Dengxue Cao ✉
Yu Gu
Wei Lin

https://doi.org/10.21278/TOF.473048022
ISSN 1333-1124
eISSN 1849-1391

FAULT DIAGNOSIS BASED ON OPTIMIZED WAVELET PACKET TRANSFORM AND TIME DOMAIN CONVOLUTION NETWORK

Summary

In the past, the method of combining signal processing and neural network was widely used in the fault diagnosis of mechanical rolling bearings to achieve the purpose of fault signal detection and identification. However, traditional methods cannot fully extract fault features. In this paper, taking mechanical rolling bearings as an example, a new rolling bearing fault diagnosis model is proposed by combining the improved wavelet packet transform with time domain convolution network. The improved Northern Goshawk algorithm is used to optimize the wavelet packet transform and analyse the fault signal; then, the time domain convolution network is used to establish a fault diagnosis model. When compared with the wavelet packet decomposition model optimized by the improved Northern Goshawk algorithm, the accuracy of the improved wavelet packet transform model in detecting inner ring faults and rolling element faults is found to be increased by 8.3% and 7.4%, respectively.

Key words: *Wavelet packet transform, Fault diagnosis, Time domain convolution network, Rolling bearings*

1. Introduction

In the era of continuous advancement in science and technology, intelligent equipment has become the main force behind factory production. Mechanical bearings are one of the essential parts of most intelligent equipment, which directly affects the stable and safe operation of the equipment [1]. When the motor bearing fails, a light failure will delay the production progress, and a serious one will endanger the life and safety of the operator. In order to prevent production efficiency from being affected by bearing failure and also to protect the safety of operators, more and more researchers have begun to focus on the fault identification of bearings [2]. When the equipment is in operation, the bearing will run at a high speed, and a series of nonlinear signals generated by the bearing operation will contain a lot of information. The failure of mechanical bearings mainly includes three aspects: inner ring failure, outer ring failure, and rolling element failure [3]. If you want to know about the service life of the bearing, you must process the vibration signal generated when the bearing is in operation, and signal processing is an essential process. The quality of signal processing determines whether the fault can be found and checked on time [4]. However, the signal may be mixed with various complex characteristic signals, and this is particularly difficult to deal with.

Signal processing has always been a topic discussed by many experts and scholars, and there are many traditional signal processing methods. Early signal processing methods [5], such as the Fourier transform, are not ideal for dealing with nonlinear signals. Later, researchers deduced wavelet transform on the basis of short-time Fourier transform. Dybała and Zimroz [6] proposed an empirical wavelet transform method to identify bearing faults, and they achieved better results. However, the wavelet transform cannot handle details. Therefore, on the basis of wavelet transform, wavelet packet transform is proposed to optimize detail processing. In the studies [7] and [8], the wavelet packet transform method is used to process bearing fault signals. In order to extract additional signal features, experts and scholars began to add optimization algorithms to signal decomposition. In the study [9], the particle swarm optimization algorithm is used to improve the accuracy of fault identification.

In further research, many researchers began to use machine learning methods for bearing fault diagnosis [10]. Thus, the convolutional neural network is used in the study [11], the back propagation neural network in the study [12], and the deep convolutional neural network and random forest ensemble learning in the study [13]. However, fault diagnosis methods combined with neural networks are easily affected by various factors. Insufficient fault feature extraction will greatly reduce the recognition accuracy. There are many human factors in the feature extraction of the traditional fault signal decomposition method. The number of decomposition layers is too small and too many layers often lead to insufficient feature extraction, which may cause a waste of resources. Therefore, it is particularly important to optimize the signal decomposition method [14]. The self-group algorithm is used to optimize the signal decomposition and find the optimal parameters. However, traditional optimization algorithms have problems such as slow convergence and are easy to fall into local optima. Aiming at this problem, an improved Northern Goshawk algorithm is proposed to optimize the wavelet packet transform. This algorithm is used to improve the convergence speed of the process of wavelet packet transform and solve the problem of local optima. At the same time, the algorithm finds the best decomposition layer number K through optimization and fully extracts the fault feature information contained in the signal.

2. Optimized wavelet packet variation

2.1 Basic wavelet theory

Jean Morlet, a French geophysicist, proposed a new signal processing method, the wavelet analysis, on the basis of the short-time Fourier transform. However, the wavelet transform method ignores the processing of the detail signal. In the fault diagnosis of a rolling bearing, the detail signal also contains a lot of fault information. Therefore, in order to solve the problem of detail signal processing, the concept of wavelet packet transform is proposed on the basis of wavelet transform. The wavelet transform decomposes the signal at low frequencies, while the wavelet packet transform also decomposes the high-frequency signal. Figure 1 shows the flow chart of the wavelet packet transform.

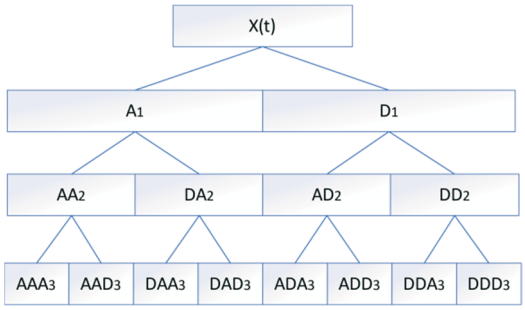

Fig. 1 Flow chart of the wavelet packet transform.

