

Zero-shot compound fault diagnosis of gearboxes via fault-coupled semantics

SU Xingliang¹, XU Ke¹, ZHANG Kaixing¹, LIAO Jianbin², LI Guoqiang², YAN Jin^{2*}

1. Silk Production Workshop, Longyan Tobacco Industrial Co., Ltd., Longyan 364021, China;

2. School of Marine Engineering, Jimei University, Xiamen 361021, China

*Corresponding author: YAN Jin (jasonyan@jmu.edu.cn)

Received: October 27, 2025

Revised: November 23, 2025

Accepted: December 19, 2025

Abstract: Gearboxes with complex structures are key components in rotating machinery, and their failures exhibit both sporadic and coupling characteristics. However, in engineering practice, gearbox fault data are extremely scarce, and the sample size for composite faults is effectively zero. Therefore, it is of great significance to investigate fault diagnosis modeling methods under scarce fault samples and zero composite fault samples. This paper designs a new loss function for single single-fault samples and zero composite fault samples based on the prior knowledge of similarity between time series monitoring signals and the difference between single faults, as well as the prior knowledge that composite faults are composed of single faults. This enables the effective optimization of the feature encoder. On this basis, a diagnostic algorithm for single faults and composite faults of gearboxes is constructed, achieving effective diagnosis of various states of gearboxes. This paper validates the proposed method using two gearbox fault experimental platforms. The results show that the proposed method can construct and obtain the gearbox fault diagnosis model under single single-fault samples and zero composite fault samples.

Key words: gearboxes; compound fault; zero shot; fault semantics; physical information network; contrastive learning

0 Introduction

Gearboxes are one of the most important components of rotating machinery^[1-3]. Accurate monitoring condition of the gearbox is essential to ensure the reliability of rotating industrial equipment. Gearboxes commonly suffer from single faults such as bearing failure, tooth wear, tooth breakage, scuffing, polishing, etc. However, these faults often occur in different parts of the gearbox and are often coupled, which makes it difficult to diagnose compound faults in practice^[4-5]. Currently, data-driven methods for gearbox fault diagnosis have achieved significant results, but these methods are often applied to industrial scenarios with very few fault samples, which makes their modelling very difficult^[6-8]. In addition, these methods do not consider compound faults, so it is urgent to strengthen research on how to establish an effective diagnostic model for compound faults with very few single fault data.

Traditional gearbox fault diagnosis methods mainly rely on signal analysis techniques and expert experience. On the basis of signal analysis, time domain analysis (e. g., peak factor, crag index) and frequency domain

analysis (e. g., Fourier transform) are usually used to extract vibration signal features, combined with threshold judgement or spectral comparison to identify the typical failure modes (e. g., tooth wear, broken teeth, etc.)^[9-10]. Time-frequency analysis methods (e. g., short-time Fourier transform, wavelet analysis) further enhance the fault feature extraction capability of non-smooth signals^[11-13]. In addition, physicochemical methods based on temperature monitoring and oil abrasive particle analysis are also widely used for comprehensive diagnosis. Although these traditional methods have the advantages of mature theory and easy implementation, they rely on artificial feature engineering and have limited generalization ability for complex faults, which provides an improved direction for the subsequent development of intelligent diagnosis techniques.

Deep neural network models such as CNN, ResNet, transformer, are used to automatically learn the deep fault features in the raw condition monitoring signals of gearboxes, which solves the empirical dependence problem of manual feature design and shows a strong advantage in end-to-end fault diagnosis^[14-16]. Since data-driven methods^[17-18] rely on the deep learning of data

samples, once the samples in the training set are not representative, the gearbox fault diagnosis model constructed by relying on such methods will encounter challenges in practical applications. Therefore, variable condition fault diagnosis and small sample fault diagnosis methods for gearboxes have been greatly developed, in which a large number of researches have been carried out to solve the variable condition fault diagnosis problem based on migration learning and the small sample fault diagnosis problem based on data generation.

As data-driven fault diagnosis methods for gearboxes continue to evolve, they need to be applied in engineering practice^[19]. However, although the techniques represented by transfer learning enable the methods studied in the academic field to be effectively applied in engineering practice to gear transmission equipment, these algorithms are still adjusted and optimized based on actual objects, and face the challenge of how to further enhance their universality and robustness in application. In recent years, physical information neural networks (PINNs)^[20-22] have been able to constrain the network model learning process by combining the physical knowledge of mechanical equipment and the generalization of the acquired physical knowledge, which makes this type of method have a strong potential for practical application. Yu et al. combined a bearing dynamics model with a CycleGAN variant to develop a rolling bearing fault diagnosis model in the absence of real bearing fault data^[23]. Sun et al. embedded fault representations from multimodal signals of rotating machinery into an attention network, which allowed the model to focus on typical fault characteristics with physical properties, further improving the performance of the model^[24]. Zhu et al. proposed a causally and physically-driven domain generalization network, PhysiCausalNet, which was capable of mining fault causality and combining it with a priori knowledge of the physics of the unknown target machine equipment to achieve effective diagnosis of unknown faults^[25].

Physical information-driven fault diagnosis methods provide a feasible path to increase the state identification performance of gearboxes in engineering practice. However, such methods still face the need to be supported by a large number of data samples, which makes it urgent to further break through the dependence on fault samples under knowledge-driven conditions. Currently, zero-sample learning has received extensive attention in the field of fault diagnosis of rotating machinery. Zhang et al. proposed a zero-sample learning method based on semantic space and 1D CNN, which

described the shared fault attributes of visible and invisible faults through semantic space, and achieved the recognition of unseen faults through similarity judgement^[26]. Xu et al. constructed a compound fault diagnosis model with zero compound fault samples based on the semantics of fault features^[27]. Liao et al. proposed an attribute-consistent generative adversarial network (ACGAN-FG) for zero-sample fault diagnosis for generating more diversified fault features, which further enhanced the robustness of the fault diagnosis model for rotating machinery fault diagnosis model robustness^[28]. For gearboxes in engineering practice, a small number of fault samples and zero fault samples are more common^[29-30].

Based on our prior knowledge of the physical characteristics of time series signals and the semantic knowledge between single faults and composite faults, we utilized a single sample of a single fault and combined it with the fact that it can be obtained in large quantities from a large amount of normal state data. We propose a new design scheme. This scheme aims to design a new network model that can achieve efficient training of the loss function of deep network models. On this basis, a diagnostic algorithm was constructed for single and compound faults in gearboxes. The main contributions of the proposed methodology are summarized as follows.

