Expert Systems With Applications 233 (2023) 120875

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A label information vector generative zero-shot model for the diagnosis of compound faults☆

Juan Xu [a,b], Kang Li [a,b], Yuqi Fan [b], Xiaohui Yuan [c,*]

[a] Key Laboratory of Knowledge Engineering with Big Data, Hefei University of Technology, Ministry of Education, China
[b] Hefei University of Technology, China
[c] University of North Texas, Denton, TX, 76203, USA

ARTICLE INFO

Keywords:
Compound fault diagnosis
Generative adversarial networks
Generative zero-shot learning

ABSTRACT

Diagnosis of compound faults remains a challenge owing to the coupling of fault characteristics and the exponential increment of the number of possible fault types. Current compound faults diagnostic methods often require a large number of training data for each type of compound fault. In real-world scenarios, training data of compound faults are usually difficult to acquire and sometimes even inaccessible. In contrast, single fault samples are much easier to obtain. Thus in this paper inspired by the idea of zero-shot learning, we present a novel label information vector generative zero-shot model to identify unknown compound faults, using only single fault samples as the training set. This model comprises several modules, namely label information vector (LIV) definition, feature extractor, and generative adversarial modules, respectively responsible for representing the prior knowledge of specific class labels for the single fault and compound fault, extraction of fault features, and mapping the relationship between the fault features and the fault LIVs. By adversarial training between the samples and LIVs of single faults, the model can generate compound fault features using the compound fault LIVs. Thus the unknown compound faults are identified by measuring the distance between the features extracted from the testing compound fault samples and the generated features from LIVs. The proposed method is evaluated on a self-built experimental platform. The results demonstrate that without any compound fault samples in the training set, the compound fault classification accuracy of the model reaches 78.10%.

1. Introduction

Bearings are important basic parts of complex rotating machines. Fault diagnosis of bearings is an important aspect of ensuring equipment safety. Compound fault means that more than two interrelated and interacting faults occur at the same time. If there exist coupling among some components, the number of compound fault grows exponentially. These compound faults exhibit different characteristics which are coupled together and hidden in the collected vibration signals. Therefore, the diagnosis of compound faults is the most challenging issue in the fault diagnosis of bearings (Deng et al., 2022; Dibaj et al., 2021; Li et al., 2015; Xu & Li, 2021).

Traditional compound fault diagnosis methods mainly include qualitative experience-based (Chatti et al., 2014; Ubar et al., 2012), analytical model-based (Mhamdi et al., 2013; Piacentino & Talamo, 2013), and signal analysis-based (Li et al., 2019; Tang et al., 2020) methods. These methods require empirical knowledge and engineering experience from experts, which is difficult to apply in industrial scenarios.

The development of machine learning, especially deep learning, has prevalently emerged in the field of compound fault diagnosis (Gao et al., 2022; Guo et al., 2019; Zhao et al., 2019). Consequently, several learning model-based methods have been proposed to automatically learn fault features and identify compound faults from raw sensory data instead of experts. However, these methods require a large number of labeled data for each type of compound fault for model training. In realistic scenarios, it is difficult and sometimes impossible to obtain samples of every compound fault type, which limits the application of such methods (Xing et al., 2022).

Given that the single fault samples are relatively easy to obtain, studies have been conducted on decoupling learning models to apply single fault samples to train models for the identification of compound faults. The idea entails training a decoupled classifier, from the single

☆ This work was supported in part by the National Key Research And Development Plan under Grant 2018YFB2000505, in part by the Key Research and Development Plan of Anhui Province under Grant 202104a04020003, and in part by the Fundamental Research Funds for the Central Universities under Grant PA2021KCPY0045.

* Corresponding author.
E-mail addresses: xujuan@hfut.edu.cn (J. Xu), 2020171187@mail.hfut.edu.cn (K. Li), yuqi.fan@hfut.edu.cn (Y. Fan), xiaohui.yuan@unt.edu (X. Yuan).

https://doi.org/10.1016/j.eswa.2023.120875
Received 30 May 2022; Received in revised form 15 June 2023; Accepted 15 June 2023
Available online 28 June 2023
0957-4174/© 2023 Elsevier Ltd. All rights reserved.

fault samples, to decouple compound faults into a combination of multiple single faults, then classify the compound faults by identifying multiple single faults. Huang et al. (2019) designed an intelligent compound fault diagnosis network based on a deep capsule network and ensemble learning. Liang et al. (2019) combine the convolutional neural network with the wavelet transform and multi-label classification (WT-MLCNN) to diagnose compound faults for the gearbox. However, these decoupling learning models need several independent loss functions for each fault class, which greatly increases model parameters and thereby aggravates its complexity and computational cost.