In the wavelet transform, the scale parameter a in the function $\Psi_{a,b}(t)$ changes the window size, and b adjusts the position.

$$\Psi_{a,b}(t) = |a|^{-1/2}\ \Psi(\frac{t-b}{a}) \tag{1}$$

In the formula (1), $\Psi_{a,b}(t)$ is the wavelet function, and $\Psi(t)$ is the basic wavelet. As a basic wavelet, $\Psi(t)$ must satisfy:

$$\int_{-\infty}^{\infty} |\Phi(w)|^2 |w|^{-1}\, dw < \infty\,, \tag{2}$$

where $\Phi(w)$ is the Fourier transform of the basic wavelet function $\Psi(t)$.

In the wavelet packet transform, the scale function $\phi_{(t)}$ and the wavelet function $\Psi(t)$ satisfy the dual scale equation:

$$\phi(t) = \sqrt{2}\sum_k h_k \phi(2t-k) \tag{3}$$

$$\Psi(t) = \sqrt{2}\sum_k g_k \phi(2t-k) \tag{4}$$

In the formula (4), g_k is the low-pass filter coefficient, h_k is the high-pass filter coefficient; g_k and h_k are orthogonal to each other.

The wavelet packet decomposition equation is expressed as:

$$d_I^{j,2n} = \sum_k h_{k-2L} d_k^{j+1,n} \tag{5}$$

$$d_I^{j,2n+1} = \sum_k g_{k-2L} d_k^{j+1,n} \tag{6}$$

In the formula (6), g_{k-2L} and h_{k-2L} are the filter coefficients.

2.2 Improved Northern Goshawk optimization algorithm

a) Basic principle

The Northern Goshawk optimization algorithm was first proposed by Mohammad Dehghan et al. in 2022 [15]. The northern goshawk is a member of the eagle family and is good at preying on a variety of prey, as shown in Figure 2. The Northern Goshawk optimization algorithm is better than the traditional particle swarm optimization algorithm in solving the problem, while the optimization algorithm and genetic algorithm are faster and more accurate. The Northern Goshawk optimization algorithm mainly includes two stages, namely the global search and local search.

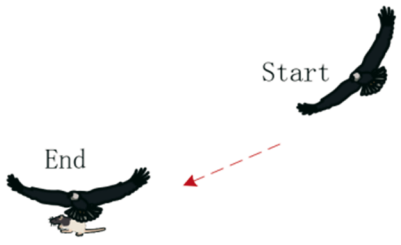

Fig. 2 Northern goshawk prey map

The population of goshawks is represented by a matrix as:

$$X=\begin{bmatrix} X_{1,1} & \cdots & X_{1,j} & \cdots & X_{1,n} \\ \vdots & . & . & . & \vdots \\ X_{i,1} & . & X_{i,j} & . & X_{i,n} \\ \vdots & . & . & . & \vdots \\ X_{n,1} & \cdots & X_{n,j} & \cdots & X_{n,n} \end{bmatrix}_{n*n}, \tag{7}$$

where X1 consists of 1 row and n columns, as follows: $X_k = [x_{k1}, x_{k2} \cdots x_{kn}]$

In the formula (7), X_k represents the position of the k-th goshawk, and X_{kn} represents the position of the k-th goshawk in the n-th dimension.

In solving the problem, the specific expression of the objective function of the northern goshawk is as follows:

$$F=\begin{bmatrix} F(X_1) \\ F(X_2) \\ \cdots \\ F(X_n) \end{bmatrix}_{n*1} \tag{8}$$

In the formula above, F represents the objective function of the goshawk population, and F_n represents the objective function of the n-th goshawk.

I. Global search

In the global search process, any northern goshawk has randomness in the selection of targets. In order to determine the best hunting range, the northern goshawk conducts a global search. The formula is as follows:

$$U_i = X_n, i = 1,2,3...,N,\ n = 1,2,3,...,N. \tag{9}$$

In the formula (9), U_i is the position of the i-th northern goshawk.

$$X_{i,j}^{NEW,U1}=\begin{cases} X_{i,j}+r(U_{i,j}-IX_{i,j}),\ F_{Ui}<F_i \\ X_{i,j}+r(X_{i,j}-U_{i,j}),\ F_{Ui}\geq F \end{cases} \tag{10}$$

In the formula above, F_{Ui} represents the target value of the i-th goshawk, $X_{i,j}^{NEW,U1}$ is the position update of the i-th goshawk in the j-th dimension, the parameter r represents [0,1], while the parameter i is a random integer of 1 or 2.

$$X_i=\begin{cases} X_i^{NEW,U1}, F_i^{NEW,U1}<F_i \\ X_i, F_i^{NEW,U1}\geq F_i \end{cases} \tag{11}$$

In the formula above, $F_i^{NEW,U1}$ is the updated target value of the i-th goshawk in the first stage, and $X_i^{NEW,U1}$ represents the latest spatial position of the i-th goshawk.