- 1) By exploiting physical priors, including the pronounced temporal coherence of condition-monitoring signals and the inherent separability of distinct gearbox fault states, we propose three novel loss functions that obviate the need for compound-fault samples during encoder training.

- 2) By jointly optimizing these losses using only a small number of single fault samples, a discriminative feature encoder capable of learning the state signals of the gearbox can be generated.

- 3) Using the resulting encoders, we construct an end-to-end diagnostic framework capable of identifying both single and compound faults in gearboxes.

1 Proposed method

A novel diagnosis method with zero compound fault sample is proposed which fuses the physical information of the similarity of timing signals, the difference in meanings between different state samples, and the semantic difference between compound faults and single faults. The proposed method is shown in Fig. 1. The proposed method consists of three main parts. Firstly, the imbalance of the sample set often leads to the failure

of effective training for deep network models, thus unable to solve the imbalance problem between normal state samples and fault state samples in actual industrial scenarios. Secondly, the single fault state sample semantic information between the single fault state sample and the compound fault state sample is used to construct the compound fault state sample. Finally,

introducing the prior similarity knowledge between time signal samples and the inherent differences among samples of different fault states, this enables the construction of a comparison matrix, the design of a comparison loss function, and the application of the designed comparison loss function to train the deep network model.

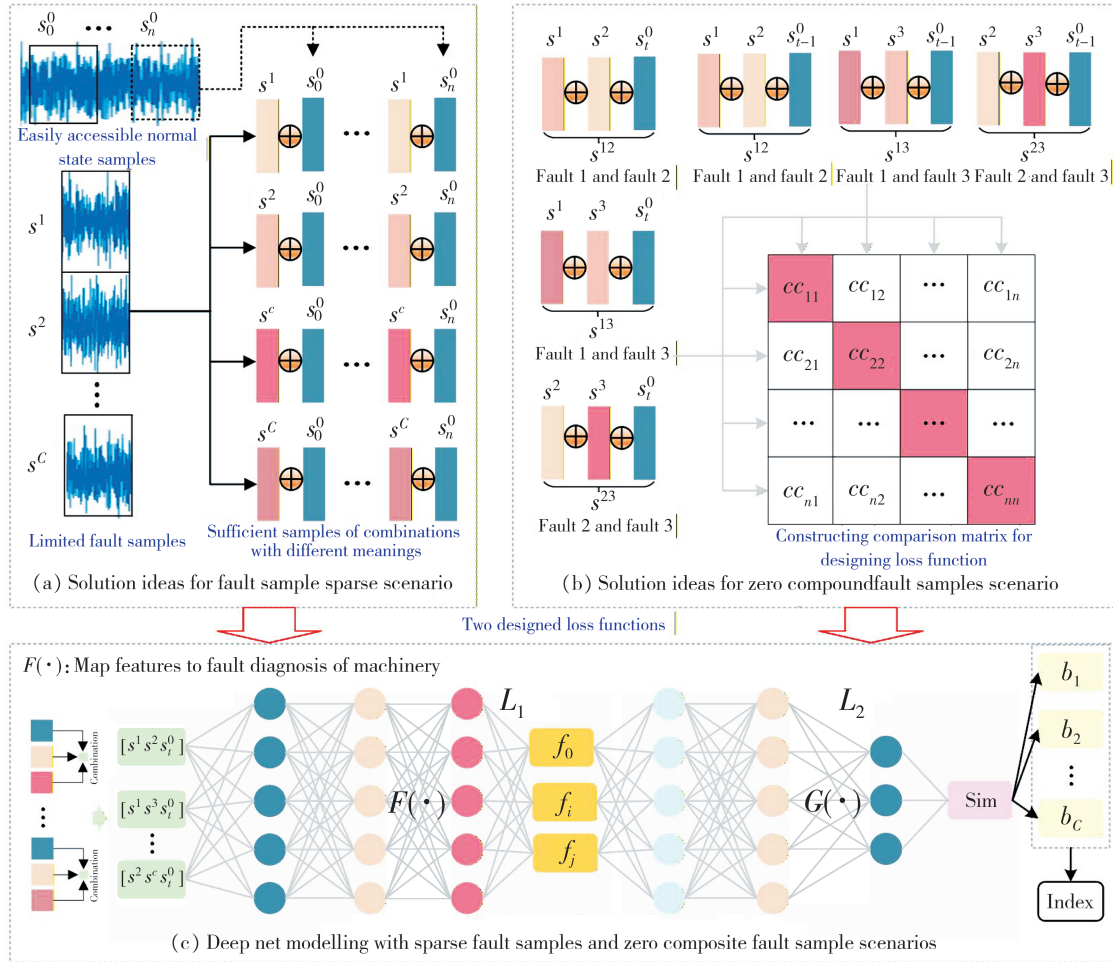


Fig. 1 Framework for diagnostic methods under zero compound fault samples

1.1 Single fault data scarcity resolution strategy

In real-world scenarios, gearboxes have a very low failure probability, and most operate under normal conditions, as they are generally designed with sufficient reliability guarantees. However, once a gearbox fails, the resulting damage is severe. Therefore, the cost of acquiring real fault state data is extremely high, which also leads to a very limited number of available fault samples. Data-driven gearbox fault diagnosis methods have received widespread attention. The core lies in conducting model training with a large number of sufficient sample data, so that the model can recognize various types of samples. However, such methods often face the problem of extreme imbalance among training

samples in practice. Considering that the normal state data of the gearbox are sufficient, and this obviously meets the demand of the data-driven model for the number of data samples, So if the single fault samples are combined with the normal state samples, then there will be no imbalance between the single fault samples and the normal state samples. Meanwhile, since the individual faulty samples are combined with the normal state samples, their information meanings are different. Therefore, they have label significance in the data-driven approach.

Let the known single fault samples be $\{s^1, s^2, \dots, s^C\}_a$, respectively, where C denotes the number of categories of fault samples, and a denotes the number of single fault samples available. Let the set of normal state samples be

$\{s_1^0, s_2^0, \dots, s_n^0\}$, where 0 denotes the normal state, n denotes the number of samples in the normal state, and $n \gg a$. Thus, the physical information can be described as

$$s^1 \oplus s_i^0 \neq s^2 \oplus s_i^0, \dots, \neq s^C \oplus s_i^0. \quad (1)$$

When $a = 1$, one gets $\{s^1 \oplus s_1^0, s^1 \oplus s_2^0, \dots, s^1 \oplus s_n^0\}, \dots, s^C \oplus s_1^0, s^C \oplus s_2^0, \dots, s^C \oplus s_n^0$. It can be seen that the number of samples for various fault types is consistent with that for the normal state, indicating that the application of the above physical information can resolve the issue of sample imbalance between mechanical equipment fault samples and normal state samples.

