quad (2)$$

where a_j denotes the amplitude of the j^{th} fault impact, g denotes the attenuation coefficient of the bearing, M denotes the number of bearing impact excitations, $\gamma(t)$ denotes the pseudocycle time; τ_j denotes the time delay caused by relative sliding, and ω_0 denotes the fault feature frequency. Fig. 2 illustrates the variation of R_e , S_r and S_k with the inner ring fault defect of rolling bearings. For a better understanding, the defect variation is converted into the variation of the signal noise ratio (SNR).

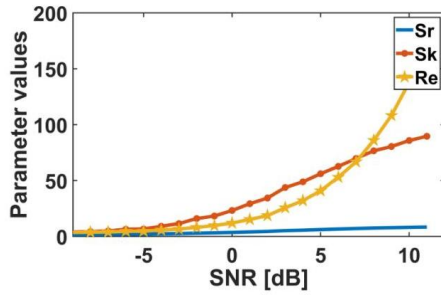


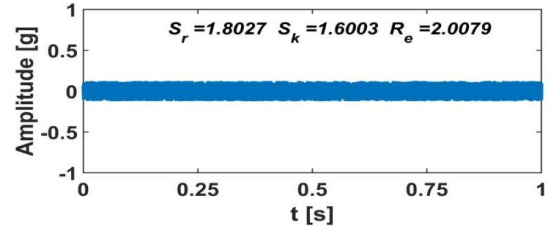
Fig. 2. S_k , S_r , R_e schematic diagram of variation with defects.

Fig. 2 shows the schematic diagram of S_k , S_r and R_e with the fault defect of the rolling bearing. The defect changes have been converted to SNR changes for better understanding.

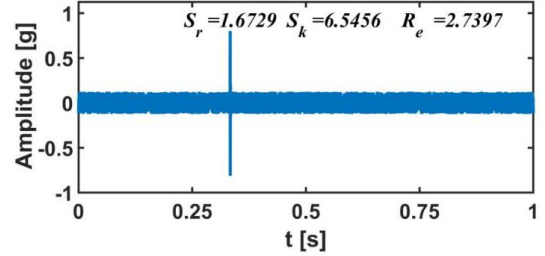
As can be seen in Fig. 2, S_r is not sensitive to the defect size, and R_e has a similar trend to S_k . It can be seen that R_e and S_k have high sensitivity to fault defect changes and can indicate bearing faults more accurately. This phenomenon can be similar to outer ring failure and rolling body failure and is not described here. As shown in Fig. 2, S_r is insensitive to defect size, while S_k and R_e exhibit similar variation trends. This indicates that S_k and R_e are highly sensitive to fault defect variation and can accurately indicate bearing faults. This phenomenon shows similar variation trends for outer race faults and rolling element faults, which will not be discussed further here. Fig. 3 illustrates the variations of S_k , S_r and R_e .

Fig. 3 shows the variation of S_k , S_r and R_e under white noise and composite noise conditions, with the amplitude normalized in the figure.

As can be seen in Fig. 3, the data show that S_r has excellent sensitivity to contingency pulse responses, while S_k shows high sensitivity to such events. In the case of a contingency interference, S_k increases more than threefold, suggesting that contingency pulses in aircraft engines significantly affect S_k . Although R_e has some sensitivity to accidental pulses, its sensitivity to incidental noise is only 18.4 % of that of S_k . Fig. 3 also shows that R_e is able to balance the sensitivity of bearing defects with accidental pulse stability.



(a) Gaussian noise S_k, S_r, R_e



(b) Gaussian noise + contingency pulse S_k, S_r, R_e

Fig. 3. S_k , S_r , R_e sensitivity to accidental noise.

E_e refers to the information entropy of the envelope signal p_j , which evaluates the sparse signal more strongly than the information entropy.

$$E_e(k) = - \sum_{j=1}^N u_k \log_2(u_k) \quad (3)$$

After normalizing R_e and E_e , the proposed compound entropy $C_e(k)$:

$$C_e(k) = \frac{\widetilde{E_e}(k)}{\widetilde{R_e}(k)} \quad (4)$$

where $\widetilde{R_e}(k)$ is the normalized R_e defined as $\widetilde{R_e}(k) = \frac{R_e(k)}{\sum_{i=1}^N R_e(i)}$, and $\widetilde{E_e}(k)$ is the normalized envelope entropy defined as $\widetilde{E_e}(k) = \frac{E_e(k)}{\sum_{i=1}^N E_e(i)}$.

Therefore, the MACE (q) proposed in this paper can be expressed as follows:

$$q = \min_{|K||\alpha|} \left\{ -\frac{1}{K} \sum_{k=1}^K C_e(k) \right\} \quad (5)$$

where K is the total number, and α is the adjustment coefficient.

Improved dung beetle algorithm (I-DBO)

DBO is an optimization algorithm that simulates the behavior of the praying mantis and offers the advantages of a smaller number of parameters and rapid global search capabilities, as described in [23]. To solve the problems of slow convergence and susceptibility to local optima, the Lévy flight strategy is incorporated into the Mantis optimization algorithm in this paper to improve its global search proficiency. The Mantis position update formula is:

$$x_i(t+1) = x_i(t) + \gamma \oplus \text{Levy}(\lambda) \quad (6)$$

where $x_i(\cdot)$ is the position function of mantis, γ is the random step; $\oplus$ is the point multiplication operation; $Levy(\lambda)$ is the search path of the compound Lévy distribution.

Composite entropy Jensen-Renyi divergence distance

After optimizing the key parameters of the VMD with I-DBO, the optimal number of parameter modes k and the penalty factor α can be determined. In order to extract the most effective modal components as the signal input for the intelligent fault diagnosis algorithm, it is necessary to screen each mode of VMD decomposition. In this paper, the idea of Jensen-Renyi divergence (JRD) distance [24], is used to define the composite entropy JRD distance $J_R(k)$.

$$J_R(k) = \frac{1}{1-\gamma} \{ [\widetilde{R}_e(k)]^\alpha + |C_c(k)|^\alpha \} \quad (7)$$

where γ is the parameter of the two probability distributions, and $C_c(k)$ is Pearson's correlation coefficient:

$$C_c(k) = \frac{\sum_{i=1}^n (u_k(i) - \bar{u}_k)(x(i) - \bar{x})}{\sqrt{\sum_{i=1}^n (u_k(i) - \bar{u}_k)^2} \sqrt{\sum_{i=1}^n (x(i) - \bar{x})^2}} \quad (8)$$

$J(k)$ with the role of flexibly regulating the complexity of the evaluation of the time series, the system is more stable when $\alpha = 1/2$. Order $J(k)$, namely $\mathbf{q} = [q(1) \ q(2) \ \dots \ q(l)]$, where $q(i) > q(i+1)$, empirically, set the threshold $M = 1$.

$$\sum_{i=1}^K q(i) > M \quad (9)$$

Fig. 4 shows the algorithm flow of the signal processing module.

