Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 11, No. 3, 2025

Acoustic Emission Diagnosis of Rolling Bearing Faults based on Optimized VMD-Transformer

Hao Chen, Yang Yu *

School of Information Science and Engineering, Shenyang University of Technology, Shenyang, Liaoning 110870, China
* **Corresponding author**: Yang Yu

Abstract: The acoustic emission signal of rolling bearing fault has poor feature data extraction effect, resulting in low fault diagnosis accuracy. This paper proposes a rolling bearing fault diagnosis method based on optimized variational mode decomposition (VMD) denoising and feature enhancement, and using the Transformer model for diagnosis. The Seahorse Optimizer (SHO) algorithm is used to optimize the VMD parameters [K, α]. The optimized VMD is used to decompose the acoustic emission signal of the bearing to obtain several intrinsic mode components (IMFs). The intrinsic mode components containing more fault information are selected according to the envelope entropy and kurtosis, and the signal is reconstructed; the reconstructed signal is input into the Transformer model to complete the diagnosis. The experimental results show that the accuracy of the optimized VMD-Transformer fault diagnosis model reaches 98.1%, which is significantly improved compared with other fault diagnosis models.

Keywords: Acoustic Emission Signa; Rolling Bearing; Sea-horse Optimizer; Transformer; Variational Module Architecture.

1. Introduction

As the core component of rotating machinery, the operating status of rolling bearings is directly related to the reliability and safety of the equipment. However, under long-term high load and complex environmental conditions, rolling bearings are prone to various faults such as wear, peeling, and cracks [1]. If these potential faults are not detected in time, they will not only lead to equipment failure, but also cause major economic losses and safety accidents. Therefore, the development of efficient and accurate rolling bearing fault diagnosis methods has important theoretical value and engineering significance.

In the field of fault diagnosis, acoustic emission signals have become an important data source for rolling bearing fault detection due to their high sensitivity to micro crack extension and internal material damage [2]. However, acoustic emission signals usually exhibit non-stationary, high noise and nonlinear characteristics, which brings great challenges to feature extraction and fault pattern recognition [3]. Commonly used classical signal processing methods such as empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), and wavelet decomposition are all affected by modal aliasing and endpoint effects.

Variational mode decomposition (VMD) is an optimized and improved signal decomposition method. It decomposes the signal into a set of narrowband modes through a variational framework. It has a good mathematical foundation and decomposition effect [4]. However, the decomposition performance of VMD is heavily dependent on the selection of its key parameters, while traditional empirical or heuristic methods are subjective and limited in parameter optimization.

In the feature extraction and classification stage, the rapid development of machine learning and deep learning technology has provided powerful tools for rolling bearing fault diagnosis. Traditional machine learning methods such as support vector machines (SVM) and random forests rely on manually designed features, have limited feature expression capabilities and are difficult to deal with complex fault modes [5-6]. In recent years, deep learning methods such as convolutional neural networks (CNNs) and long short-term memory networks (LSTMs) have shown excellent performance in fault diagnosis. However, CNNs are insufficient in extracting global features of time series, while LSTMs are susceptible to gradient vanishing and low training efficiency due to their recursive structure [7-8].

The Transformer model has attracted widespread attention for its success in the field of natural language processing. Its architecture based on the attention mechanism can flexibly model long-distance dependencies and effectively process large-scale data. In recent years, the application of Transformer in time series analysis has gradually emerged, and it has shown strong feature modeling capabilities in mechanical fault diagnosis [9].

Based on the above background, this paper proposes a rolling bearing acoustic emission fault diagnosis method based on optimized variational mode decomposition (VMD) and Transformer. First, the SHO algorithm [10] is used to globally optimize the parameter combination [K, α] of VMD to overcome the subjectivity of traditional parameter selection and improve the accuracy and stability of signal decomposition; the decomposed modal components are used to select components with rich fault information to reconstruct the signal. The position encoding and multi-head attention mechanism enable the Transformer model to use the bearing fault acoustic emission signal directly as the input of the model to realize rolling bearing fault diagnosis. Experiments show that this method has a high fault diagnosis accuracy.