1.2 Definition of semantics under zero compound fault samples

For gearboxes, coupled faults occur in most scenarios, making it crucial to conduct accurate diagnosis of compound faults. However, considering the strong coupling nature of compound faults in practical applications, it is extremely challenging to perform reliability experiments on such faults. Therefore, establishing an accurate compound fault diagnosis model under the condition of zero compound fault samples is of great significance.

Compound faults are composed of the combination of multiple single faults. Therefore, by using single faults to obtain samples that can represent compound faults, the semantic representation of compound faults is defined. Specifically, let the single fault be $\{s^1, s^2, \dots, s^C\}$. The compound fault is represented as

$$s^i \oplus s^j \rightarrow s^{ij}, \quad (2)$$

where $i, j \in \{1, 2, \dots, C\}, i \neq j$. It is possible to achieve the characterization of compound faults under the condition that a single fault sample is obtained.

1.3 Diagnostic algorithm under zero compound fault samples

By incorporating the defined physical information and compound fault semantics, the issue of training dataset imbalance is resolved under the practical conditions of scarce single fault samples and zero compound fault samples. Existing data-driven methods indicate that the key lies in achieving effective training of the deep network model. Therefore, how to implement such a deep network model under the scenario of scarce single fault samples and zero compound fault samples is crucial. Furthermore, after obtaining the trained deep network model, how to apply it to the fault diagnosis of single faults and compound faults in mechanical equipment is

another key challenge.

Hence, the training of the deep network model is implemented with a single fault sample and zero compound fault samples, which mainly includes the following steps.

Step 1 The physical information that the timing signals of mechanical equipment have the high degree of similarity is introduced. Specifically, two normal state samples (s_{t-1}^0, s_t^0) are selected from the timing signals during the normal operation of mechanical equipment, and then the physical information is expressed as

$$H(s_{t-1}^0) \cong H(s_t^0), \quad (3)$$

where $H(\cdot)$ denotes the information entropy. It can be seen that $s_{t-1}^0 \cong s_t^0$.

Step 2 Introduce the compound fault semantic formulation, i. e., $s^1 \oplus s^2 \rightarrow s^{12}$, $s^1 \oplus s^3 \rightarrow s^{13}$, and $s^i \oplus s^j \rightarrow s^{ij}$. To solve the problem of the imbalance between the fault samples and the normal state samples, the compound fault semantics is introduced to couple the fault 1 with the fault 2 as an example.

$$s_{t-1}^0 \oplus s^1 \oplus s^2 \rightarrow s_{t-1}^{12}, \quad (4)$$

$$s_t^0 \oplus s^1 \oplus s^2 \rightarrow s_t^{12}. \quad (5)$$

Combining with Eq. (3), we can obtain

$$s_{t-1}^{12} \cong s_t^{12}. \quad (6)$$

Step 3 Introduce the compound fault semantic samples, construct the deep network model, and carry out network model training using the compound fault semantic samples. For the network model, the key lies in constructing an appropriate loss function. The network model consists of two parts: the encoder is $F(\cdot)$ and the fully connected network layer is $G(\cdot)$. The compound fault semantic sample is $\{s_{t-1}^0 \oplus s^i \oplus s^j\}_m$, and its similar compound fault semantic sample is $\{s_t^0 \oplus s^i \oplus s^j\}_m$, where $t \in [1, 2, n]$, $j \in [1, 2, C]$, and m is the compound fault category.

For m compound faults with the following semantics, different compound faults have different semantic meanings between them, and have the same semantics in the normal state samples $t-1$ vs. t . Taking m as 2 for example, the two compound faults are $s^1 \oplus s^2$ vs. $s^1 \oplus s^3$, respectively, which can be obtained by

$$f_{t-1}^{12} = F(\{s_{t-1}^0 \oplus s^1 \oplus s^2\}_1), \quad (7)$$

$$f_t^{12} = F(\{s_t^0 \oplus s^1 \oplus s^2\}_1), \quad (8)$$

$$f_{t-1}^{13} = F(\{s_{t-1}^0 \oplus s^1 \oplus s^3\}_2), \quad (9)$$

$$f_t^{13} = F(\{s_t^0 \oplus s^1 \oplus s^3\}_2), \quad (10)$$

where $f_{t-1}^{12} \cong f_t^{12}$ and $f_{t-1}^{13} \cong f_t^{13}$, and introducing the compound fault sample semantics, it follows that

$$f_{t-1}^{12} \neq f_{t-1}^{13}. \quad (11)$$

Jointly with the above semantic formulation, it can be seen that $cc(f_{t-1}^{12}, f_t^{12}) \rightarrow 1$ and $cc(f_{t-1}^{12}, f_{t-1}^{13}) \rightarrow 0$. Therefore, if we unite the Eqs. (7) to (10), the constructed comparison matrix is

$$M = \begin{bmatrix} cc(f_{t-1}^{12}, f_t^{12}) & \cdots & cc(f_{t-1}^{1n}, f_t^{12}) \\ \vdots & & \vdots \\ cc(f_{t-1}^{12}, f_t^{1n}) & \cdots & cc(f_{t-1}^{1n}, f_t^{1n}) \end{bmatrix}. \quad (12)$$

Step 4 The comparison matrix (12) embodies a priori knowledge that diagonal elements are maximized and off-diagonal elements are minimized. The loss function for maximizing diagonal elements is

$$L_1 = - \sum_{i,j}^C cc(f_{t-1}^{ij}, f_t^{ij}),$$

$$\text{s.t. } i \neq j; i, j \in \{1, 2, \dots, C\}, t \in \{1, 2, \dots, n\}. \quad (13)$$

Moreover, the loss function for the minimal representation of the non-diagonal elements is

$$L_2 = \sum_{i,j,n,m}^C cc(f_{t-1}^{ij}, f_t^{nm}),$$

$$\text{s.t. } ij \neq nm, t \in \{1, 2, \dots, n\}. \quad (14)$$

Furthermore, based on the inherent prior knowledge among normal conditions, individual faults, and combined faults, it can be concluded that distinct differences exist among the three scenarios.

$$f^{ij} = F(s^0 \oplus s^i \oplus s^j), \quad (15)$$

$$G(f^{ij}) \rightarrow \begin{cases} 0, & i \neq j, \\ 1, & i = j. \end{cases} \quad (16)$$

Thus, combining Eqs. (15) and (16), a new loss function can be defined as

$$L_3 = \sum_{i \neq j}^C G(f^{ij}) - \sum_{i=j}^C G(f^{ij}). \quad (17)$$

Step 5 Joint L_1 , L_2 , and L_3 , the contrastive loss function can be established as

$$L = \alpha L_1 + \beta L_2 + \gamma L_3. \quad (18)$$

Setting L as an optimised loss function with $F(\cdot)$ for the encoder and $G(\cdot)$ for the fully connected network layer, the above network model is trained and optimised by applying normal state samples, single fault samples, and compound fault semantic samples.