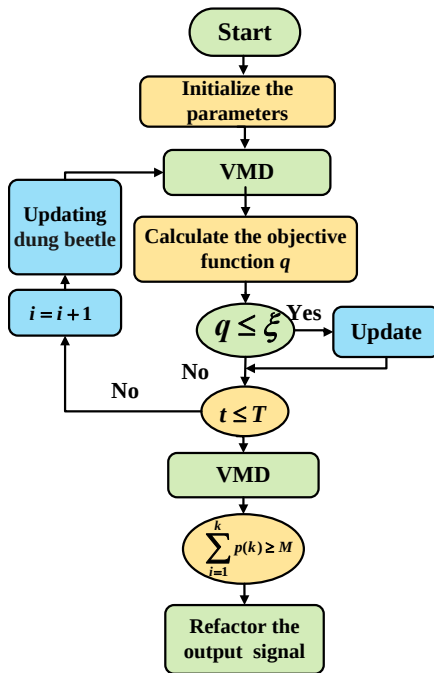


Fig. 4. Schematic diagram of the variation with defects.

B. Signal conversion method

Gramian angular field

$$x'(i) = \frac{[x(i) - \max(x)] + [x(i) - \min(x)]}{\max(x) - \min(x)} \quad (10)$$

The Gramian angular field (GAF) can convert one-dimensional waveforms into two-dimensional image mapping in the Cartesian coordinate system. Since it can well reflect the temporal characteristics of signals and has the advantage of fewer parameters to be adjusted and no need for human selection of wavelet basis, the GAF is highly appreciated by the academic community [25]. In the Gramian angular field mapping, the time series x should be scaled, the scaled time series is $x'(i)$.

From $x'(i)$ to Gramian angular field mapping, the amplitude of the signal to the cosine angle β , which takes the value range of $[0, \pi]$, and the temporal relationship to the radius r , the specific mapping relationship is as follows:

$$\begin{cases} \theta(i) = \arccos[x'(i)], & -1 \leq x'(i) \leq 1 \\ r(i) = \frac{t(i)}{N} & t(i) \in N \end{cases} \quad (11)$$

As shown in (11), the GAF can clearly express the temporal relationship of the signal. After the time series of the signal is mapped to the polar coordinates, the time series can be calculated by the angle and angle difference of each time point, so that GAF and field ($\mathbf{G}^S$) and GAF difference field ($\mathbf{G}^D$) appear. They can be expressed as follows:

$$\mathbf{G}^S = \begin{bmatrix} \cos(\alpha_{1,1}) & \cos(\alpha_{1,2}) & \dots & \cos(\alpha_{1,n}) \\ \vdots & \vdots & \ddots & \vdots \\ \cos(\alpha_{n,1}) & \cos(\alpha_{n,2}) & \dots & \cos(\alpha_{n,n}) \end{bmatrix} \quad (12)$$

$$\mathbf{G}^D = \begin{bmatrix} \sin(\beta_{1,1}) & \sin(\beta_{1,2}) & \dots & \sin(\beta_{1,3}) \\ \vdots & \vdots & \ddots & \vdots \\ \sin(\beta_{n,1}) & \sin(\beta_{n,2}) & \dots & \sin(\beta_{n,n}) \end{bmatrix}$$

where: $\alpha_{i,j} = \theta(i) + \theta(j)$, $\beta_{i,j} = \theta(i) - \theta(j)$

As shown in (12), the length n of the diagonal matrix $\mathbf{G}^S$ and $\mathbf{G}^D$ after GAF mapping can maintain the integrity and timing of the original sequence, and $\mathbf{G}^S \in \mathbb{R}^{n \times n}$, $\mathbf{G}^D \in \mathbb{R}^{n \times n}$. It can be analyzed that $\mathbf{G}^S$ and $\mathbf{G}^D$ respond to different angles of the time series $\mathbf{x}$, and there are both connections and differences between them. To ensure the integrity of learning, the two angular fields should be learned simultaneously to better extract their features.

C. Fault diagnosis method

Parallel fusion attention convolutional neural network

The $\mathbf{G}^S$ and $\mathbf{G}^D$ obtained from the GAF conversion contain a large amount of information between the signal and time, and the two-dimensional CNN can effectively mine its deep relationship with good properties. The two-dimensional CNN builds a deep convolutional network by alternating multidimensional convolution, which mainly consists of

a convolutional layer, an activation layer, a pooling layer, a fully connected layer, and a Softmax classifier. The specific operation method can be found in [26] and is not described here.

During model operation, CNN disregards the correlation between each channel and spatial information, which leads to problems such as uneven resource distribution and lower resolution accuracy. The attention mechanism, on the other hand, focuses on prominent features, helping to improve the model training speed and resolution accuracy. Currently widely used attention mechanisms include the Channel Attention Module (CAM) and the Spatial Attention Module (SAM) [27]. The CAM evaluates the importance of each feature map and assigns different weights depending on its importance. The expression for the channel attention mechanism in this paper is:

$$A^{CAM} = S \left[W_d \left(W_{mp}(X^{CAM}) \right) \right] \times X^{CAM} \quad (13)$$

where X^{CAM} is the input of the channel attention mechanism, $S(\cdot)$ is the Softmax function, W_d is the convolution operation, and W_{mp} is the average pooling operation.

Then the output Z^{CAM} of the channel attention mechanism can be expressed as:

$$Z^{CAM} = A^{CAM}(X^{CAM}) \odot^{CAM} X^{CAM} \quad (14)$$

where $\odot^{CAM}$ is the channel multiplication operation. Its structure is shown in Fig. 5.