II. Local search

The local search stage is to capture the prey more accurately. The attack range of a goshawk is assumed to be R, and the mathematical formula is as follows:

$$X_i^{NEW,U2} = X_{i,j} + R(2r-1)X_{i,j} \tag{12}$$

In the formula above, $X_{i,j}^{NEW,U2}$ is the latest spatial position of the goshawk in the *j*-th dimension in the second stage, and the parameter r is a random number in [0,1].

$$R = 0.02(1-\frac{t}{T}) \tag{13}$$

In the formula (13), *t* is the current number of iterations, and *T* is the maximum number of iterations.

$$X_i = \begin{cases} X_i^{NEW,U2}, F_i^{NEW,U2} < F_i \\ X_i, F_i^{NEW,U2} \geq F_i \end{cases} \tag{14}$$

In the formula (14), $X_i^{NEW,U2}$ is the new position of the *i*-th goshawk, and $F_i^{NEW,U2}$ is the function value of the *i*-th goshawk after the second update.

b) Improve the rules

On the basis of the Northern Goshawk optimization algorithm, equation 12 describes the local search of the algorithm, where the change in the parameter R determines the change in the position of the northern goshawk, so as to coordinate the search ability of the algorithm. The larger the R value, the stronger the global search ability of the algorithm. The smaller the R value, the stronger the local search ability of the algorithm. From the formula 13, one can see that the parameter R has a linear decreasing change characteristic. This change is difficult to adapt when dealing with multi-dimensional complex optimization problems. Improving global search capabilities can avoid premature algorithms when dealing with multi-dimensional problems. Therefore, according to the characteristics of the cosine function, the formula 13 is improved as follows:

$$R' = 0.2\cos(\frac{\pi}{2}\cdot(1-\frac{t}{T})) \tag{15}$$

In the formula above, the parameter R has the characteristic of a cosine change. In the early stage of algorithm iteration, the parameter R changes slowly to improve the global search ability. In the later stage of the algorithm iteration, the R value decreases rapidly, which enhances the local search ability.

The updated formula is:

$$X_i^{NEW,U2} = X_{i,j} + R'(2r-1)X_{i,j} \tag{16}$$

In order to prove the superiority of the improved Northern Goshawk algorithm, different benchmark functions are used in this study for comparative experiments. The self-group optimization (SGO) algorithm, whale optimization (WO) algorithm, and the grey wolf optimization (GWO) algorithm are compared with the improved Northern Goshawk optimization (NGO) algorithm. The improved SGO, WO and GWO algorithms are used for further comparative tests. Figure 3 shows the convergence curves for 500 iterations. Due to space limitations, only representative experiments are shown in this paper. The benchmark functions used in the experiments are shown below.

$$f(1) = \sum_{i=1}^{n-1}(|x_i + 0.5|)^2 \tag{17}$$

$$f(2) = \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n}\cos(\frac{x_i}{\sqrt{i}}) + 1 \tag{18}$$

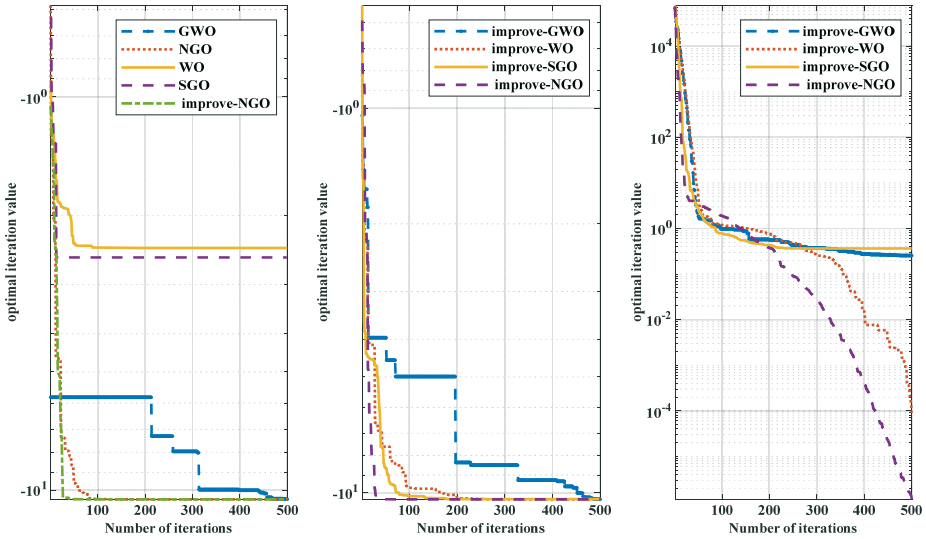

Fig. 3 Algorithm test convergence graph: **(a)** comparison experiment between the four optimization algorithms of SGO, WO, GWO, and NGO and the improved Northern Goshawk optimization algorithm under the f (1) benchmark function; **(b)** comparison experiment between the improved SGO, the improved WO and the improved GWO algorithms and the improved Northern Goshawk optimization algorithm under the f (1) benchmark function; **(c)** comparison experiment between the improved SGO, the improved WO and the improved GWO algorithms and the improved Northern Goshawk optimization algorithm under the f (2) benchmark function.

Note: SGO-self-group optimization; WO-whale optimization; GWO-grey wolf optimization; NGO-northern goshawk optimization

From the comparison of the convergence curves in Figure 1 one can see that the improved Northern Goshawk optimization algorithm is superior to the unimproved Northern Goshawk optimization algorithm and other optimization algorithms in terms of convergence speed. Other optimization algorithms, such as the GWO algorithm, have the problem of local optimal solution in the iterative process of the test function, while the improved Northern Goshawk optimization algorithm can effectively solve this kind of problem.