Step 6 By combining the deep network model trained based on normal state samples with individual samples of various faults, diagnostic algorithms for single faults and composite faults can be designed.

Set trained feature encoder as $M(\cdot)$, and one online monitored condition data as the s_x .

Set the baseline sample as $s^0 \oplus s_0^i \oplus s_0^j$, and the test sample as $s^0 \oplus s_0^i \oplus s_x$.

Step 1 Calculate the similarity between the tested sample and baseline samples of each category.

$$\text{sim}_m^i = M(s^i \oplus s^i \oplus s_x), i = 0, 1, 2, \dots$$

$$\text{sim}_{fe}^i = cc(M(s^i \oplus s^i \oplus s^i), M(s^i \oplus s^i \oplus s_x)), i = 0, 1, 2, \dots$$

where $cc(\cdot)$ the cosine similarity calculation operation.

Step 2 Perform fault detection and fault identification.

if $\text{argmin}\{\text{sim}_m^i\} = 0$ or $\text{argmax}\{\text{sim}_{fe}^i\} = 0$:

s_x is normal condition sample.

else:

s_x is abnormal condition sample.

if $(\text{argmin}\{\text{sim}_m^j\} = j$ or $\text{argmax}\{\text{sim}_{fe}^j\} = j)$ and $j \neq 0$:

s_x is single fault condition.

$p_x = j$

else:

$p_x = j + 4$.

2 Case studies and experimental results

2.1 Wind power gearboxes case study

2.1.1 Experimental setup description and data preparation

The fault dataset obtained from the gearbox experiments was used to validate the method proposed in this case. It mainly consists of the prime mover, gearbox, asynchronous motor, and data acquisition unit. Single fault experiments and compound fault experiments were carried out using this experimental platform. Among them, three kinds of single faults were selected, which were tooth blade falling off (C2), outer ring broken (C4), gear box fixed disc loosening (C6), and two kinds of compound faults, which were tooth blade falling off + outer ring broken (C2+C4), tooth blade falling off + fixed disc loosening (C2+C6). During each fault test, the vibration signal of the gearbox was collected by the NI-cDAQ-9174/9234 vibration sensors in both directions of the X-axis and Y-axis, with a sampling frequency of 10 240 Hz, and 1 228 800 vibration data points are collected.

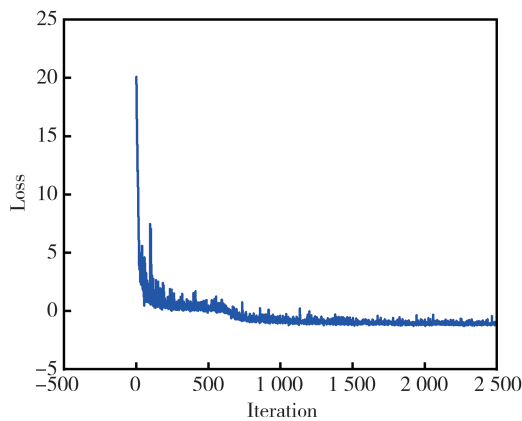
In order to verify the effectiveness of the method proposed in this paper, four typical faults of the wind turbine gear box were selected for the verification of the fault diagnosis method for rotating machinery under extreme data imbalance conditions. The data set of the wind turbine gear box faults was obtained using the experimental platform, where the sample data points in the normal state were 2 048, totaling 2 000 data samples, and the data samples for each fault state were 1 000. The detailed information of the data set is shown in Table 1.

Table 1 Validation dataset for gearbox condition diagnosis methods with zero compound fault samples

Condition	Fault type	Description	Number of training samples	Number of testing samples
Normal	Seen	Nor	1 000	1 000
Loss of teeth	Seen	LT	1	1 000
Outer ring fault	Seen	OIF	1	1 000
Loose gearbox retaining disk	Seen	LGRD	1	1 000
Compound fault 1	Unseen	CF1	0	1 000
Compound fault 2	Unseen	CF2	0	1 000

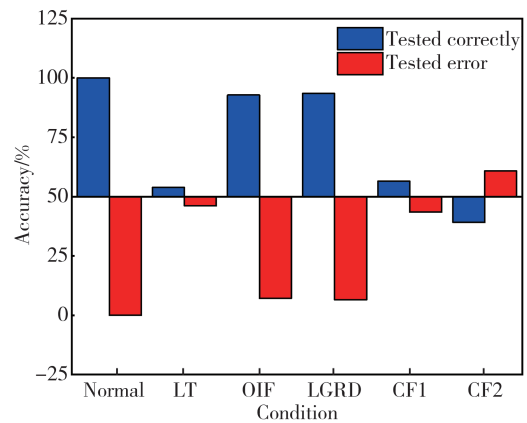
2.1.2 Discussion and analysis of results

The effectiveness of the loss function designed with zero compound fault samples and few single fault samples is verified, as shown in Fig. 2. Thus, even if there is only one sample for each individual fault type, the effective training of each deep network can still be achieved. This shows that the priori knowledge embedded in the model training process can solve the zero-compound fault sample limitations and the extreme imbalance between single fault samples and normal state samples, which has a convenient modelling advantage in practical engineering applications.

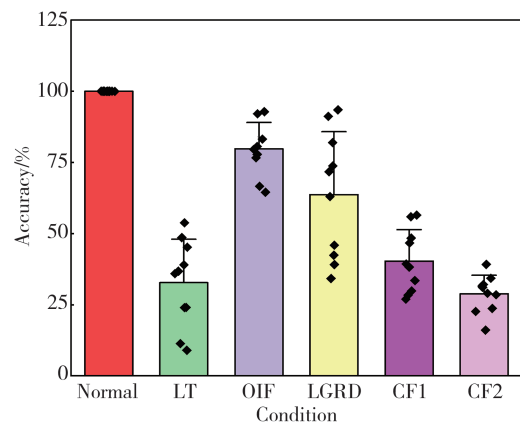
**Fig. 2** Training process

The test results of the proposed fault diagnosis algorithm with zero compound fault samples and single fault samples are shown in Fig. 3(a). When both blue and red bars are relatively long, it indicates strong recognition capability for that particular category. It can be seen that the test accuracy is 100% for normal state. The test accuracy is more than 50% for fault LT, OIF, LGRD, CF1. The test accuracy is only 53.8% for fault LT. And the fault test accuracies of compound fault CF1 and CF2 are 56.5% and 39.2%. Therefore, although the accuracy rate of fault identification for wind turbine gearboxes is no more than 80%, the construction cost of its fault diagnosis model is low, which meets the requirements of actual modeling. Fig. 3(b) shows that for the normal state of the wind power gear box. The test result of the normal state has achieved a 100% accuracy rate. However, there is still considerable room for

improvement in the recognition accuracy of other fault states. Additionally, the test results indicate that the proposed fault detection and identification method tends to classify the test samples as the normal state. This leads to errors when distinguishing between single faults and compound faults. Therefore, future research should focus on optimizing the identification algorithms for single faults and compound faults, such as determining more appropriate threshold settings.