In contrast, zero-shot learning (ZSL) models identify unknown classes using data from the known classes, supplemented by prior knowledge to represent the specific class labels of the unknown classes, to train a certain learning model (Rahman et al., 2020; Xu et al., 2022a; Zhuo & Ge, 2021). Thus the classification problem of unknown classes is transformed into a supervised learning problem, without additional data collection, complex model combinations, and increased computational effort.

Inspired by the idea of ZSL, using the samples of single faults to build a zero-shot learning compound fault diagnosis model becomes a promising attempt to enable the compound fault diagnosis with no or extremely scarce examples.

To the best of our knowledge, there are limited studies on the zero-shot learning model for fault diagnosis of bearings. Xing et al. (2022) proposed a label description space embedded model for zero-shot diagnosis of compound faults (LDS-IFD). In our previous works (Xu et al., 2022b), we proposed a zero-shot learning compound fault diagnosis model of bearings (ZLCFDM). The two studies both established a label information vector (LIV) space to construct the relationship between single faults and compound faults, then embedded all the fault features and label information vectors into a certain space and finally classified the compound fault samples according to the similarity measurement.

Noted that the aforementioned two models embed high-dimensional fault features into a low-dimensional label information vector space. When the distance between different fault features is very close, it is difficult for the models to find the fault label information vectors closest to the fault feature, thus the classification ability of the models will be seriously degraded (Shigeto et al., 2015).

Different from the aforementioned embedded ZSL models, we attempt a new perspective from the generative model to tackle the dimensional constraint problem of embedded models. Specifically, through adversarial training of the single fault samples, the model focuses on learning the mapping from the signal-derived fault LIVs to generate the correspondingly high-dimensional fault features, such that they have a matched dimensionality with the feature extracted from the original samples. Then the trained model generates features of compound faults by using the corresponding LIVs. The compound faults are identified by computing the Euclidean distance between the extracted features from samples and the generated features from signal-derived LIVs.

The main contributions of this article are summarized as follows.

1. We propose a novel generative zero-shot compound fault diagnosis framework using only single fault samples. The training process of our model generates high-dimensional fault features from the signal-derived fault LIVs via adversarial training of both for facilitating the classification of the model. The trained model identifies the compound faults by computing the Euclidean distance between the extracted features from examples and the generated features from signal-derived LIVs.
2. A new fault LIV definition method is designed to obtain a representative set to bridge the known single faults and the unknown compound faults. The single fault LIVs are designed from thresholding peaks of the vibration signals, whereas the LIVs of the compound faults are derived from the fault semantics of the single fault. This enables learning from single-fault examples and classifying the compound faults. The feasibility of our method is confirmed by the experimental results.

The remainder of this article is organized as follows. Section 2 introduces the related work. Section 3 presents our proposed zero-shot model. Section 4 discusses the experimental results including comparison studies with the state-of-the-art methods. Section 5 concludes this paper with a summary.

2. Related work

ZSL classification is an effective method for solving the problem of missing class labels when the classes of the training set and the testing set are disjoint (Feng & Zhao, 2020; Lampert et al., 2013). The basic idea of zero-shot learning is to use some data of known classes, supplemented by relevant common knowledge information or prior knowledge as attribute labels (hereinafter referred to as label information vector), to train a certain learning model and finally identify the data of unknown classes (Demirel et al., 2019; Ding et al., 2018).

In ZSL models, label information vectors are a kind of prior knowledge to represent the specific categories of objects. Thus LIVs that bridge the known and unknown classes are a core problem of ZSL. The quality of the LIV greatly impacts the performance of models. Previous LIVs definition approaches mainly include manual definition-based (Chen et al., 2019; Lampert et al., 2013) and learning-based (Mikolov et al., 2013; Pennington et al., 2014).

Manual definition-based methods design each dimension of LIVs by the prior knowledge to express certain attributes of the class. Lampert et al. (2013) proposed an object attributes-based approach to defining semantic vectors, where high-level descriptions of objects such as color, shape, and size are used. Chen et al. (2019) used the attributes such as fault signal mean value, standard deviation, and peak-to-peak value to define the fault semantic vectors. Both of these methods treat each attribute as one dimension of the semantic vector and use one-hot encoding to indicate whether each attribute is present in the description of the fault class. Ultimately, these attributes are defined as the semantic vectors for the faults. However, such definitions of LIV are not specific or clear enough to capture all the characteristics of bearing faults and may suffer from subjectivity or inconsistency in practice.

Correspondingly, learning-based methods use some pre-trained models (e.g., Word2Vec and GloVe), to learn the vector representation for each class as its LIVs. Mikolov et al. (2013) proposed an improved Skip-gram model that learns the mapping to represent the input center word as a vector and the surrounding words as other vectors, which can capture semantic and contextual relationships between words. Pennington et al. (2014) proposed a word representation tool called global vectors for word representation to form the semantic vector. The tool is based on a global word frequency statistical method, which represents a word as a vector of real numbers that captures semantic characteristics such as similarity and analogy between words. All of these vectors form a semantic vector space. However, if we use the learning-based method to define the fault LIV, similar words of fault classes, e.g., inner ring faults, outer ring faults, and rolling body faults are converted into the LIV of the classes by the pre-trained models. The high similarity of their words makes them close to each other in the semantic space, which is difficult to distinguish between them and greatly degrades the classification ability of the model.