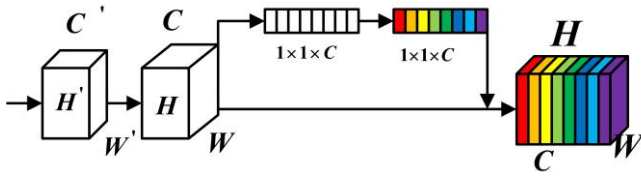


Fig. 5. Schematic diagram of the channel attention mechanism.

The spatial attention mechanism is mainly that the input spatial information is transformed into another space by the spatial transformation module and while retaining the key position information, the output of each spatial information is weighted to distinguish the importance of the position. The channel attention mechanism A^{SAM} can be expressed as follows:

$$A^{SAM} = W_q(X^{SAM}) S[\delta(W_{d1}(X^{SAM}) W_{d2}(X^{SAM}))] \quad (15)$$

where X^{SAM} is the input for the spatial attention mechanism; $S(\cdot)$ is the Softmax function; W_q is the convolutional operation; W_{d1} and W_{d2} are the fully connected neural network layers; $\delta(\cdot)$ is the tensor-by-element square root operation. The output Z^{SAM} of the same spatial attention mechanism can be expressed as:

$$Z^{SAM} = A^{SAM}(X^{SAM}) \odot^{SAM} X^{SAM} \quad (16)$$

where $\odot^{SAM}$ is a spatial multiplication operation. Its structure is shown in Fig. 6.

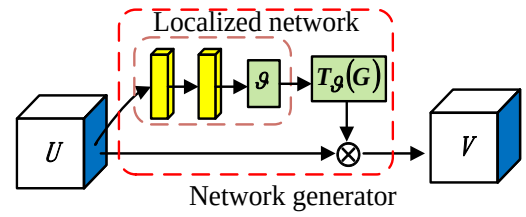


Fig. 6. Schematic diagram of the spatial attention mechanism.

Overall, CAM focuses more on the "what" and SAM focuses more on the "what position". Different information of attention can train different deep models.

3. THE MACE + PFACNN FAULT DIAGNOSIS METHOD

Based on the above analysis, this paper proposes a fault diagnosis model of CNN based on the minimum average compound entropy and the parallel fusion attention mechanism CNN. The model consists of a signal processing module, a signal conversion module and a fault diagnosis classification module. In the signal processing module, the $C_e(k)$, which is composed of the envelope entropy and Rayleigh entropy, is used as the fitness function for VMD parameter optimization to optimize the key parameters of the VMD. In the signal transformation module, the GAF, which can better reflect the signal time domain correlation, is used to transform one-dimensional time domain signals into two-dimensional image features. A PFACNN is proposed for the fault diagnosis classification module. The framework utilizes the parallel CNN algorithm and develops the channel attention mechanism and spatial attention mechanism in the parallel CNN to extract the image model more completely. The model structure is shown in Fig. 7:

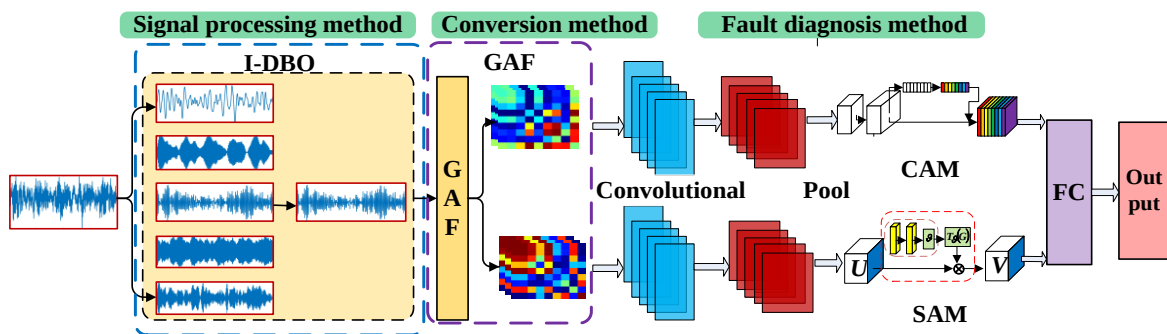


Fig. 7. MACE + PFACNN model structure diagram.

4. VERIFICATION OF EXPERIMENTAL DATA

A. MACE-VMD signal processing

The bearing fault signal data from Case Western Reserve University (CWRU) has become the standard experimental data to verify the bearing fault extraction algorithm [28]. The data signal was obtained from the vibration signal at the driving end of the experimental system with a sampling frequency of 12000 Hz. In this test, discharge processing technology is used to determine the single point damage of the rolling bearing. The bearing model is 6205-2RS JEM SKF ball bearing, and the structural parameters are shown in Table 1. Fig. 8 shows the time domain waveform of the outer ring fault. The MACE is used as the fitness function of the improved Mantis algorithm and Fig. 9 shows the fitness function change curve. The amplitude is normalized.

Table 1. Experimental parameters.

Inner diameter [mm]	Pitch diameter [mm]	Thickness [mm]	Outer diameter [mm]	Rolling diameter [mm]	Contact angle [°]
25	39	15	52	8	0

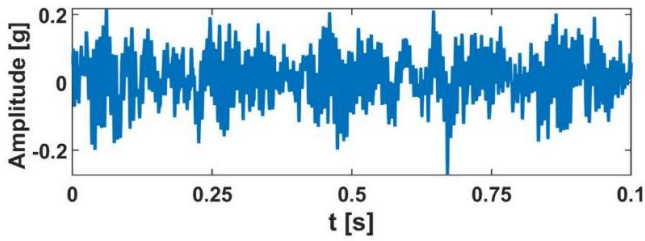


Fig. 8. Time domain waveform of outer ring signal.

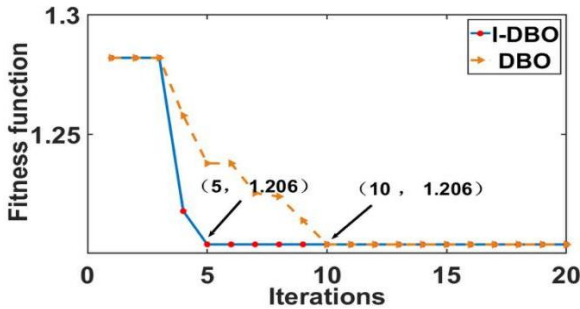


Fig. 9. Fitness function.