2.3 Optimized wavelet packet transform

The parameters of wavelet packet transform are optimized by using the improved Northern Goshawk optimization (improve-NGO) algorithm. In the traditional wavelet packet transform, when the number of decomposition layers K is small, the signal decomposition may be insufficient. The trend item may be mixed with other interference items, resulting in a larger envelope entropy. When an appropriate K value is taken, the envelope entropy of the trend item becomes smaller. Therefore, the algorithm uses envelope entropy as the fitness function in the process of wavelet packet transform optimization. The wavelet packet transform decomposition is optimal when the minimum entropy (local envelope entropy) in the decomposed signal is minimum.

The wavelet packet transform parameter optimization process is shown in Figure 4. Firstly, the position vector [K, α] of the population is initialized, and the fitness of each northern goshawk is calculated by using the envelope entropy as the fitness function. Then, by judging the convergence factor, an iterative formula is selected for iterative update until the termination condition is met and the optimal parameter combination is obtained.

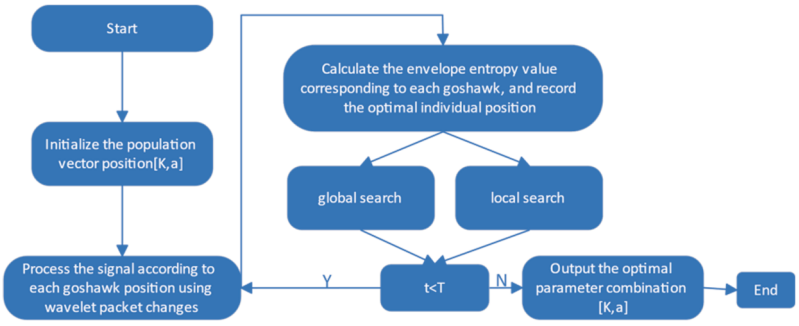

Fig. 4 Wavelet packet transform optimization flowchart

3. Temporal Convolutional Networks (TCNs)

Temporal convolutional network, or the time domain convolution network [16], is a model for time series processing proposed by Shaojie Bai et al. in 2018. The time domain convolution network model is a variant of the convolutional neural network. The network has a strong parallelism and does not occupy a lot of memory [17]; it fully solves the problems of gradient disappearance and gradient explosion in a recurrent neural network [18]. The TCN model mainly consists of dilated causal convolution (Dilated Causal Conv), weight normalization (Weight Norm), activation function (Relu), and dropout and residual connections. The TCN model structure is shown in Figure 5:

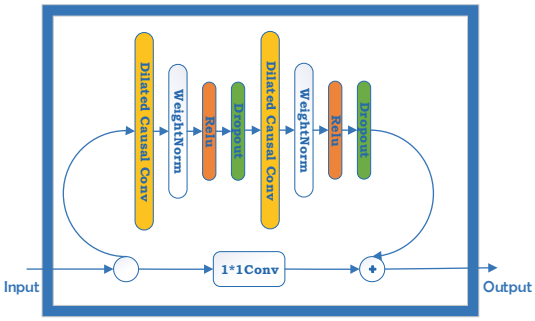

Fig. 5 TCN model structure

Dilated causal convolution can be simplified to three parts: dilation, causality, and convolution. Convolution refers to a smoothing operation of a convolution kernel conducted on data. Causality refers to the fact that at the time T of the n-th layer, the data only has influence on the $(n-1)$ layer at the time T and the value before the time T. This means that an element of the output sequence in a causal convolution depends only on the elements of the input sequence before it. Therefore, causal convolution is a strict time-constrained model. Figure 6 shows a causal convolution with a sampling interval of 1.

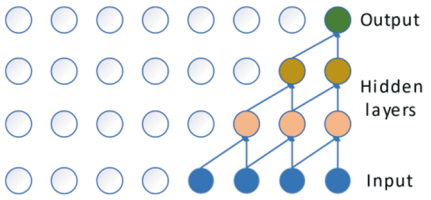

Fig. 6 Schematic diagram of causal convolution

On the basis of causal convolution, there is a sampling interval when convolution is allowed, which constitutes a dilated convolution, shown in Figure 7.

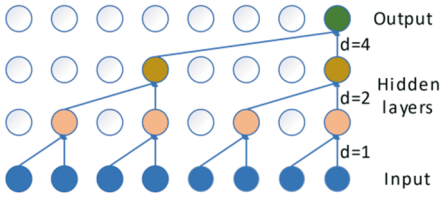

Fig. 7 Schematic diagram of dilated convolution

In the dilated convolution, the parameter d represents the sampling interval. In general, the sampling interval becomes larger as the number of layers increases. In Figure 7, we can see that the sampling interval increases exponentially with the increase in the number of layers.

The weight normalization in the TCN model is accelerated by rewriting the network weights, and the TCN model uses the Relu activation function. The Relu activation function can make network training faster. The Relu activation function formula is as follows:

$$f(x) = \max(0, x) \tag{19}$$

The dropout layer discards neural network units with a certain probability during the network training process to prevent the network from overfitting and achieve the effect of regularization.

In the case of not introducing the remaining connection modules, as the network model continues to deepen and the causal convolutions are continuously superimposed, the network model will inevitably fall into the phenomenon of gradient disappearance. To address this phenomenon, the TCN model introduces a residual module into the network. The residual module can make the model not increase the convolution module when it is optimal, fully solve the problem of gradient explosion and disappearance, and ensure the consistency of input and output.