(a) Types of wind turbine gearbox failures



(b) Results of ten fault tests

Fig. 3 Wind turbine gearbox troubleshooting test results

To verify that the designed loss function can be applied to the training of different networks, efficient former, ResNet, and CNN were used as feature encoders. By the constructed loss function (Eq. (18)), α , β , and γ are set to 1, and normal state samples and individual fault samples of the wind turbine gearbox are used to train several feature encoders, which are then

applied to the fault diagnosis algorithm. Fig. 4 presents the optimization process and fault diagnosis test results of three feature encoders under the designed loss function. From the training process of CNN, Resnet, and efficient former feature encoders as shown in Fig. 4(a), 4(c), and 4(e), it can be seen that effective optimization of network encoders can be achieved under zero compound fault samples, single fault samples, and normal state samples. In addition, by using trained feature encoders, effective recognition of single and compound faults in gearboxes can be achieved, as shown in Fig. 4(b), 4(d), and 4(f). However, from the test results of its single and compound faults, it can be seen that there is still considerable room for improvement in its performance. In CNN-based fault diagnosis, the optimal test accuracies for single faults are only 53.8%,

92.8%, and 93.5%, while the test accuracies for compound faults are only 56.5% and 39.2%, indicating a large room for improvement. In the fault diagnosis based on efficient former, the accuracy of normal state testing is 73.8%, the accuracies of single fault testing are 70.4%, 92.8%, 93.5%, and the accuracies of compound fault testing are 57.0%, 77.6%. Comparing the fault diagnosis methods based on CNN and efficient former, the former has stronger ability to identify normal states and a higher risk of overfitting. The latter has better recognition ability for single faults and compound faults and good generalization ability.

In practical engineering, it is obvious that the efficient former has more advantages. However, the inherent defect of high false alarm rate still needs to be further improved.

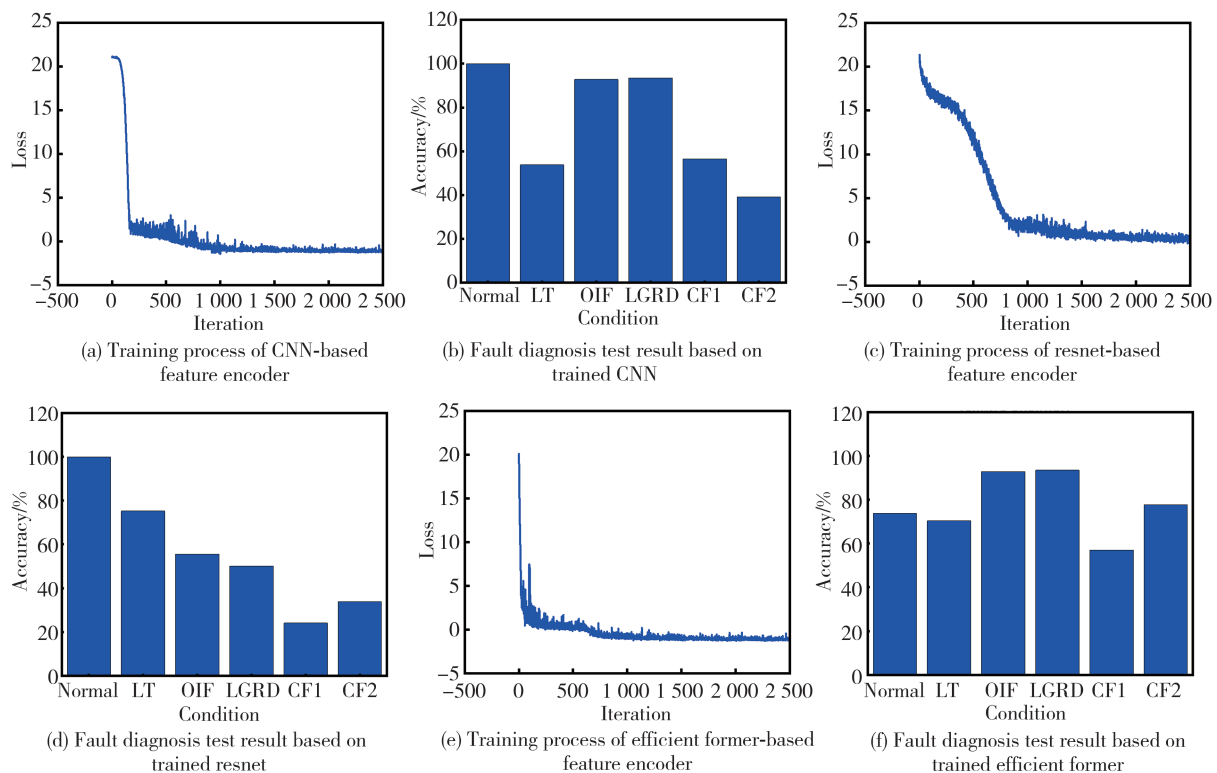


Fig. 4 Training process of three network feature encoders with designed loss functions and their test results

Table 2 presents the ten validation results of fault diagnosis methods based on CNN, ResNet, and efficient former, as well as their computational resource consumption. All methods were performed on the Ubuntu system, Python 3.9, Python 2.5.1, and Nvidia 4090 graphics card. Three different methods are adopted for fault diagnosis and detection. By comparison, it can be seen that the fault diagnosis and detection method based on CNN has the best effect, with the shortest time consumption. The fault diagnosis response time is approximately 0.2 s, which can meet the requirements of long-term engineering

applications.