All in all, the aforementioned two LIV definition methods are inapplicable fault classifications because they have drastically different label information from vibration signals collected from machinery. More importantly, the mapping of machinery faults to a label information space is ill-posed. The LIV of vibration signal for fault diagnosis has not yet been well defined. Hence, this paper proposes a novel fault LIV definition method for fault diagnosis.

3. Methodology

We formulate the compound fault diagnosis problem as a ZSL problem. The training set of single fault $D_s = \{x^i, y^i, S^i \mid x^i \in x_s, y^i \in Y_s, S^i \in S_s\}_{i=1}^N$ includes K classes and N samples, where $x_s = \{x^1, x^2, \dots, x^N\}$ and $Y_s = \{c_1, c_2, \dots, c_K\}$ are the training samples and labels, S^i is the LIV of the ith training sample with dimension $C \times 1$, $S_s = \{S^1, S^2, \dots, S^N\}$ is the label set of single faults.

The testing set for vibration signal of compound faults is given by $D_u = \{\hat{x}^i, \hat{y}^i, \hat{S}^i \mid \hat{x}^i \in x_u, \hat{y}^i \in Y_u, \hat{S}^i \in S_u\}_{i=1}^M$, including L classes and M samples, where $x_u = \{\hat{x}^1, \hat{x}^2, \dots, \hat{x}^M\}$ is the testing samples, $Y_u = \{\hat{c}_1, \hat{c}_2, \dots, \hat{c}_L\}$ represents the class label set, $\hat{S}^i$ is the LIV of the ith testing sample with dimension $C \times 1$, $S_u = \{\hat{S}^1, \hat{S}^2, \dots, \hat{S}^M\}$ represents the label information set of all compound faults in the set space S, $S = S_s \cup S_u$. Notably, Y_u as an unknown class of the compound faults, is not involved in the model training and testing. It is only used to evaluate the accuracy of classification results of the compound faults. The dataset and class sets satisfy

$$I(p(x_s); p(x_u)) = 0; \quad Y_s \cap Y_u = \emptyset; \quad \text{and } \varphi(S_s) = S_u$$

where x_s and x_u follow different distributions and Y_s and Y_u are disjoint. The mutual information of distributions $p(x_s)$ and $p(x_u)$ is calculated with $I(\cdot)$. The compound label S_u is obtained based on function $\varphi(\cdot)$ with single label S_s.

In the training phase, a model $F(x_s, S_s; \theta)$ is trained with examples x^i and labels S^i of single faults, that learns a mapping function

$$F(x_s, S_s; \theta) : (x^i, S^i) \rightarrow y^j \tag{1}$$

where $\theta = arg\min \sum_i^N Dis(x^i, S^i)$ and $Dis(\cdot)$ is the distance metric function. Note that compound fault samples are not used in the training process.

In the testing phase, samples $\hat{x}^j$ and labels $\hat{S}^j$ of the compound faults are input into the model, which makes a prediction of the fault class $\hat{c}_k$, $\hat{c}_k \in Y_u$:

$$F(x_u, S_u; \theta) : (\hat{x}^j, \hat{S}^j) \rightarrow \hat{c}_k \tag{2}$$

3.1. Model structure

Fig. 1 depicts our proposed label information vector generative zero-shot method that is composed of five parts: data preprocessing module, label information vector definition module, feature extractor module, generative adversarial module, and classifier module.

To extract the useful information from the vibration signals, Wavelet Transform (WT) is used to derive the details in the data preprocessing module:

$$WT(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} \phi(t) * \psi(\frac{t-b}{a}) dt \tag{3}$$

where $\psi(t) = e^{(-t^2/2)} cos(5t)$ and a is the scale factor and b is the translation factor, $\phi(t)$ is the time-domain vibration signal sequence. We use WT to convert every set of 256 time-domain data points into a 64×64 2D binary time–frequency domain image. The center rate is 0.8125.