2. VMD Parameter Optimization

2.1. Variational Mode Decomposition

Variational Mode Decomposition (VMD) is a data-driven adaptive signal decomposition method proposed by Dragomiretskiy and Zosso in 2014. VMD decomposes the input signal into several modal components (IMFs) with limited bandwidth through variational derivation, which can effectively suppress modal aliasing and end-point effects.

The core idea of VMD is to use variational optimization methods to decompose the input signal f(t) into several modal components uk(t) with limited frequency bands and minimize the bandwidth of each component. The main steps are as follows:

Construct a constrained variational model, where each IMF uk(t) has a specific center frequency ωk. VMD minimizes the bandwidth of each modal component by finding an optimal set of {uk, ωk}, that is

$$min\sum_{k}\left\|\frac{d}{dt}\left[u_k\left(t+jH\left(u_k(t)\right)\right)e^{-jw_k t}\right]\right\|^2 \quad (1)$$

Here, H(.) is the Hilbert transform, which is used to resolve the signal into a single sideband signal.

The Lagrange multiplier method is used to construct the optimization problem. By constraining $\sum_k u_k=f(t)$ so that the sum of the components of the modal decomposition is equal to the original signal, the Lagrangian function is constructed:

$$L(\{u_k\}\{\omega_k\})=\alpha\sum_{k}\left\|\frac{d}{dt}\left[u_k(t+jH(u_k))e^{-j\omega_k t}\right]\right\|^2+\|f-\sum_k u_k\|^2+\langle\lambda f-\sum_k u_k\rangle \quad (2)$$

Among them, α is the tuning parameter and λ is the Lagrange multiplier.

The alternating direction multiplier method (ADMM) is used to iteratively solve uk(t) and ωk. The calculation is converted to the frequency domain through Fourier transform and updated in each iteration: the modal components uk are calculated, the center frequency ωk is updated, and the Lagrange multiplier λ is updated until the convergence condition is met, that is:

$$\left\|u_k^{(n)}-u_k^{(n-1)}\right\|^2<\epsilon \quad (3)$$

Among them, ϵ is the set convergence threshold.

2.2. Seahorse Optimization Algorithm

The Seahorse Optimization Algorithm (SHO) is a swarm intelligence optimization algorithm inspired by the natural behavior of seahorses. It mainly simulates the movement of seahorses in water. The core of SHO is to simulate the movement pattern of seahorses to update the population position, so as to find the optimal solution in the search space. It mainly includes position update strategy and fitness evaluation.

1. Initialize the population. Assuming the search space is D-dimensional and the population size is N, the position of the hippocampus individual can be expressed as:

$$X=\{X_1,X_2,\dots,X_N\} \quad (4)$$

Among them, each individual Xi is a D-dimensional vector:

$$X_i=\{X_{i1},X_{i2},\dots,X_{iD}\} \quad (5)$$

The population is usually initialized in a random manner:

$$X_{i,j}=\{X_{min,j}+\text{r}.(X_{max,j}-X_{min,j}) \quad (6)$$

Where $X_{min,j}$ and $X_{max,j}$ represent the minimum and maximum values of the j-th dimension variable, respectively, and r is a random number between [0,1].

2. Position update formula, The search process of SHO is mainly divided into the exploration phase and the exploitation phase, and the following formula is used to update the individual position:

(1) Exploration phase (global search)

$$X_i^{t+1}=X_i^t+S\cdot R \quad (7)$$

Where: S is the step factor, R is the random direction vector, usually, the step factor is defined as:

$$S=e^{-\lambda t}\cdot(X_{best}-X_i^t)+\beta\cdot(X_r^t-X_s^t) \quad (8)$$

Where : λ is the convergence factor , Xbest is the current optimal solution , X_r^t and X_s^t are two randomly selected hippocampal individuals , and β is the control factor .

$$X_i^{t+1}=X_i^t+\gamma\cdot(X_{best}-X_i^t)+\delta.R \quad (9)$$

Where : γ is the local search weight , δ is the perturbation factor , and R is the random perturbation .