In order to further improve the success rate of rolling bearing fault diagnosis, a rolling bearing fault diagnosis model based on the combination of optimized wavelet packet transform and time domain convolution network is proposed in this paper (improve-WPT-TCN model). A flowchart of the experiment is shown in Figure 8:

Note: WPT- wavelet packet transform; TCN-temporal convolution network; FC-fully connected.

Fig. 8 Experimental flow chart

In this experimental process, a fully connected (FC) layer module is added to output the results after the processing by the TCN network module. The FC layer plays a classification role in this network model. In the FC layer module, two fully connected layers will be used. The fully connected layer also uses the Relu activation function to speed up network training. The output value of the last fully connected layer will be passed to an output, which uses the Softmax logistic regression for the final result classification. The Softmax function can map the output real number domain to [0, 1], which is used to represent the effective real number space of the probability distribution. The Softmax formula is expressed as follows:

$$f(t) = \frac{1}{1+e^{-t}} \tag{20}$$

4. Experimental verification

4.1 Experimental data processing

The data acquisition platform, shown in Figure 9, is an online simulation platform. The structure includes the motor bearings required for the experiment, platform support frame, signal waveform display, bearing simulation fault work area, and data transmission line. The platform can realize a collection of fault signals required for the experiment. The experiment collects three kinds of faulty bearing signals and normal signals to verify the algorithm. The collected data includes 40 sets of normal signals, 120 sets of inner ring fault signals, 120 sets of outer ring fault signals, and 120 sets of rolling bearing signals. The experimental bearing is a 2612-type rolling bearing (single-row, radial, short cylindrical roller bearing); the spindle speed, *n*, is1045 r/min and the accelerometer sampling frequency, *f*, 1024 Hz.

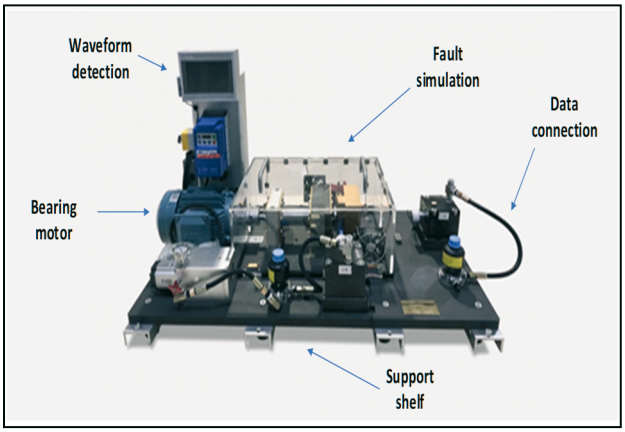

Fig. 9 Cloud experiment data acquisition platform

The rolling bearing signals recorded by the data in this experiment include normal signals, rolling element fault signals, inner ring fault signals and outer ring fault signals. Figure 10 below shows the original signal of the inner ring fault as an example.

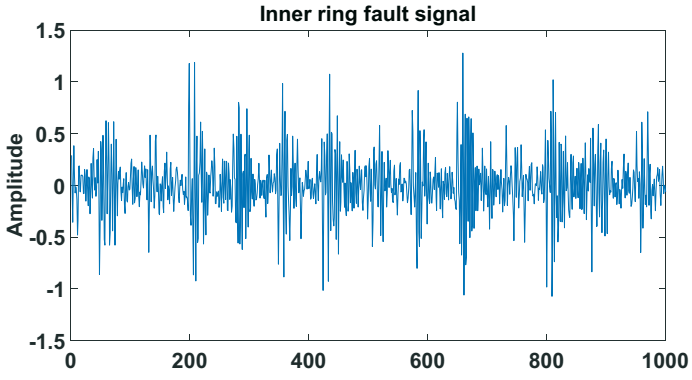

Fig. 10 Inner ring fault original signal

After the signal is collected, the improved NGO algorithm is used to optimize the wavelet packet transform and analyse the signal. Taking the fault signal of the inner ring as an example, the wavelet packet transform is optimized to decompose the fault signal of the inner ring of the rolling bearing. The improved North Goshawk algorithm optimization shows that the optimal decomposition layer number K is eight layers, and the optimal alpha value is 2000. Figure 11 shows the eight frequency band signals decomposed after algorithm optimization.

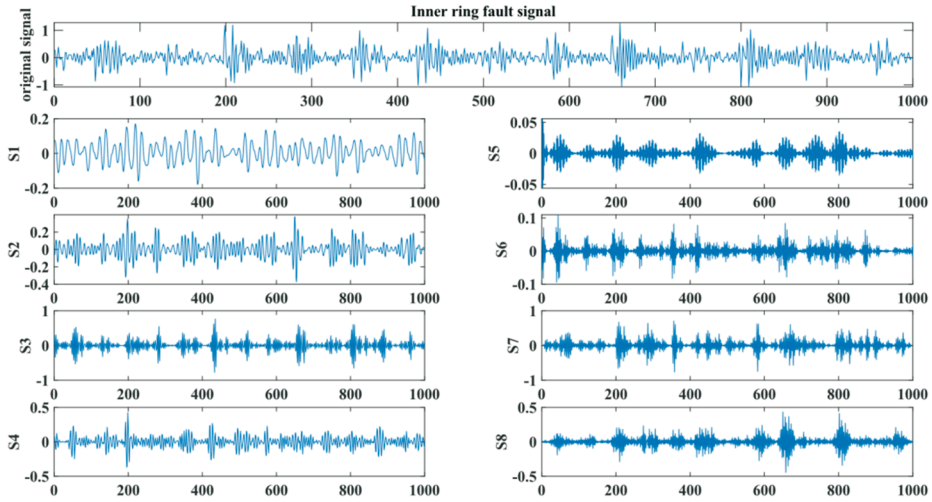

Fig. 11 The result of the wavelet packet transform of a normal signal

In order to best show the superiority of the improved Northern Goshawk algorithm, the optimized decomposed components of the algorithm are displayed through the Hilbert marginal spectrum, as shown in Figure 12.