As can be seen in Fig. 8, the fault signal contains a lot of noise and it is difficult to recognize the fault information. As can be seen in Fig. 9, C_e gradually decreases with increasing iteration, and the optimization results of the two algorithms are basically the same. However, I-DBO can reach the minimum value at the 5th iteration, while the dung beetle algorithm needs to reach the minimum value of the fitness function at the 10th iteration. The faster convergence rate can significantly reduce the model training time and improve the efficiency. The best optimized parameter of the improved Mantis optimization algorithm is $[\alpha \ k] = [1360 \ 8]$. The VMD is decomposed into 8 modal components according to the best optimized parameters. Fig. 8 shows $\tilde{C}_c$, $\tilde{R}_e$ and J_e .

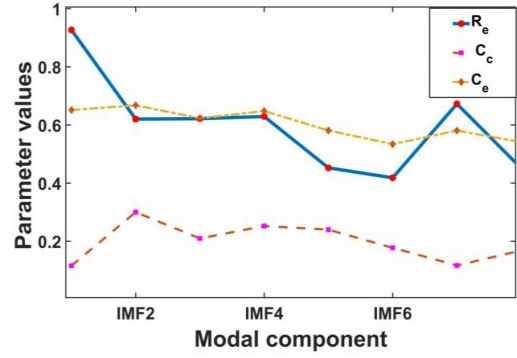
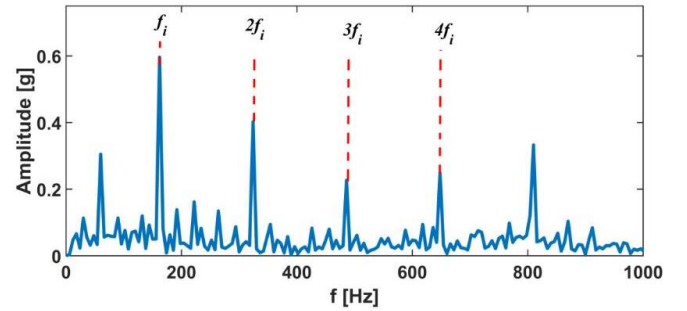
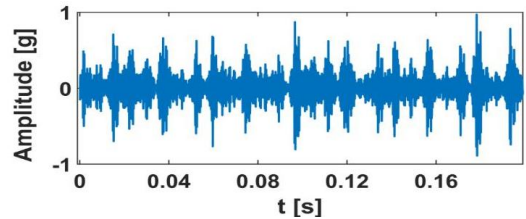


Fig. 10. Entropy change curve.

As can be seen in Fig. 10, R_e and C_c of the individual modal components do not exhibit identical variation trends. In IMF1, for example, R_e is very high, but C_c is relatively small. Therefore, it is not comprehensive to evaluate the effective mode based on R_e or C_c alone. For this reason, the evaluation index of the composite JRD distance is established in this paper. As can be seen in Fig. 10, for the composite entropy evaluation index is for IMF2 and IMF4, so IMF2 and IMF4 are selected as the effective IMF component, and the signal reconstruction. The amplitude is normalized.



(a) Frequency domain analysis results



(b) Time domain analysis results

Fig. 11. Analysis results after MACE-VMD.

As can be seen in Fig. 11(a), the reconstructed signal can clearly show the fault frequency information of the fourth order bearing inner ring. The repeated impact phenomenon shown in Fig. 11(b) is also clearer than Fig. 8. Fig. 11 proves the effectiveness of the proposed signal feature extraction method.

B. Experimental data verification of multiple fault classification

To verify the classification accuracy of the proposed fault diagnosis model, inner ring, outer ring, and rolling element faults with a size of 0.18 mm, 0.36 mm and 0.54 mm were selected, along with a set of normal bearing faults, totaling 10

datasets. The data sample length is 2048, with 100 data points selected for training for each set. The training set randomly selects 70 sets, while the test set randomly selects 30 sets. The types of datasets and the selection of training and test sets are shown in Table 2. The diagnostic results of the model are shown in Table 3.

Table 2. Dataset classification.

Fault location	Failure diameter [mm]	Tag	Dataset	
			Training set	Test set
Regular	0	1	70	30
	0.18	2	70	30
Inner ring	0.36	3	70	30
	0.54	4	70	30
Outer ring	0.18	5	70	30
	0.36	6	70	30
	0.54	7	70	30
Rolling body	0.18	8	70	30
	0.36	9	70	30
	0.54	10	70	30

Table 3. Fault diagnosis results.

Category	Accuracy [%]	Category	Accuracy [%]
1	100	6	100
2	100	7	100
3	100	8	99.3
4	100	9	99.3
5	100	10	100

As shown in Table 3, MACE + PFACNN accuracy for the 10 categories in the CRWC dataset, 9 categories achieved 100 % accuracy, recall and F1 score, and only category 8 was misclassified. The fault accuracy of the MACE + PFACNN model is 99.3 %. The high accuracy classification of multi-category fault signals shows the good classification performance of this algorithm.

C. Ablation experiments

To test the influence of the signal processing module, the signal transformation module and the fault diagnosis module on model feature classification extraction, the model was ablated by deleting or replacing modules. The classification results of the different module combinations are shown in Table 4:

Table 4. Results of ablation experiment.

Module			Model	Accuracy [%]
1	2	3		
×	×	√	A	91.2
×	√	√	B	93.7
√	×	√	C	97.4
√	√	×	D	95.3
√	√	√	E	98.9

As can be seen in Table 4, the signal feature extraction module has the largest impact on the overall diagnostic accuracy in the ablation experiments. Using the same CNN diagnostic module, the Gramian angular field transformation can improve the diagnostic accuracy by 2.5 %, while using the signal feature extraction module can increase the diagnostic accuracy by 5.2 %. The combination of the signal feature extraction module and the fault diagnosis module outperforms the other module combinations. The experiments show the advantages of the proposed fault signal feature extraction module and the rationality of the proposed model.

D. Experimental data analysis of model noise resistance

To demonstrate the advantages of this algorithm, we will use the same dataset to analyze the noise resistance performance of popular algorithms from recent years. The comparative experiments are mainly divided into two categories: one is a combination of one-dimensional signal processing module + CNN, and the other is a combination of two-dimensional signals + CNN.

The one-dimensional signal processing module + CNN uses ensemble empirical mode decomposition (EEMD) to preprocess the signal and takes the classical correlation coefficient as the basis for extracting the effective modal components. The modal components with the highest correlation coefficient are selected as input features of the CNN. Three network modules, including CNN, CNN support vector machine (CNNSVM), and CNNBIGRU [29]-[31], were selected.