3. Transformer Theory

3.1. Self-Attention Mechanism

The basic idea of the self-attention mechanism is that for each element in the input sequence, its similarity with all other elements in the sequence is calculated and attention weights are assigned according to the similarity. Specifically, given an input sequence The specific calculation formula of $X\in R^{n\times d}$ is as follows:

$$\text{attention}(Q,K,V)=\text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (10)$$

Where: Q, K, V are query matrix, key matrix, and value matrix respectively, $Q=XW_Q$, $K=XW_K$, $V=XW_V$. W_Q , W_K , $W_V \in R^{d\times d_k}$, d_k is the dimension of the latent space for attention calculation .

3.2. Multi-head Self-Attention Mechanism

The multi-head attention mechanism is an extension of the self-attention mechanism, which aims to capture different feature subspaces by computing multiple attention heads in parallel. Each head calculates its own independent attention weight and weighted sum, so that it can focus on different parts or information of the input sequence. By splicing the results of these different heads together and then undergoing linear transformation, a comprehensive representation is finally formed. Specifically, the multi-head attention mechanism can help the model learn information on multiple subspaces simultaneously, thereby improving the model's expressiveness and learning ability.

Assume we have h heads, denoted as head 1, head 2, head h, and the output of each head is the weighted sum of the self-attention mechanism. The output of the multi-head attention mechanism can be expressed as:

$$\textit{M}\text{ultiHead}(Q,K,V)=\textit{Concat}(\text{head}_1,\text{head}_1\dots.\text{head}_h)W_O \quad (11)$$

Among them: head i = Attention (XW_Q^i,XW_K^i,XW_V^i) i is the output of the i- th head, is the linear transformation matrix of the head; W_Q^i,W_K^i,W_V^i is the linear transformation matrix of the head; W_O is the linear transformation matrix of the final output; Attention is a self-attention mechanism, which represents a weighted summation process.

3.3. Positional Encoding

Since the Transformer model does not have a loop structure like RNN, it cannot automatically perceive the time sequence of the input data. For bearing signals, timing information is crucial. Therefore, it is necessary to add position information to the input data to ensure that the model can correctly interpret the time evolution of the signal. Positional Encoding uses sine and cosine functions. Waveforms of different frequencies can provide position information at different scales:

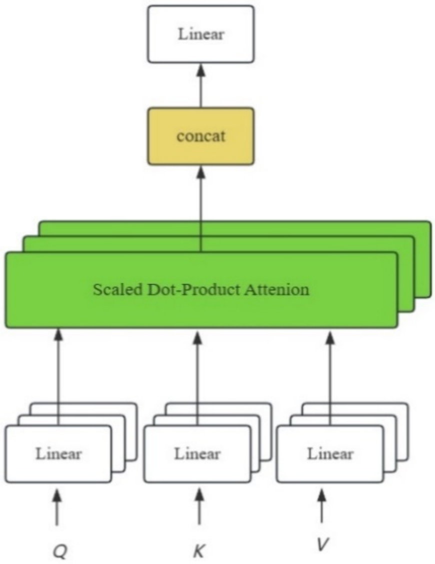

Figure 1. Multi-head Self-Attention Mechanism

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{2i/D}}\right) \quad (12)$$

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{2i/D}}\right) \quad (13)$$

Among them: pos represents the position of the current data point in the time series, D is the dimension of the input feature, and through sine/cosine waves of different scales, points at different positions have unique representations, thereby encoding time series information.