Finally, it was compared with the fault diagnosis method based on advanced machine learning, including extreme gradient boosting (XGBoost), support vector machine (SVM), least absolute deviations (LAD), bernoulli naive bayes (BNB), gaussian naive bayes (GNB), decision tree, and principal component analysis (PCA). Table 3 presents the comparison results of the proposed method in two scenarios. In scenario 1, a one-time training approach is adopted, which means that each type of fault sample is a participant in the training process, and the fault samples

gradually expand from a single sample to 1 000 samples to balance the various types of samples. In scenario 2, 1 000 samples are trained for each type of fault. Six comparison indicators are used to evaluate the performance of the comparison methods, including the fault detection and diagnosis results under a single fault sample, the fault

detection and diagnosis results based on fully supervised learning, and the duration of training and testing. From the comparison results, it can be seen that the proposed method has more significant advantages in fault diagnosis and fault detection, indicating that this method is advanced and does not consume more computing resources.

Table 2 Results of ten tests for fault diagnosis based on different feature encoders

Encoder	Fault diagnosis/%	Fault detection/%	Training time/s	Testing time/ms
CNN-based	57.23±7.48	82.35±10.60	185.19±1.54	26.8±0.20
Resnet-based	42.53±6.92	66.61±12.18	395.23±1.52	23.9±0.07
Efficient former-based	49.65±6.08	71.03±9.58	211.84±7.40	36.6±2.60

Table 3 Comparison of proposed method with other state-of-the-art machine learning methods.

Method	1-shot fault diagnosis/%	1 000-shot fault diagnosis/%	1-shot fault detection/%	1 000-shot fault detection/%	Training time/s	Testing time/ms
Ours	64.71	—	92.95	—	182.72	27.060
GMM	16.67	16.67	83.33	16.67	190.50	3.700
GMM+PCA	16.67	16.30	83.33	81.95	1.27	0.001
LAD	16.67	16.81	16.67	68.52	160.10	0.050
LAD+PCA	16.67	16.50	83.33	83.33	2.06	0.001
BNB	16.67	19.00	16.67	76.12	1.25	0.400
BNB+PCA	17.05	17.58	83.33	83.33	2.19	0.001
GNB	16.67	64.15	16.67	81.62	0.87	1.260
GNB+PCA	16.67	23.68	16.83	83.25	1.93	0.001
SVM	16.67	67.06	16.67	96.30	22.54	20.500
SVM+PCA	16.67	24.95	16.67	83.33	3.10	0.001
DT	16.67	17.35	16.67	74.23	11.53	0.040
DT+PCA	16.67	100	16.73	82.03	3.14	0.001
XGBoost	16.67	26.60	16.67	83.45	25.93	0.040
XGBoost+PCA	16.67	64.47	16.67	83.27	1.94	0.001

2.2 Turbine gearbox case study

2.2.1 Description of test bench

This case uses wind turbine gearbox fault data collected from the power transmission testing platform of the national renewable energy laboratory (NREL) to verify the proposed method. The turbine of this platform is a 750 kW, three-bladed, upwind turbine, and the turbine generator operates at 1 800 r/m and 1 200 r/m. The turbine is equipped with three types of single faults on key components such as the planetary carrier, low-speed shaft, intermediate speed shaft, high-speed shaft, and bearings, namely scuffing, overheating, and fretting corrosion. Compound faults include compound fault 1 (fretting corrosion, scuffing, polishing wear), compound fault 2 (assembly damage, scuffing, dents), compound fault 3

(assembly damage, dents), and compound fault 4 (scuffing and polishing, fretting corrosion). In this paper, faults such as scuffing, assembly damage, dents, and fretting decay are set as benchmark single faults. While fretting, decay, scuffing, polishing wear, and assembly damage, scuffing, and dents are considered compound faults.

The obtained data is collected by acceleration sensors. Assuming the baseline faults are s^1 , s^2 , and s^3 , and the compound faults are represented by $s^0 \oplus s^1 \oplus s^2$ and $s^0 \oplus s^1 \oplus s^7$, respectively. Similarly, in this section, 1 001 samples were selected as the benchmark failure samples, among which 1 was used as the training sample and 1 000 were used as the test data. 1 000 samples were taken for compound faults, all of which were used as test data. The details of the compound fault dataset are shown in Table 4.

Table 4 Validation dataset for gearbox condition diagnosis methods with zero compound fault samples

Condition	Fault type	Description	Number of training samples	Number of testing samples
Normal	Seen	Nor	1 000	1 000
Scuffing	Seen	Sc	1	1 000
Assembly damage and dents	Seen	Add	1	1 000
Fretting corrosion	Seen	Fc	1	1 000
Fretting corrosion, scuffing, polishing wear	Unseen	CF1	0	1 000
Assembly damage, scuffing, dents	Unseen	CF2	0	1 000

2.2.2 Results and discussion

The optimization process of the proposed method for

CNN, ResNet, and efficient former feature encoders under single reference fault samples and zero compound fault

samples of NREL turbine gearbox is shown in Fig.5. It can be seen that the loss function constructed by introducing temporal signals and fault semantic knowledge is also applicable to turbine gearboxes, and can effectively optimize three typical network feature encoders under single

and compound fault samples. From the process of loss function variation, it can be seen that the loss function of the efficient former feature encoder converges quickly, followed by ResNet, which makes them more prone to overfitting compared to CNN feature encoders.

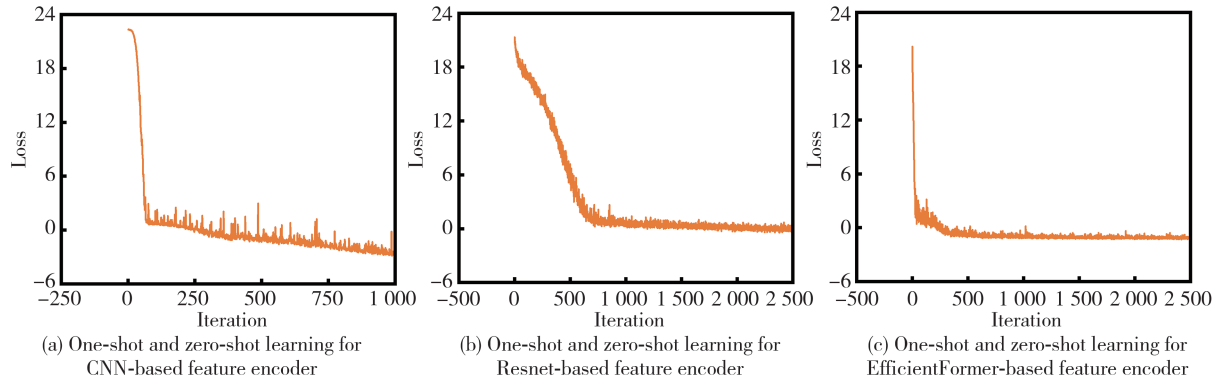


Fig. 5 Optimization process of NREL turbine gearbox single fault samples and compound fault samples with three feature encoders obtained from designed losses

Fig.6 presents ten validation results of using three types of feature encoders obtained through training for gearbox fault diagnosis. It can be seen that the fault diagnosis model based on CNN has the best performance in identifying single faults and compound faults. Compared to ResNet and EfficientNet, CNN has a simpler network architecture and lower risk of overfitting. This renders CNN-based fault diagnosis models less susceptible to overfitting during training scenario design. This characteristic is precisely the

reason for its outstanding fault diagnosis performance. Consequently, overfitting constitutes a critical factor affecting the performance of the proposed method in practical applications. The key research challenge lies in preventing training overfitting under conditions where only a minimal number of fault samples are available. A future direction for enhancing the method involves incorporating additional physical information and physical mechanisms to mitigate overfitting.