Two-dimensional signal + CNN selects three types of models with higher accuracy for comparative verification: grayscale image + CNN, continuous wavelet transform + CNN, and Gramian angular field feature parameters + parallel CNN. Fig. 12 shows the experimental results of the proposed algorithm and the comparison algorithms. Table 5 shows the accuracy, recall rate, and F1 results of the five methods under 0 dB working conditions. All data is the average value after 50 Monte Carlo experiments.

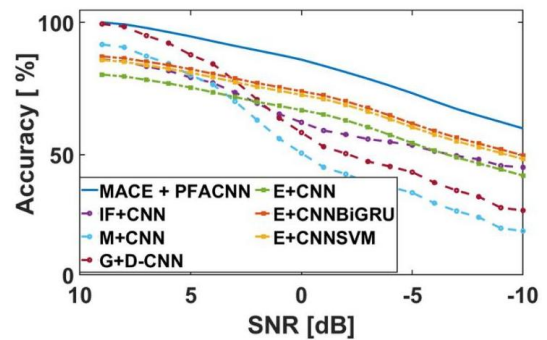


Fig. 12. Analysis results of noise resistance.

Table 5. Results of 0 dB white noise comparison test.

Model	MACE + PFACNN	IF+ CNN	M+ CNN	E+ CNN	E+ CNNSVM
Accuracy [%]	89.2	64.5	51.6	67.8	72.4
Recall rate [%]	89.8	64.9	52.1	68.2	72.9
F1 [%]	89.7	64.6	51.8	67.7	72.6

As can be seen in Fig. 12, the IF + CNN method with missing signal processing module, M + CNN and G + DCNN bearing fault diagnosis model can maintain good fault diagnosis accuracy under high signal-to-noise ratio conditions, but the low noise threshold accuracy and the fault diagnosis accuracy decreases rapidly below 3 dB. Improving the model diagnosis module can improve the base stage diagnosis accuracy, but cannot improve the fault diagnosis credibility threshold. The fault diagnosis model with the fault signal feature parameter extraction module can improve the confidence threshold of the diagnosis model, but the overall fault diagnosis accuracy is low due to the difference between one-dimensional and two-dimensional signals and the lack of fusion attention mechanism in neural network resolution. The proposed algorithm adds the fault signal feature extraction module based on the minimum average fusion entropy and the CNN fault diagnosis module of the parallel fusion attention mechanism, so that it has better diagnostic confidence threshold and diagnostic accuracy. As shown in Table 5, under the condition of 0 dB, MACE + PFACNN has the highest accuracy, precision and F1. The fault diagnosis model equipped with a signal processing module has much higher accuracy than the Missing model.

To further study the diagnostic accuracy of the complex noise model, a random periodic occasional pulse signal and white noise are added to the original signal to test the diagnostic accuracy of the algorithm and the comparison algorithm. The results are shown in Fig. 13. Table 6 shows the accuracy, recall rate, and F1 results of the five methods under 5 dB working conditions. All data is the average value after 50 Monte Carlo experiments.

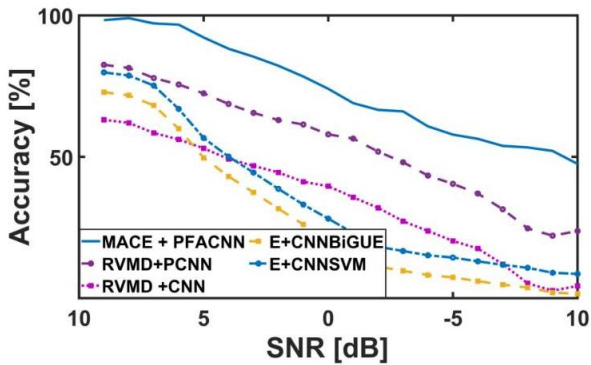


Fig. 13. Analysis results of the ability to resist complex noises.

As can be seen in Fig. 13, the diagnostic accuracy of all algorithms decreases after the addition of sporadic impact noise. The two types of algorithms that replaced the signal feature extraction module are less robust to sporadic impact noise and fail to accurately extract the signal's characteristic components. Both the diagnostic accuracy threshold and the overall accuracy have decreased significantly. In contrast, the three algorithms using the minimum average fusion entropy signal feature extraction module proposed in this paper show a more stable decrease. The MACE + PFACNN model is over 15 % higher than the comparison model. It can also be observed that the model accuracy of the missing signal transformation module is higher than that of the missing fault diagnosis module.

Table 6. Experimental results of 5 dB complex noise comparison.

Model	MACE + PFACNN	RVMD+ DCNN	RVMD+ CNN	E+CNN BiGRU	E+CNN SVM
Accuracy [%]	91.3	80.1	52.1	54.2	48.3
Recall rate [%]	91.6	80.5	52.7	54.9	48.8
F1 [%]	91.4	80.3	52.5	54.6	48.6

As shown in Table 6, under the condition of 5 dB complex noise, the fault accuracy, recall rate and F1 of the MACE + PFACNN algorithm are much higher than those of the comparison algorithm. Table 6 demonstrates the great advantages of the algorithm in this paper under the condition of complex noise.

E. Experimental data analysis of the model generalization ability

In actual production, it is unavoidable that bearings operate under variable load conditions. To verify the generalization ability of this model under variable load conditions, this paper uses bearing data from Xichang University for validation, selecting three bearing types with capacities ranging from 1 to 3 HP. Both the model in this paper and the comparison model perform 50 Monte Carlo simulations, taking the average of the training results. The analysis results of the model in this paper are shown in Fig. 14, and the experimental results are shown in Table 7.

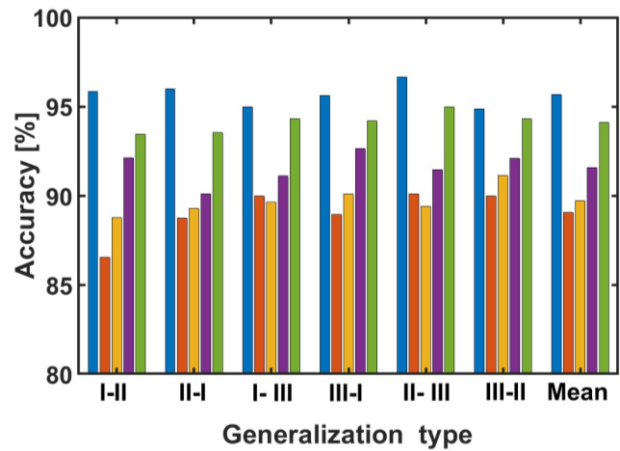


Fig. 14. Analysis results of the generalization ability.