The "Acknowledgment" (no "s) section appears immediately after the conclusion. If applicable, this is where you indicate funding for the work. The preferred spelling of the word "acknowledgment" in American English is without an "e" after the "g." Avoid expressions such as "One of us (S.B.A.) would like to thank ... ." Instead, write "We thank ... ." Sponsor and financial support acknowledgments are included in the acknowledgment section. For example: This work was supported in part by the U.S. Department of Commerce under Grant BS123456 (sponsor and financial support acknowledgment goes here). Researchers that contributed information or assistance to the article should also be acknowledged in this section. Also, if corresponding authorship is noted in your paper, it will be placed in the acknowledgment section. Note that the acknowledgment section is placed at the end of the paper before the reference section.

4. Optimizing VMD-Transformer Fault Diagnosis Model

This section will introduce in detail the rolling bearing acoustic emission diagnosis method based on optimized VMD and transformer model. This section is divided into three parts: optimization of VMD parameters, IMF selection and signal reconstruction and diagnosis model structure.

4.1. Optimizing VMD Parameters

In the process of VMD signal decomposition, parameters K and α have a great influence on the decomposition quality of the signal. It is difficult to get the appropriate parameters by manually setting the parameters, which affects the quality of VMD decomposition. As an emerging swarm intelligence optimization method, the Hippocampus Optimization Algorithm (SHO) has good global search capabilities, fast convergence speed and robustness. Therefore, the Hippocampus Optimization Algorithm is used here to optimize the VMD parameters to obtain the optimal parameter combination.

This paper uses the minimum envelope entropy as the fitness function. Envelope entropy is an indicator to measure the complexity and uncertainty of signal energy distribution. It reflects the degree of chaos of the signal by analyzing the probability distribution of the signal envelope. Lower envelope entropy usually means that the signal energy distribution is more concentrated and the structure is clearer, while higher envelope entropy indicates that the energy distribution of the signal is more dispersed and may be interfered by noise. After VMD decomposes the bearing acoustic emission signal, the IMF component with more noise often leads to an increase in envelope entropy. Using the minimum envelope entropy as the objective function helps to suppress the influence of noise on the decomposition results, thereby highlighting the main features of the signal. The process of SHO optimizing VMD parameters is shown in Figure 2.

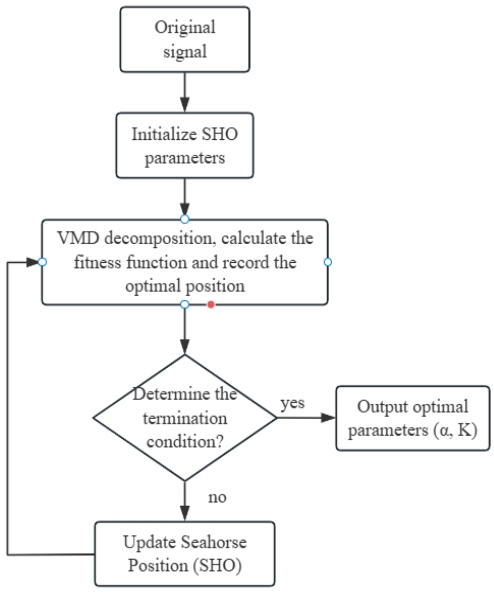

Figure 2. SHO optimization process of VMD parameters

4.2. IMF Selection and Signal Reconstruction

After the original signal of the bearing is decomposed by optimized VMD, the envelope entropy is used as an indicator to reduce the impact of noise on the signal. In order to improve the reconstruction effect of the signal and obtain more fault information, the kurtosis value is used to select the IMF for the second time.

When a bearing fails, it is usually accompanied by intermittent shocks. Kurtosis is a statistical indicator to measure the degree of signal peaks. IMFs with larger kurtosis may contain more fault characteristic components, while IMFs with lower kurtosis may only be stable noise or trend signals. Kurtosis calculation formula:

$$K = \frac{E[(X-\mu)^4]}{(E[(X-\mu)^2])^2} \quad (14)$$

Where X is the signal data, μ is the mean, and E[.] represents the expected operation.