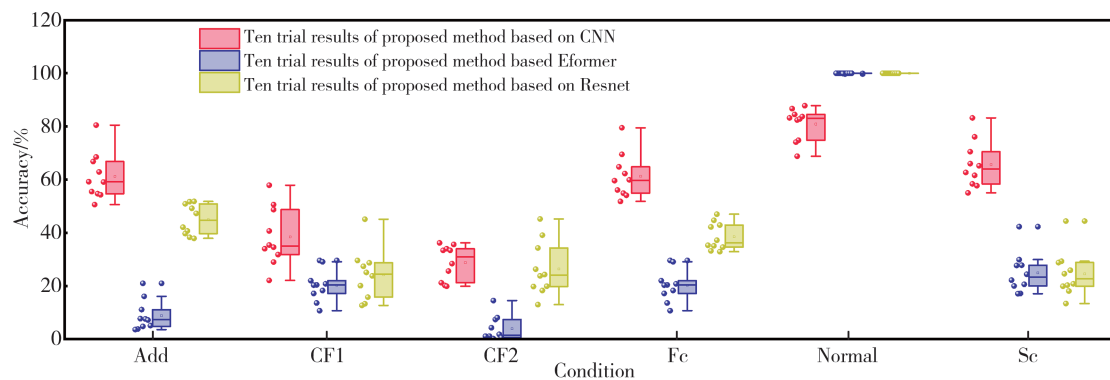


Fig. 6 Ten trail results for normal state, single fault, and compound fault using three training-obtained feature encoders combined with the proposed fault diagnosis algorithm

To further verify the progressiveness of the method in this paper, the advanced deep neural networks vision transformer (ViT), DenseNet, ResNet, CNN, and advanced machine learning models XGBoost, decision tree (DT), SVM are selected as the comparison models. The comparison models were trained using supervised learning methods, where each model was trained on a dataset containing 1 000 samples for each type of fault state, and tested using a dataset containing 1 000 samples for each type of fault state. The comparison results of multiple comparison models are

shown in Table 5.

It can be seen that the proposed method achieves better performance than advanced machine learning models such as XGBoost, SVM, and DT by utilizing a single sample of a single fault and zero compound fault samples. Compared with other advanced deep network models, the proposed method achieves satisfactory performance. It should be pointed out that the proposed method requires few fault samples, which is more advantageous in engineering practice.

Table 5 Comparison results of proposed method with other supervised learning methods

Method	FDI/%	FDE/%	Training time/s	Testing time/ms
Ours	57.57	82.89	186.05	26.60
ViT	34.72	99.67	4 508.70	8.39
DenseNet	33.83	99.67	784.90	2.46
Resnet	34.56	99.50	785.03	2.65
CNN	35.78	99.39	684.20	2.36
XGBoost	40.42	83.33	43.62	0.06
Decision tree	35.58	79.48	709.44	0.05
SVM	37.15	83.33	88.87	35.60
GNB	38.26	83.33	0.911	0.30
BNB	16.32	60.65	0.64	0.20
LAD	16.53	83.33	339.20	0.05
GMM	16.67	50.47	194.60	7.50

3 Conclusions

This paper proposes three novel contrastive learning loss functions that combine physical information of the strong temporal nature of gearbox vibration signals with semantic knowledge between single and compound faults. By combining normal state data with extremely rare single fault data, the established loss function can effectively train different feature encoders under zero compound fault samples and extremely rare single fault samples. Based on the trained feature encoders, an end-to-end diagnostic algorithm for single and compound faults is proposed. The proposed method was validated on two gearbox experimental platforms. From the comparison results between the proposed method and advanced machine learning-based approaches, it can be seen that the performance of the proposed method outperforms other advanced machine learning in both fault detection and diagnosis, and which provides a feasible path for constructing intelligent fault detection and diagnosis methods for mechanical equipment under zero compound fault sample and extreme sample imbalance conditions.

Acknowledgement

This study was supported by the Natural Science Foundation of Fujian Province (No. 2024J08064), the Natural Science Foundation of Xiamen (No. 3502Z202471042).

Declaration of conflicting interests

The authors have no conflicts of interest related to this publication.

References

- [1] YU J B, WANG S Y, WANG L, et al. Gearbox fault diagnosis based on a fusion model of virtual physical model and data-driven method. *Mechanical Systems and Signal Processing*, 2023, 188: 109980.
- [2] XIA J Y, HUANG R Y, CHEN Z Y, et al. A novel digital twin-driven approach based on physical-virtual data fusion for gearbox fault diagnosis. *Reliability Engineering & System Safety*, 2023, 240: 109542.
- [3] ZHOU P, CHEN S Q, HE Q B, et al. Rotating machinery fault-induced vibration signal modulation effects: a review with mechanisms, extraction methods and applications for diagnosis. *Mechanical Systems and Signal Processing*, 2023, 200: 110489.
- [4] JIANG W X, WU J, ZHU H P, et al. Paired ensemble and group knowledge measurement for health evaluation of wind turbine gearbox under compound fault scenarios. *Journal of Manufacturing Systems*, 2023, 70: 382-394.
- [5] YANG H B, WANG Z Y, XU M, et al. Improved deep transfer learning and transmission error based method for gearbox fault diagnosis with limited test samples. *Mechanical Systems and Signal Processing*, 2025, 230: 112593.
- [6] REN Z J, LIN T T, FENG K, et al. A systematic review on imbalanced learning methods in intelligent fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 3508535.
- [7] PAN T, CHEN J, ZHANG T, et al. Generative adversarial network in mechanical fault diagnosis under small sample: a systematic review on applications and future perspectives. *ISA Transactions*, 2022, 128: 1-10.
- [8] LIU P J, GUO Z C, HE L, et al. Fault diagnosis method of AC motor rolling bearing based on heterogeneous data fusion of current and infrared image. *Journal of Measurement Science and Instrumentation*, 2024, 15(4): 558-570.
- [9] BAI Y H, CHENG W D, WEN W G, et al. Application of time-frequency analysis in rotating machinery fault diagnosis. *Shock and Vibration*, 2023, 2023: 9878228.
- [10] PENG Y Y, QIAO W, CHENG F Z, et al. Wind turbine drivetrain gearbox fault diagnosis using information fusion on vibration and current signals. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 3518011.
- [11] WANG S B, CHENG C Y, ZHOU J H, et al. Reassignment-enable reweighted sparse time-frequency analysis for sparsity-assisted aeroengine rub-impact fault diagnosis. *Mechanical Systems and Signal Processing*, 2023, 183: 109602.
- [12] CHEN X W, FENG Z P. Iterative generalized time-frequency reassignment for planetary gearbox fault diagnosis under nonstationary conditions. *Mechanical Systems and Signal Processing*, 2016, 80: 429-444.
- [13] SYED S H, MURALIDHARAN V. Feature extraction using Discrete Wavelet Transform for fault classification of planetary gearbox-A comparative study. *Applied Acoustics*, 2022, 188: 108572.
- [14] VASHISHTHA G, CHAUHAN S, KUMAR S, et al.