Table 7. Generalization experiment results – Accuracy [%].

Model	MACE + PFACNN	E+ CNN	E+ SVM	IF+ CNN	M+ DCNN
I-II	95.8	86.5	88.7	92.1	93.4
II-I	96.0	88.7	89.3	90.1	93.5
I-III	94.9	89.9	89.6	91.1	94.3
III-I	95.6	88.9	90.1	92.6	94.2
II-III	96.6	90.1	89.4	91.4	94.9
III-II	94.8	90.0	91.1	92.1	94.3
Mean	95.6	89.0	89.7	91.5	94.1

In Fig. 14, I-II indicates that the model was trained with I as the source domain and tested for its generalization ability with II as the target domain. Table 7 shows the accuracy [%] of each method. As can be seen in Fig. 14 and Table 7, the average accuracy of E + CNN and E + SVM is below 90 %. The average accuracy of IF + CNN and M + DCNN is slightly higher than that of the former two. The proposed MACE + PFACNN model has an accuracy of 95.67 %, which is 1.53 % higher than the highest M + DCNN model among the comparison algorithms. Fig. 14 and Table 7 show that the MACE + PFACNN model has good generalization ability.

5. BENCH TEST VERIFICATION

A. Experimental verification of fault classification bench

The test equipment consists of the SpectraQuest rotating mechanical fault test bench, dynamic acceleration sensors from the Yangzhou Branch, and Bentley displacement sensors in conjunction with the LMS data acquisition equipment. The experimental speed is set to 2700 rpm, and MBER-12 K is used as the fault bearing model. Composite faults, which include ball, inner ring, and outer ring faults, are detected. The speed is kept constant at 2700 rpm and the LMS SCADAS mobile data acquisition system is used to acquire the vibration signals from the acceleration sensor at a sampling frequency of 12.8 kHz. Fig. 15 shows the layout of the experiment. Various faults are simulated in this experiment, including roller faults, inner ring faults, outer ring faults, and their combinations. Table 8 is the classification table of the data set. Fig. 16 is the confusion diagram and classification visualization diagram of MACE+PFACNN fault diagnosis.

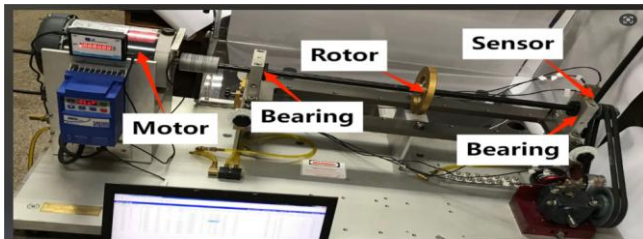


Fig. 15. Experimental layout.

1. Motor, 2. Coupling, 3. Acceleration sensor, 4. Bearing housing I, 5. Spindle, 6. Rotor, 7. Acceleration sensor, 8. Bearing housing II, 9. Bearing I, 10. Bearing II.

As can be seen in Fig. 16 and Fig. 17, the MACE + PFACNN algorithm showed high classification accuracy and its classification accuracy for the five categories reached 100 %. More importantly, the classification accuracy is still 100 %.

Table 8. Dataset classification.

Fault location	Failure diameter [mm]	Tag	Dataset	
			Training set	Test set
Regular	—	1	70	30
Inner ring	—	2	70	30
Outer ring	90°	3A	70	30
	135°	3B	70	30
Regular	—	4	70	30
Compound failure	outer 90°	5A	70	30
	outer 135°	5B	70	30

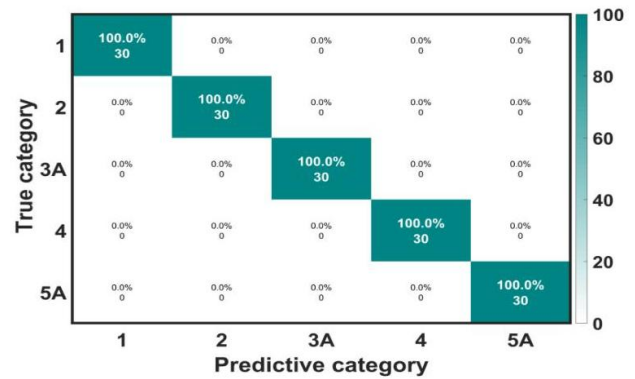


Fig. 16. Fault diagnosis results.

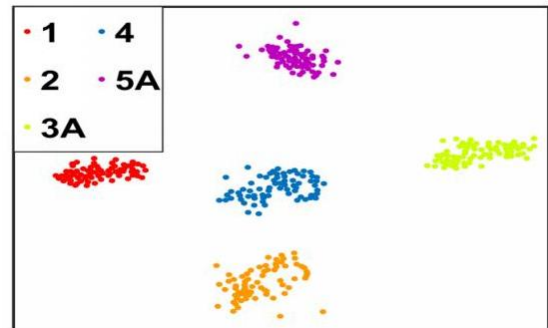


Fig. 17. Visualization results after diagnosis.

B. Generalization ability bench test verification

To further investigate the generalization ability of the bench test model, we obtained two single-fault test datasets with generalized abilities by changing the fault location in the outer ring: Dataset 3A and Dataset 3B, and two composite-fault test datasets with generalized abilities: Dataset 5A and Dataset 5B. Both our method and the comparison method used Monte Carlo simulations for 50 trials, and the average of the training results was taken as the final result. Fig. 18 shows four sets of generalized ability test data, with one operating condition as the source domain and another as the target domain. Specifically, Dataset 3A-3B indicates that Dataset 3A is the source data and Dataset 3B is the target data, with other numbered datasets following similar conventions. The generalization ability test results are shown in Table 9. Table 9 shows the accuracy [%] of each method.