4.3. Diagnostic Model Structure

After the bearing signal is decomposed and reconstructed by optimized VMD, the time series signal cannot be directly used as the input of the Transformer model because the Transformer itself does not have built-in time sequence perception capabilities. Therefore, positional encoding is required to tell the model the time sequence of the input. As input data, the time series signal must first be embedded in time segments and divided into multiple time segments. Each segment is mapped to a high-dimensional space through a linear transformation. In order to enable the model to capture global information, the [CLS] Token is added, which is a special marker for global feature learning. It is added to the beginning of the input sequence and represents the global features of the entire input signal. At the same time, positional encoding is added to each segment to ensure that the Transformer can understand the time sequence in the input signal. After being processed by the Transformer encoder, it enters the classifier to achieve fault classification.

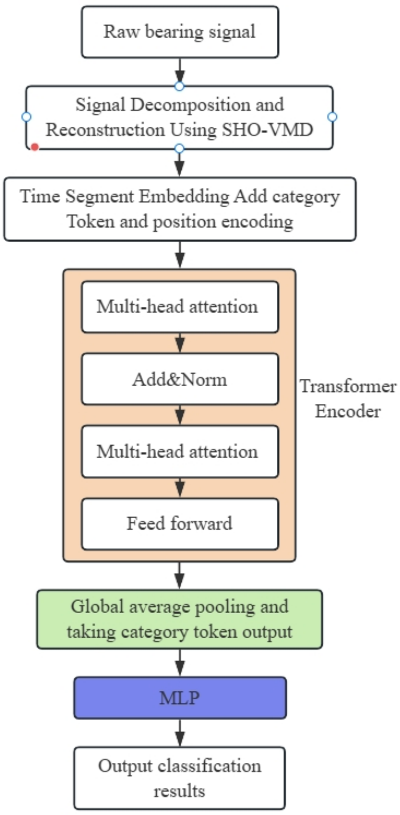

Figure 3. SHO-VMD-Transformer model

5. Experimental Analysis

5.1. Experimental Data

The data of this experiment is collected by using the rolling bearing acoustic emission detection experimental platform as shown in Figure 4 to collect acoustic emission signals of rolling bearings in various states. Five acoustic emission signals were collected under normal bearing conditions, rolling element failure, cage failure, outer ring failure, and inner ring failure. The motor speed was 600r/min. 200 samples were collected for each bearing state, of which 120 samples were used to train the model, 40 samples were used to verify the model, and the remaining 40 samples were used for model testing. The length of each sample is 1024.

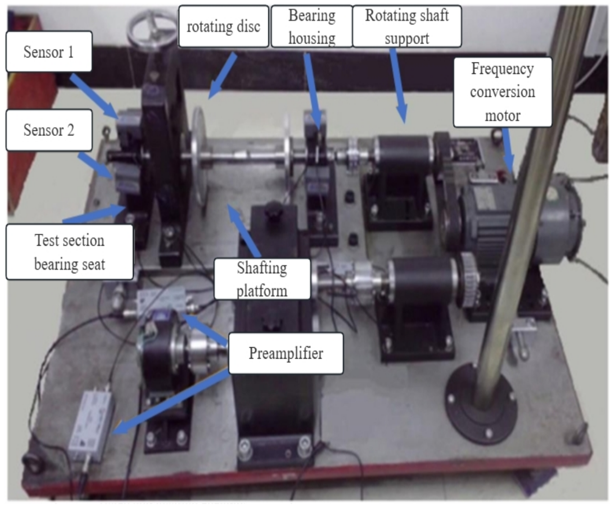

Figure 4. Experimental platform for acoustic emission testing of rolling bearings