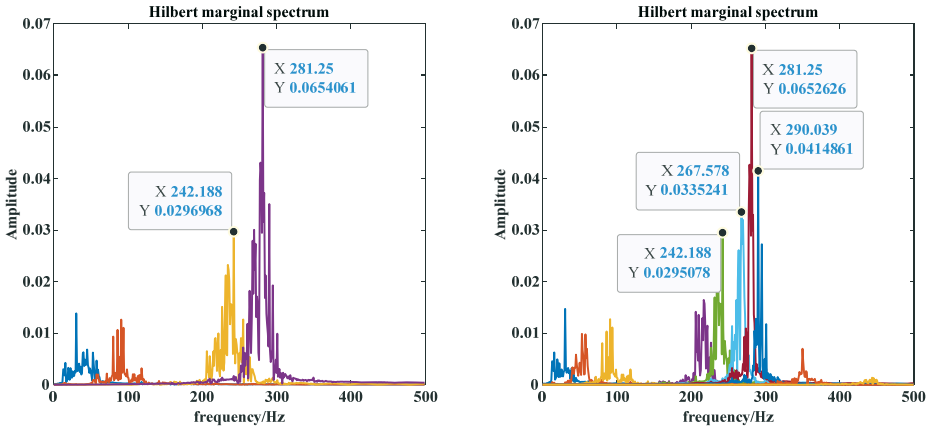

Fig. 12 Comparison of Hilbert transform. (a) Hilbert marginal spectrum after traditional wavelet packet decomposition; (b) The optimized wavelet packet decomposition of Hilbert marginal spectrum.

Figure 12 show that the main frequency domains where inner ring faults occur are around 242 Hz and 281 Hz. However, after the algorithm-optimized wavelet packet decomposition, the main fault frequencies are around 242 Hz, 267 Hz, 281 Hz, and 290 Hz. After comparison, it can be found that the component fault characteristics decomposed after optimization are more complete, and the decomposition of fault signals is more detailed, which is conducive to improving the accuracy of fault identification.

After the rolling bearing fault signal is decomposed, the eigenvector energy spectrum of the fault signal is constructed. The expression of the total energy E in the wavelet packet transform is:

$$E = \sum_{i=1}^{8} Q_3^i, Q_3^i = \sum_{i=1}^{8} (Q_3^i)^T P_3^i \tag{21}$$

In the formula above, P_j^i is the i-th wavelet obtained by the decomposition of the j-th layer of wavelet packets.

Figure 13 shows the energy spectrum of the four types of fault eigenvectors optimized by the improved wavelet packet transform of the rolling bearing. The feature extraction values of the first 5 groups of data are plotted, as shown in Table 1.

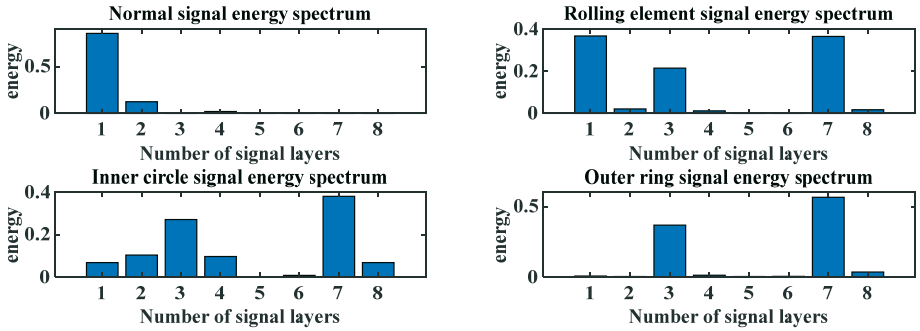

Fig. 13 Energy spectra of the fault feature vectors: (a) the normal signal energy spectrum; (b) the inner ring fault signal energy spectrum; (c) the rolling element fault signal energy spectrum; (d) the outer ring fault signal energy spectrum.

Note: 1-8 in the figure represent the energy of each layer after decomposing a signal into 8 layers.

Table 1 Feature extraction

Group	Eigenvalue 1	Eigenvalue 2	Eigenvalue 3	Eigenvalue 4	Eigenvalue 5	Eigenvalue 6	Eigenvalue 7	Eigenvalue 8
1	0.851675	0.124016	0.00248	0.020371	1.97E-05	0.000296	0.000339	0.000803
2	0.87448	0.104041	0.002334	0.017793	2.27E-05	0.000262	0.000296	0.000771
3	0.831864	0.140136	0.002691	0.023596	3.26E-05	0.000377	0.000363	0.000941
4	0.864769	0.111874	0.002377	0.01952	2.48E-05	0.000301	0.000319	0.000816
5	0.893891	0.088444	0.001729	0.014868	2.16E-05	0.000175	0.000267	0.000603

4.2 Training results

The eigenvector energy spectrum is obtained through the improved wavelet packet transform, and the eigenvector energy spectrum is used as the input training sample. Before the Improve-WPT-TCN fault diagnosis model training, four faults and one normal state of the rolling bearing are binary coded, and the coding results, shown in Table 2, are suitable for subsequent model training.