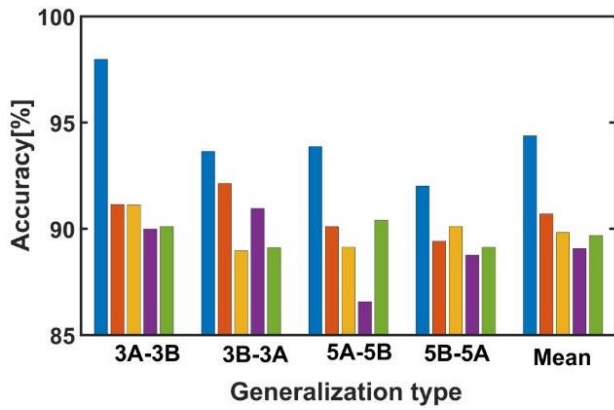


Fig. 18. Analysis results of generalization ability.

Table 9. Generalization experiment results – Accuracy [%].

Model	MACE + PFACNN	E+ CNN	E+ SVM	IF+ CNN	M+ DCNN
3A-3B	97.98	91.14	91.12	89.99	90.11
3B-3A	93.64	92.13	88.96	90.96	89.11
5A-5B	93.87	90.11	89.13	86.57	90.40
5B-5A	92.01	89.41	90.11	88.76	89.13
Mean	94.37	90.69	89.83	89.07	89.68

As shown in Fig. 18 and Table 9, the average classification accuracy of the four comparison models is about 90 %, while the accuracy of the model proposed in this paper is 94.37 %, which is 3.68 % higher than that of the best-performing E + CNN model among the comparison algorithms. Moreover, the classification accuracy of single faults in this model is higher than that of composite faults. Fig. 18 and Table 9 show that the model proposed in this paper has good generalization abilities.

6. CONCLUSION AND PROSPECTS

A. Conclusion

Given the low classification accuracy, low noise resistance, and weak generalization of intelligent diagnostic models for rotating machinery bearings, this paper focuses on the comprehensive optimization of the fault diagnosis models. The paper proposes an innovative fault signal feature extraction module based on MACE. This algorithm uses minimum composite entropy as a fitness function for the VMD key parameter optimization algorithm, which improves the quality of fault data. A novel PFACNN is also presented, which improves the accuracy of fault diagnosis. The main conclusions of this paper are:

Experimental data and bench tests show that the proposed MACE + PFACNN bearing fault diagnosis model achieves excellent classification accuracy with rates of 99.3 % and 100 %, respectively.

The fault signal processing improvement module based on MACE possesses high fault sensitivity and robustness to complex noise and effectively extracts fault signal features. It has a significant impact on the classification accuracy of the entire model.

Compared with the comparison model, the MACE + PFACNN bearing fault diagnosis model exhibits superior noise resistance, especially in the presence of composite noise, with resistance much higher than that of the comparison model.

In addition, the MACE + PFACNN bearing fault diagnosis model shows robust generalization ability under variable load conditions and variable fault locations. The classification accuracy for these two conditions is 1.53 % and 3.68 % higher than that of the best-performing comparison model, respectively.

B. Prospects

The discussion of the generalization ability of the fault diagnosis model proposed in this paper refers only to the generalization ability of faults in different positions. The generalization ability at variable speed has not yet been investigated. Further research on the generalization ability of the fault diagnosis model for variable speed will be conducted. It should be a good method to convert the variable speed problem into the constant speed problem by using appropriate signal processing methods, e.g., tacholless order tracking (TLOT).

The fault diagnosis model proposed in this paper has good accuracy only for fault classification, but it has not addressed the solution of bearing fault degree. Further research on the fault diagnosis model will be conducted in the future to evaluate the fault degree.

REFERENCES

- [1] Zhang, T., Xu, F., Jia, M. (2022). A centrifugal fan blade damage identification method based on the multi-level fusion of vibro-acoustic signals and CNN. *Measurement*, 199, 111475. <https://doi.org/10.1016/j.measurement.2022.111475>
- [2] Wang, X., Mao, D., Li, X. (2021). Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network. *Measurement*, 173 (6), 108518. <https://doi.org/10.1016/j.measurement.2020.108518>
- [3] Xu, Z., Ye, T., Chen, L., Gao, Y., Chen, Z. (2025). Health-radar: Noncontact multitarget heart rate variability detection using FMCW radar. *IEEE Sensors Journal*, 25 (1), 405-418. <https://doi.org/10.1109/JSEN.2024.3494755>
- [4] Luo, J., Chen, Y., Huang, Q., Zhang, S., Zhang, X. (2023). Joint application of VMD and IWOA-PNN for gearbox fault classification via current signal. *IEEE Sensors Journal*, 23 (12), 13155-13164. <https://doi.org/10.1109/JSEN.2023.3269594>
- [5] Chang, Y., Bao, G., Cheng, S., He, T., Yang, Q. (2021). Improved VMD-KFCM algorithm for the fault diagnosis of rolling bearing vibration signals. *IET Signal Processing*. <https://doi.org/10.1049/sil2.12026>
- [6] Liu, G., Chen, B. (2021). The pressure characteristics analysis of oil pulsation flow based on VMD. *Scientific Reports*, 11, 17390. <https://doi.org/10.1038/s41598-021-96860-0>