5.2. Experimental Data

In order to improve the diagnostic effect of the model, we must first solve the modal aliasing of VMD decomposition. Since the artificial parameters K and α cannot obtain a better decomposition effect, the VMD parameters K and α are optimized by the SHO algorithm, and the envelope entropy is used as the fitness function to improve the decomposition effect. The range of the number of decompositions K is set to [3, 10], and the range of the penalty factor α is set to [200, 2500]. Taking the inner ring fault as an example, the optimization curve is shown in Figure 5. As the number of iterations increases, the fitness value gradually decreases and converges to 8.3062. The optimal parameter combination [K, α] is [6, 2231]. The parameter combination is brought into VMD to obtain the decomposed IMF frequency domain diagram. As shown in Figure 6, the central frequencies of each IMF are evenly distributed, and there is no endpoint effect and modal aliasing phenomenon, which proves the effectiveness of the SHO-VMD method.

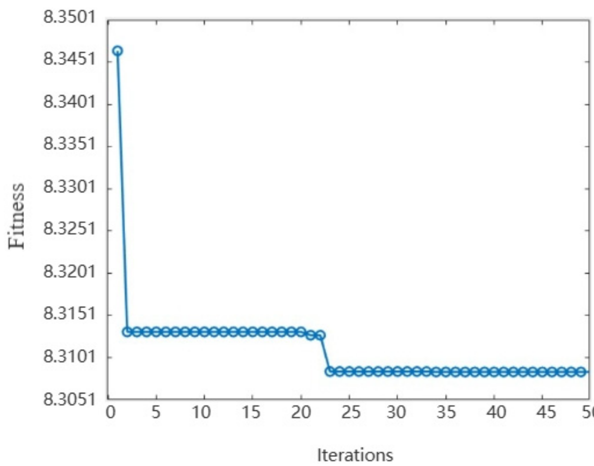

Figure 5. SHO optimizes VMD parameters

The envelope entropy and kurtosis value of each IMF component are calculated, as shown in Table 2. From the data in the table, it can be seen that IMFs with larger kurtosis often have smaller envelope entropy values, and the minimum envelope entropy and maximum kurtosis appear at IMF7 at the same time. According to the joint screening criteria of envelope entropy and kurtosis, IMF6 is selected as the characteristic signal for signal reconstruction.