Table 2 Fault type number

Fault type	Numbering
Normal	0001
Inner ring failure	0010
Outer ring failure	0100
Rolling element failure	1000

After the binary numbering is completed, the previously obtained eigenvector energy spectrum data is imported into the Improve-WPT-TCN fault diagnosis network model for 500 rounds of model training. In this paper, the first 90% of the experimental data is used for training, and the last 10% is used for testing.

After 500 rounds of training, the trained model is used to test 10% of the fault data samples (including 4 normal signal samples, 12 inner ring fault signal samples, 12 outer ring fault signal samples, and 12 rolling element fault signal samples). In order to better reflect the superiority of the Improve-WPT-TCN model, this experiment uses the unimproved wavelet packet decomposition model (WPT-TCN) to conduct a control experiment and compare it with the test results of the Improve-WPT-TCN model. The final test result is shown in Figure 14.

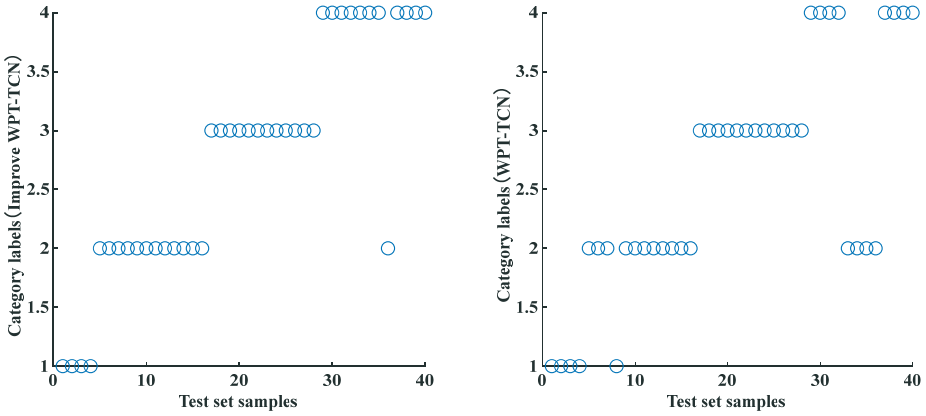

Fig. 14 Model test renderings: (a) the test results of the improved wavelet packet decomposition; (b) the test performance of the unimproved wavelet packet decomposition.

Table 3 shows the diagnostic effects of four faults.

Table 3 Fault diagnosis effect table

Combined name	Failure 1 Accuracy /%	Failure 2 Accuracy /%	Failure 3 Accuracy /%	Failure 4 Accuracy /%	Global accuracy /%
Improve WPT-TCN	100%	100%	100%	91.7%	97.9%
WPT-TCN	100%	91.7%	100%	83.3%	93.8%

From Figure 14 and Table 3, it can be seen that the recognition of no fault and outer ring fault has achieved the same accuracy. In the case of the inner ring fault and rolling element fault, the improved wavelet packet transform has higher recognition accuracy than the wavelet packet transform.

In order to test the capability of the improved Northern Goshawk algorithm to optimize the wavelet packet transform, different neural network algorithms combined with the wavelet packet transform are used for fault identification, and the optimized test results are compared. Different neural network test results are shown in Table 4.

Table 4 Different neural network test results

Combined name	Test set global accuracy /%
WPT-CNN	92.6%
Improve WPT-CNN	93.8%
WPT-SVM	90.3%
Improve WPT-SVM	91.9%
WPT-LSTM	92.5%
Improve WPT-LSTM	93.7%

Note: CNN-convolutional neural network; SVM-support vector machines; LSTM-long short-term memory.

It can be seen from Table 4 that when the wavelet packet transform is combined with different neural networks for fault identification, the improved Northern Goshawk algorithm is used to optimize the wavelet packet transform, and the final recognition accuracy is improved. This also shows the feasibility of the improved Northern Goshawk algorithm to optimize the wavelet packet transform.

In order to further compare the performance of the improved Northern Goshawk algorithm, the WO algorithm and the SGO algorithm are used to optimize the wavelet packet transform and identify faults. The optimized test results are shown in Table 5.

Table 5 Test results of different optimization algorithms

Combined name	Test set global accuracy /%
WPT-TCN	95.5%
Improve WPT-TCN	97.9%
WO-WPT-TCN	96.4%
SGO-WPT-TCN	95.8%

It can be seen from Table 5 that the improved Northern Goshawk algorithm used to optimize the wavelet packet transform exhibits better final recognition accuracy than other types of optimization algorithms. This also shows that the improved Northern Goshawk algorithm has more advantages than other types of optimization algorithms.

5. Conclusion

This paper proves the feasibility of the improved Northern Goshawk optimization algorithm to optimize the wavelet packet transform through experimental results and shows that the performance of the improved Northern Goshawk optimization algorithm is better than that of other algorithms.