- [7] Song, M., Wang, J., Zhao, H., Li, Y. (2022). Bearing failure of reciprocating compressor sub-health recognition based on CAGOA-VMD and GRCMDE. *Advances in Mechanical Engineering*. <https://doi.org/10.1177/16878132221082975>
- [8] Wang, Y., Zhang, Q., Cao, R., Zhang, S., Li, S., Xu, D. (2023). The single-channel blind source separation based on VMD and Tukey M estimation for rolling bearing composite fault diagnosis. *Measurement & Control*, 56 (9-10). <https://doi.org/10.1177/00202940231174405>
- [9] Jiang, X., Wang, J., Shi, J., Shen, C., Huang, W., Zhu, Z. (2019). A coarse-to-fine decomposing strategy of VMD for extraction of weak repetitive transients in fault diagnosis of rotating machines. *Mechanical Systems and Signal Processing*, 116, 668-692. <https://doi.org/10.1016/j.ymssp.2018.07.014>
- [10] Song, D., Feng, G., Qi, T., Wang, H., Pan, D., Zhang, L. (2025). A new combined transient electromagnetic noise reduction method of VMD-SVD-WTD based on improved dung beetle optimization algorithm with multi-strategy fusion. *Measurement Science and Technology*, 6 (1), 016105. <https://doi.org/10.1088/1361-6501/ad7e42>
- [11] Glowacz, A., Sulowicz, M., Kozik, J., Piech, K., Glowacz, W., Li, Z., Brumercik, F., Gutten, M., Korenciak, D., Kumar, A., Lucas, G. B., Irfan, M., Caesarendra, W., Liu, H. (2024). Fault diagnosis of electrical faults of three-phase induction motors using acoustic analysis. *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 72 (1), e148440. <https://doi.org/10.24425/bpasts.2024.148440>
- [12] Chao, Q., Tao, J., Wei, X., Wang, Y., Meng, L., Liu, C. (2020). Cavitation intensity recognition for high-speed axial piston pumps using 1-D convolutional neural networks with multi-channel inputs of vibration signals. *Alexandria Engineering Journal*, 59 (6), 4463-4473. <https://doi.org/10.1016/j.aej.2020.07.052>
- [13] Bao, X., Fan, T., Shi, C., Yang, G. (2021). One-dimensional convolutional neural network for damage detection of jacket-type offshore platforms. *Ocean Engineering*, 219, 108293. <https://doi.org/10.1016/j.oceaneng.2020.108293>
- [14] Zeng, D., Wu, D., Huang, H., Peng, B., Wei, Y., Du, H., Zhang, P., Shi, H., Lu, Q., Cai, X. (2025). Online identification of laser welding penetration through multi-photoelectric decomposition-reconstruction and shifted-windows-based transformer deep learning framework. *Measurement*, 247, 116872. <https://doi.org/10.1016/j.measurement.2025.116872>
- [15] Wang, L., Deng, X., Ge, P., Dong, C., Bethel, B. J., Yang, L., Xia, J. (2022). CNN-BiLSTM-attention model in forecasting wave height over south-east China seas. *Computers, Materials & Continua*, 73 (1), 2151-2168. <https://doi.org/10.32604/cmc.2022.027415>
- [16] Han, R., Wang, J., Wan, Y., Bao, J., Jiang, X., Zhang, Z., Han, B., Ji, S. (2024). Attention mechanism guided sparse filtering for mechanical intelligent fault diagnosis under variable speed condition. *Measurement Science and Technology*, 35, 042001. <https://doi.org/10.1088/1361-6501/ad197a>
- [17] Xiang, L., Wang, P., Yang, X., Hu, A., Su, H. (2021). Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism. *Measurement*, 175 (8), 109094. <https://doi.org/10.1016/j.measurement.2021.109094>
- [18] Xu, Z., Li, C., Yang, Y. (2021). Fault diagnosis of rolling bearings using an improved multi-scale convolutional neural network with feature attention mechanism. *ISA Transactions*, 110, 379-393. <https://doi.org/10.1016/j.isatra.2020.10.054>
- [19] Zhang, W., Yang, J., Bo, X., Yang, Z. (2024). A dual attention mechanism network with self-attention and frequency channel attention for intelligent diagnosis of multiple rolling bearing fault types. *Measurement Science and Technology*, 35 (3), 036112. <https://doi.org/10.1088/1361-6501/ad1811>
- [20] Yang, A., Xie, B., Liu, Y., Wang, L., Li, J. (2024). A network security situation prediction for consumer data in the internet of things using variational mode decomposition (VMD) and fused CNN-BiLSTM-attention. *IEEE Transactions on Consumer Electronics*, 70 (1), 1122-1133. <https://doi.org/10.1109/TCE.2023.3323546>
- [21] Ding, M., Shi, Z., Du, B., Wang, H., Han, L. (2021). A signal de-noising method for a MEMS gyroscope based on improved VMD-WTD. *Measurement Science and Technology*, 32 (9), 095112. <https://doi.org/10.1088/1361-6501/abfe33>
- [22] Zhang, Z., Liu, B., Liu, Y., Zhang, H. (2022). Fault feature-extraction method of aviation bearing based on maximum correlation Renyi entropy and phase-space reconstruction technology. *Entropy*, 24 (10), 1459. <https://doi.org/10.3390/e24101459>
- [23] Xue, J., Shen, B. (2023). Dung beetle optimizer: A new meta-heuristic algorithm for global optimization. *The Journal of Supercomputing*, 79, 7305-7336. <https://doi.org/10.1007/s11227-022-04959-6>
- [24] Gholami, R., Hodtani, G. A. (2020). A more general information theoretic study of wireless location verification system. *IEEE Transactions on Vehicular Technology*, 69 (9), 9938-9950. <https://doi.org/10.1109/TVT.2020.3003806>
- [25] Li, Y., Zhang, H., Ma, S., Cheng, G., Yao, Q., Zuo, C. (2024). A novel method based on stepwise variational modal decomposition and gramian angular difference field for bearing health monitoring. *Arabian Journal for Science and Engineering*, 49 (11), 15773-15786. <https://doi.org/10.1007/s13369-024-09320-y>
- [26] Qi, L., Zhang, Q., Xie, Y., Zhang, J., Ke, J. (2024). Research on wind turbine fault detection based on CNN-LSTM. *Energies*, 17 (17), 4497. <https://doi.org/10.3390/en17174497>

- [27] Qu, D., Wang, S., Liu, H., Meng, Y. (2022). A car-following model based on trajectory data for connected and automated vehicles to predict trajectory of human-driven vehicles. *Sustainability*, 14 (12), 7045. <https://doi.org/10.3390/su14127045>
- [28] Case Western Reserve University. *Bearing Data Center*.
- [29] Muthulakshmi, V., Shajin, F. H., Rose, J. D., Rajesh, P. (2024). Generative adversarial networks classifier optimized with water strider algorithm for fake tweets detection. *IETE Journal of Research*, 70 (2), 1556-1571. <https://doi.org/10.1080/03772063.2023.2172466>
- [30] Wang, X., Liu, W. Y., Zhang, Y., Gu, H. (2020). A novel wind turbine fault diagnosis method based on generator current analysis. *Journal of Vibroengineering*, 22 (8). <https://doi.org/10.21595/jve.2020.21336>
- [31] Yao, Z., Zhang, T., Wang, Q., Zhao, Y. (2022). Short-term power load forecasting of integrated energy system based on attention-CNN-DBILSTM. *Mathematical Problems in Engineering*. <https://doi.org/10.1155/2022/1075698>

Received February 28, 2025

Accepted May 27, 2025