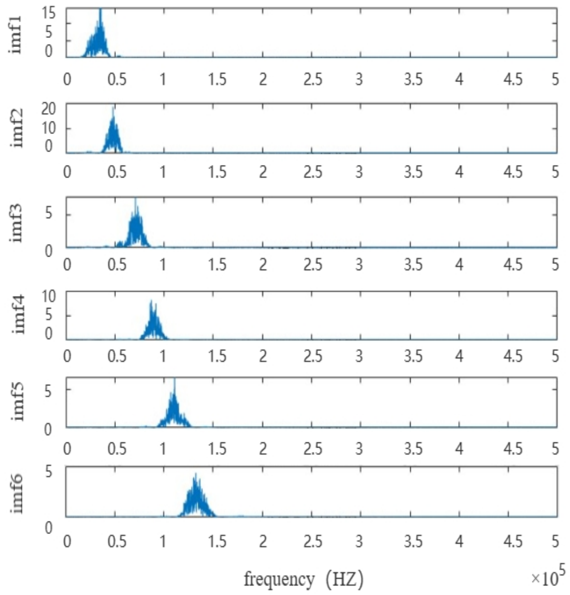

Figure 6. IMF frequency domain diagram

Table 1. Envelope entropy and kurtosis of each IMF

IMF	1	2	3
Envelope Entropy	8.3777	8.3605	8.3162
Kurtosis	5.5825	7.7501	9.0418

IMF	4	5	6
Envelope Entropy	8.3717	8.3255	8.3062
Kurtosis	8.3479	7.1871	13.2897

5.3. Model Diagnostic Performance Analysis

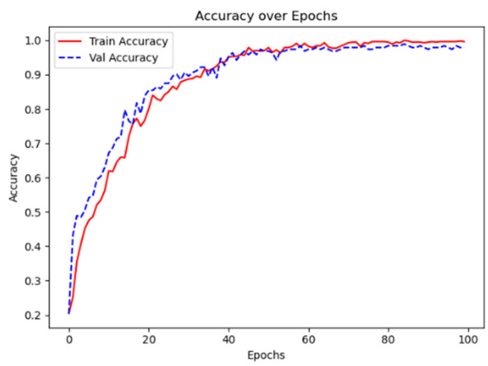

Figure 7. Model accuracy curve

The parameter optimizer is selected as Adam optimizer, and the loss function is cross entropy loss. The learning rate is set to 0.00003, Batch Size=64, Dropout=0.2, and the number of iterations is 100 for model training. The model training and verification results are shown in Figures 6 and 7. It can be seen from the figure that the accuracy of the model gradually increases in the early stage of training. After 60 rounds, the training accuracy gradually increases to about 99% for the training set and about 97% for the verification set and gradually stabilizes. The same trend can also be seen from the loss curve. It can be seen that the method proposed in this paper can accurately identify the fault state of rolling bearings.

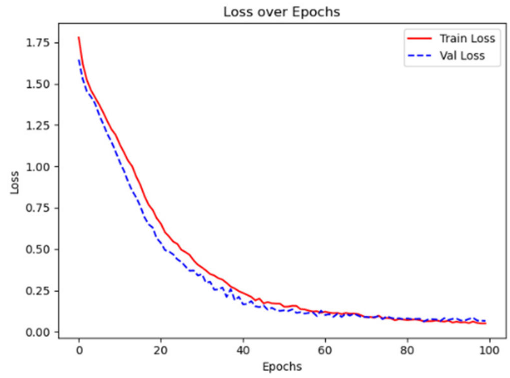

Figure 8. Loss Curve

In addition, this paper uses the accuracy of the confusion matrix as the evaluation index of the model's diagnostic performance. The test data set is input into the classification model, and the performance of the model algorithm is measured by dividing the number of all correctly predicted samples by the total number of samples. The specific results are in the confusion matrix, where each column represents the predicted value and each row represents the true value. Each element in the confusion matrix represents the number of times a sample is predicted to be a certain category. The confusion matrix is composed of the following table 1. In order to verify the reliability and effectiveness of the above model, this paper selects the accuracy of the confusion matrix as the evaluation index of the model, that is, the ratio of the number of samples predicted correctly by the model to the total number of samples. The higher the accuracy, the better the model prediction and classification effect. The calculation formula is as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \qquad (15)$$

Table 2. Confusion Matrix

Actual condition	Prediated Condition	
	Positive	Positive
Positive	TP	FN
Negative	FP	TN

The confusion matrix of the model is shown in Figure 8, where the ordinate represents the true label and the abscissa represents the predicted label. Each row of the ordinate represents the proportion of this type of fault that is misclassified as other types of faults; and each column of the abscissa shows the proportion of the five samples that are judged as this type of fault. As can be seen from the figure, the overall prediction effect of the model is good. Only one sample of the second type of fault (inner ring fault) is considered to be a third type of fault (outer ring fault), and the rest are all accurately predicted. It can be seen that the SHO-VMD-Transformer model has a good diagnostic effect.