By comparing the test results, one can conclude that the model combining the improved packet transform and time domain convolution network (Improve-WPT -TCN) and the unimproved wavelet packet decomposition model achieve the same detection effect for the normal state (fault 1) and the outer ring fault (fault 3) in the fault diagnosis of rolling bearings. However, in the diagnosis of inner ring fault (fault 2) and rolling element fault (fault 4), the diagnosis effect of the Improve-WPT-TCN model is significantly better than that of the unimproved wavelet packet decomposition model: the detection accuracy is increased by 8.3% and the global accuracy is increased by 4.1%.

REFERENCES

[1] Xu, Y.G.; Zhang, K.; Ma, C.Y.; Li, X.; Zhang, J. An improved empirical wavelet transform and its applications in rolling bearing fault diagnosis: *Appl. Sci.* **2018**, 8, 2352. https://doi.org/10.3390/app8122352

[2] Soldat, N.; Mitrovic, R.; Atanasovska, I.; Tomovic, R. A methodology for analyzing radial ball bearing vibrations: *Transactions of FAMENA*.**2020**,44(1),13-28. https://doi.org/10.21278/TOF.44102

[3] Hoang, D.T.; Kang, H.J. A survey on deep learning-based bearing fault diagnosis: *Neurocomputing* **2019**, 335, 327–335. https://doi.org/10.1016/j.neucom.2018.06.078

[4] Wu, J.D.; Liu, C.H. An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network: *Expert Syst. Appl.* **2009**, 36, 4278–4286. https://doi.org/10.1016/j.eswa.2008.03.008

[5] Shao, LQ.; Zhang, QH.; Yuan, PH. A dimensionless immune intelligent fault diagnosis system for rotating machinery: *Transactions of FAMENA*.**2022**,46(2),23-36. https://doi.org/10.21278/TOF.462032721

[6] Dybała, J.; Zimroz, R. Rolling bearing diagnosing method based on empirical mode decomposition of machine vibration signal: *Appl. Acoust*. **2014**, 77, 195–203. https://doi.org/10.1016/j.apacoust.2013.09.001

[7] Yen, G.G.; Lin, K.C. Wavelet packet feature extraction for vibration monitoring: *IEEE Trans. Ind. Electron*. **2000**, 47, 650–667. https://doi.org/10.1109/41.847906

[8] Zarei, J.; Poshtan, J. Bearing fault detection using wavelet packet transform of induction motor stator current: *Tribol. Int*. **2007**, 40, 763–769. https://doi.org/10.1016/j.triboint.2006.07.002

[9] Fan, Y.R.; Zhang, C.; Xue, Y.; Wang, J.; Gu, F. A bearing fault diagnosis using a support vector machine optimised by the self-regulating particle swarm: *Shock Vib*. **2020**, 2020, 9096852. https://doi.org/10.1155/2020/9096852

[10] Zhu, J.; Chen, N.; Peng, W. Estimation of bearing remaining useful life based on multiscale convolutional neural network: *IEEE Trans. Ind. Electron.* **2018**, 66, 3208–3216. https://doi.org/10.1109/TIE.2018.2844856

[11] Wang, H.; Xu, J.W.; Yan, R.Q.; Gao, R.X. A new intelligent bearing fault diagnosis method using SDP representation and SE-CNN: *IEEE Trans. Instrum. Meas* **2020**, 69, 2377–2389. https://doi,org/10.1109/TIM.2019.2956332

[12] Lu, Q.; Yang, R.; Zhong, M.; Wang, Y. An improved fault diagnosis method of rotating machinery using sensitive features and RLS-BP neural network: *IEEE Trans. Instrum. Meas.***2020**, 69, 1585–1593. https://doi,org/10.1109/TIM.2019.2913057

[13] Xu, G.; Liu, M.; Jiang, Z.; Söffker, D.; Shen, W. Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning: *Sensors.* **2019**, 19, 1088. https://doi.org/10.3390/s19051088

[14] Su, NQ Fault feature extraction of bearings for the petrochemical industry and diagnosis based on high-value dimensionless features br.; Zhou, ZJ.; Zhang, QH.; Hu, SL.; Chang, XX. *Transactions of FAMENA.***2022**,46(4),31-44. 10.21278/TOF.464036521

[15] Gu, Y.K.; Zhou, X.Q.; Yu, D.P.; Shen, Y.J. Fault diagnosis method of rolling bearing using principal component analysis and support vector machine: *J. Mech. Sci. Technol.* **2018**, 32, 5079–5088. https://doi.org/10.1007/s12206-018-1004-0

[16] Dehghani M.; Hubálovský Š.; Trojovský P. Northern Goshawk Optimization: A New Swarm-Based Algorithm for Solving Optimization Problems: *IEEE Access*. **2021**, 9: 162059-162080. https://doi.org/10.1109/ACCESS.2021.3133286

[17] Shaojie Bai J.; Zico Kolter.; Vladlen Koltun. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling: *Computer Science*. **2018**, 9: 162-178. https://doi.org/10.48550/arXiv.1803.01271

[18] Zhang, W.; Li, C.; Peng, G.; Chen, Y. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load: *Mech. Syst. Signal Process.* **2018**, 100, 439-453. https://doi.org/10.1016/j.ymssp.2017.06.022

Submitted: 09.11.2022

Accepted: 05.5.2023

Dengxue Cao*
Yu Gu
Prof. Wei Lin
School of electrical and electronic engineering, Shanghai Institute of Technology, Shanghai 201418, China
*Corresponding author:
dengxuecao@163.com