- Intelligent fault diagnosis of worm gearbox based on adaptive CNN using amended gorilla troop optimization with quantum gate mutation strategy. Knowledge-Based Systems, 2023, 280: 110984.
- [15] ZHANG K, TANG B P, DENG L, et al. A hybrid attention improved ResNet based fault diagnosis method of wind turbines gearbox. Measurement, 2021, 179: 109491.
- [16] ZHANG J B, LI G Y, LIU R J, et al. Compound fault transfer diagnosis of gearboxes based on improved transformer network under small datasets and variable working conditions. IEEE Transactions on Instrumentation and Measurement, 2025, 74: 3543018.
- [17] VALLACHIRA S, ORKISZ M, NORRLÖF M, et al. Data-driven gearbox failure detection in industrial robots. IEEE Transactions on Industrial Informatics, 2020, 16(1): 193-201.
- [18] CHEN Y G, JIA S L, ZHU J, et al. Fault diagnosis for on-board equipment of train control system based on BERT+CNN_BiLSTM. Journal of Measurement Science and Instrumentation, 2024, 15(1): 120-127.
- [19] SINGH V, GANGSAR P, PORWAL R, et al. Artificial intelligence application in fault diagnostics of rotating industrial machines: a state-of-the-art review. Journal of Intelligent Manufacturing, 2023, 34(3): 931-960.
- [20] KARNIADAKIS G E, KEVREKIDIS I G, LU L, et al. Physics-informed machine learning. Nature Reviews Physics, 2021, 3(6): 422-440.
- [21] NI Q, JI J C, HALKON B, et al. Physics-Informed Residual Network (PIResNet) for rolling element bearing fault diagnostics. Mechanical Systems and Signal Processing, 2023, 200: 110544.
- [22] YANG Y K, WANG X F, LI J F, et al. A data-physic driven method for gear fault diagnosis using PINN and pseudo-dynamic features. Measurement, 2024, 236: 115124.
- [23] YU Y N, TANG L H, LIU Z P, et al. A novel bearing fault data generation strategy combining physical modeling and CycleGAN variant for fault diagnosis without real samples. IEEE Transactions on Instrumentation and Measurement, 2025, 74: 3505717.
- [24] SUN D Y, LI Y B, LIU Z, et al. Physics-inspired multimodal machine learning for adaptive correlation fusion based rotating machinery fault diagnosis. Information Fusion, 2024, 108: 102394.
- [25] ZHU P, MA S, HAN Q K, et al. Deep contrastive transfer learning for rotating machinery fault diagnosis. IEEE Transactions on Instrumentation and Measurement, 2025, 74: 3501510.
- [26] ZHANG S Y, WEI H L, DING J L. An effective zero-shot learning approach for intelligent fault detection using 1D CNN. Applied Intelligence, 2023, 53(12): 16041-16058.
- [27] XU J, ZHOU L, ZHAO W H, et al. Zero-shot learning for compound fault diagnosis of bearings. Expert Systems with Applications, 2022, 190: 116197.
- [28] LIAO W J, WU L K, XU S H, et al. A novel zero-shot learning method with feature generation for intelligent fault diagnosis. IEEE Transactions on Industrial Informatics, 2025, 21(4): 3386-3395.
- [29] YANG J L, WANG C D, WEI C A. A novel Brownian correlation metric prototypical network for rotating machinery fault diagnosis with few and zero shot learners. Advanced Engineering Informatics, 2022, 54: 101815.
- [30] MA C Y, WANG X Z, LI Y B, et al. Broad zero-shot diagnosis for rotating machinery with untrained compound faults. Reliability Engineering and System Safety, 2024, 242(C): 109704.

零样本下基于故障耦合语义的齿轮箱复合故障诊断方法

苏兴亮¹, 许可¹, 张楷兴¹, 廖建彬², 黎国强², 闫锦^{2*}

1. 龙岩烟草工业有限责任公司, 福建 龙岩 364021;

2. 集美大学 轮机工程学院, 福建 厦门 361021

摘要: 具有复杂结构的齿轮箱是旋转机械中的一类重要设备, 其故障发生具有偶发性与耦合性。然而, 工程实际中齿轮箱故障数据极少, 针对其复合故障的样本数量甚至为零。因此, 研究极少故障样本与零复合故障样本下故障诊断建模方法具有重要意义。本文通过时序监测信号间的相似性与单一故障之间互异性的先验知识, 以及复合故障由单一故障组成的先验知识, 设计了一种全新的单个单一故障样本和零复合故障样本下的损失函数, 实现了特征编码器的有效优化。在此基础之上, 构建齿轮箱单一故障与复合故障的诊断算法, 实现了齿轮箱各种状态的有效诊断。本文分别采用两个齿轮箱故障实验平台对本文方法进行验证, 结果表明所提方法能够在单个单一故障样本与零复合故障样本下, 构建获得齿轮箱故障诊断模型。

关键词: 齿轮箱; 复合故障; 零样本; 故障语义; 物理信息网络; 对比学习

引用格式: SU Xingliang, XU Ke, ZHANG Kaixing, et al. Zero-shot compound fault diagnosis of gearboxes via fault-coupled semantics. Journal of Measurement Science and Instrumentation, 2026, 17(2): 320-330. DOI: 10.62756/jmsi.1674-8042.2026027