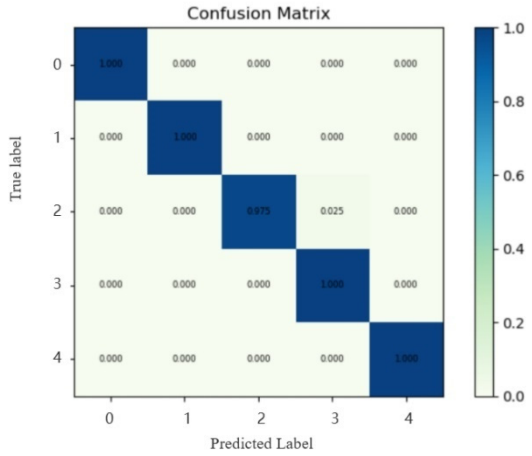

Figure 9. Model confusion matrix

In order to prove the diagnostic performance of the model proposed in this paper, comparative experiments were conducted with SVM, CNN, Transformer, VMD-Transformer and other methods. Without VMD, the bearing signal was directly input into the Transformer. Due to the interference of noise and other factors, the diagnostic results were not ideal. After VMD, although the diagnostic accuracy can be improved to a certain extent by combining the Transformer, the diagnostic effect is not particularly ideal due to the influence of VMD parameters. SHO-VMD-Transformer can obtain a reconstructed signal with significantly better features for fault diagnosis by optimizing VMD, and the fault recognition rate can reach 98.1%. Therefore, the fault diagnosis performance of this model is more superior.

Table 3. Envelope entropy and kurtosis of each IMF

Model	fault identification rate
SHO-VMD-Transformer	98.1%
VMD-Transformer	93.8%
CNN	87.2%
SVM	82.3%

6. Conclusion

In the study of rolling bearing fault diagnosis, this paper proposes a diagnostic method that uses the SHO algorithm to optimize VMD combined with Transformer. By optimizing the VMD parameters K and α through the SHO algorithm, the subjectivity of traditional parameter selection can be effectively overcome. Then, the modal components with rich fault signals are screened out by using kurtosis and envelope entropy for signal reconstruction.

By using time series embedding and position encoding, the acoustic emission signal (time series signal) of the bearing fault is directly input into the Transformer model for fault diagnosis, making full use of the global dependency modeling capability of the Transformer, enhancing the model's automatic learning and feature extraction capabilities, and improving the accuracy and robustness of the diagnosis. After experimental verification, the fault recognition rate of SHO-VMD-Transformer reached 98.1%, which is 93.8% higher than that of VMD-Transformer. It proves that the method is feasible.

References

[1] Smith, W. A., & Randall, R. B. (2015). Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. Mechanical Systems and Signal Processing, 64, 100-131.

[2] Jin Rongshun, Shen Gongtian, Wang Qiang, et al. Fault diagnosis of rolling bearing acoustic emission signals of large cable car based on EMD and approximate entropy [J]. Journal of China University of Metrology, 2018, 29 (04): 417-423.

[3] Xu Zhonglin, Li Guolu, Dong Tianshun, et al. Research progress in acoustic emission signal analysis and processing methods [J]. Materials Review, 2014, 28 (09): 56-60+73.

[4] K. Dragomiretskiy and D. Zosso, "Variational Mode Decomposition," in IEEE Transactions on Signal Processing, vol. 62, no. 3, pp. 531-544, Feb.1, 2014.

[5] Xu Ke, Chen Zonghai, Zhang Chenbin, et al. Rolling bearing fault diagnosis based on empirical mode decomposition and support vector machine [J]. Control Theory and Applications, 2019, 36 (06): 915-922.

[6] Zhang Yu, Chen Jun, Wang Xiaofeng, et al. Application of random forest in rolling bearing fault diagnosis [J]. Computer Engineering and Applications, 2018, 54 (06): 100-104+114.

[7] Chen Renxiang, Huang Xin, Yang Lixia, et al. Rolling bearing fault diagnosis based on convolutional neural network and discrete wavelet transform [J]. Journal of Vibration Engineering, 2018, 31 (05): 883-891.

[8] Chen Baojia, Chen Xueli, Shen Baoming, et al. Application of CNN-LSTM deep neural network in rolling bearing fault diagnosis [J]. Journal of Xi'an Jiaotong University, 2021, 55 (06): 28-36.

[9] Qiu Dawei, Liu Zichen, Zhou Yiqing, et al. Rolling bearing fault type identification based on Transformer neural network [J]. High Technology Letters, 2021, 31 (01): 1-11.

[10] Zhao, S., Zhang, T., Ma, S. et al. Sea-horse optimizer: a novel nature-inspired meta-heuristic for global optimization problems. Appl Intell 53, 11833–11860 (2023).