The label information vector is defined based on the vibration signals of a single fault. Assume that we have a set of C data points of the vibration signals (e.g., the vibration signal of the outer ring fault)

$$\phi = (x_1^i, x_2^i, \dots, x_k^i, \dots, x_C^i). \tag{4}$$

where x_k^i is the kth data point, i is the fault class. The signal ϕ contains samples that are greater than one period of the vibration signal to ensure that the LIV contains the effects of fault. The LIV of a single fault S^i is

$$S^i = \{a_1^i, a_2^i, \dots, a_k^i, \dots, a_C^i\} \tag{5}$$

where a_k^i is the kth dimension of the S^i. We select C data points of the original vibration signal to obtain the maximum from the data points x_k^i in all single faults ϕ as the threshold μ, i.e., $\mu = \max(\phi)$. We obtain the LIV S^i with C dimension for any single fault as follows:

$$a_k^i = \begin{cases} 0, & x_k^i \le \mu/5 \\ 1, & x_k^i \ge \mu/5, x_k^i \le 2\mu/5 \\ 2, & x_k^i \ge 2\mu/5, x_k^i \le 3\mu/5 \\ 3, & x_k^i \ge 3\mu/5, x_k^i \le 4\mu/5 \\ 4, & x_k^i \ge 4\mu/5, x_k^i \le \mu \end{cases} \tag{6}$$

Similarly, we obtain the LIV set of single faults $S_s = \{S^1, S^2, \dots, S^N\}$.

Combining the theoretical relationships between the compound fault and its corresponding single faults, the LIV set of the compound fault S_u is expressed as:

$$S_u = \{\hat{S}^{1,2}, \hat{S}^{1,3}, \dots, \hat{S}^{1,2,\dots,R}\} \tag{7}$$

$$\hat{S}^{i,j,\dots,R} = \left\{ a_k \mid a_k = \max(a_k^i, a_k^j, \dots, a_k^R), a_k^i \in S^i \right\} \tag{8}$$

where i, j, R belong to different fault classes.

The feature extraction is achieved with a CNN. We use labeled single fault samples to train the CNN feature extractor and obtain the mapping function $f(\cdot)$ for extracting single fault features v_s, $v_s = \{v_s^i \mid v_s^i = f(WT(a, b); \theta_1)\}$. The unlabeled compound faults are processed to obtain the compound fault features v_u, $v_u = \{v_u^j \mid v_u^j = f(WT(a, b); \theta_1)\}$.

The generative adversarial module of our method consists of a discriminator D and a generator G that performs a zero-sum game mechanism. The discriminator D maximizes the loss function

$$\begin{aligned} L_D = {} & E_{v \sim p_r}[D_w(v)] - E_{\hat{v} \sim p_g}[D_w(G_{\theta_2}(S, z))] \\ & - \lambda E_{\tilde{v} \sim p_{\tilde{v}}}[(\|\Delta_{\tilde{v}} D_w(\tilde{v})\|_2 - 1)^2] \end{aligned} \tag{9}$$

while network G minimizes

$$L_G = E_{\hat{v} \sim p_g}[D_w(G_{\theta_2}(S, z))] \tag{10}$$

where $E(\cdot)$ denotes expectation, p_r is the data distribution of extracted fault features, p_g is $\hat{v} = G_{\theta_2}(S, z)$, the generated fault features data distribution, $p_{\tilde{v}}$ samples uniformly from the data distribution p_r and generator distribution p_g. z is the Gaussian distribution noise, v and $\hat{v}$ are the fault feature of the real sample and the generated sample, respectively, $\tilde{v}$ is uniformly sampled from v and $\hat{v}$, S is a LIV, λ is the gradient penalty coefficient.

The generator uses the single fault label information S_s and Gaussian noise z to generate single fault features $\hat{v}_s$, $\hat{v}_s = \{\hat{v}_s^i \mid \hat{v}_s^i = G_{\theta_2}(S^i, z), S^i \in S_s\}$ and S^i is the LIV of the ith single fault sample. The discriminator decides the real single fault features v_s and the generated features $\hat{v}_s$ to guide the generator to create high-quality single fault features. The trained generator uses compound fault label information S_u and Gaussian noise z as inputs to generate compound fault features $\hat{v}_u$, $\hat{v}_u = \{\hat{v}_u^j \mid \hat{v}_u^j = G_{\theta_2}(\hat{S}^j, z), \hat{S}^j \in S_u\}$, where $\hat{S}^j$ is the LIV of the jth compound fault sample.

The classification maps $F(G(S, z), v; \theta)$ of fault features v extracted by FE to the generated fault features by the generator. The model optimizes parameter θ and matches the fault features v with the generated fault features $G(S, z)$.

Using the fault features generated from the compound fault label information as the center point, the Euclidean distance between the compound fault features of the test samples extracted by the FE and the center point is calculated, and the test sample features of the compound fault are labeled as the class with the nearest center point. The formula is as follows:

$$Y_u = \left\{ \hat{c}_k \mid k = H(min(Dis(v_u^i, \hat{v}_u^j))), i = 1, 2, \dots, M \right\} \tag{11}$$

$$H(d_k) = k, \qquad \text{if } \ d_k = min(Dis(v_u^i, \hat{v}_u^j)) \tag{12}$$

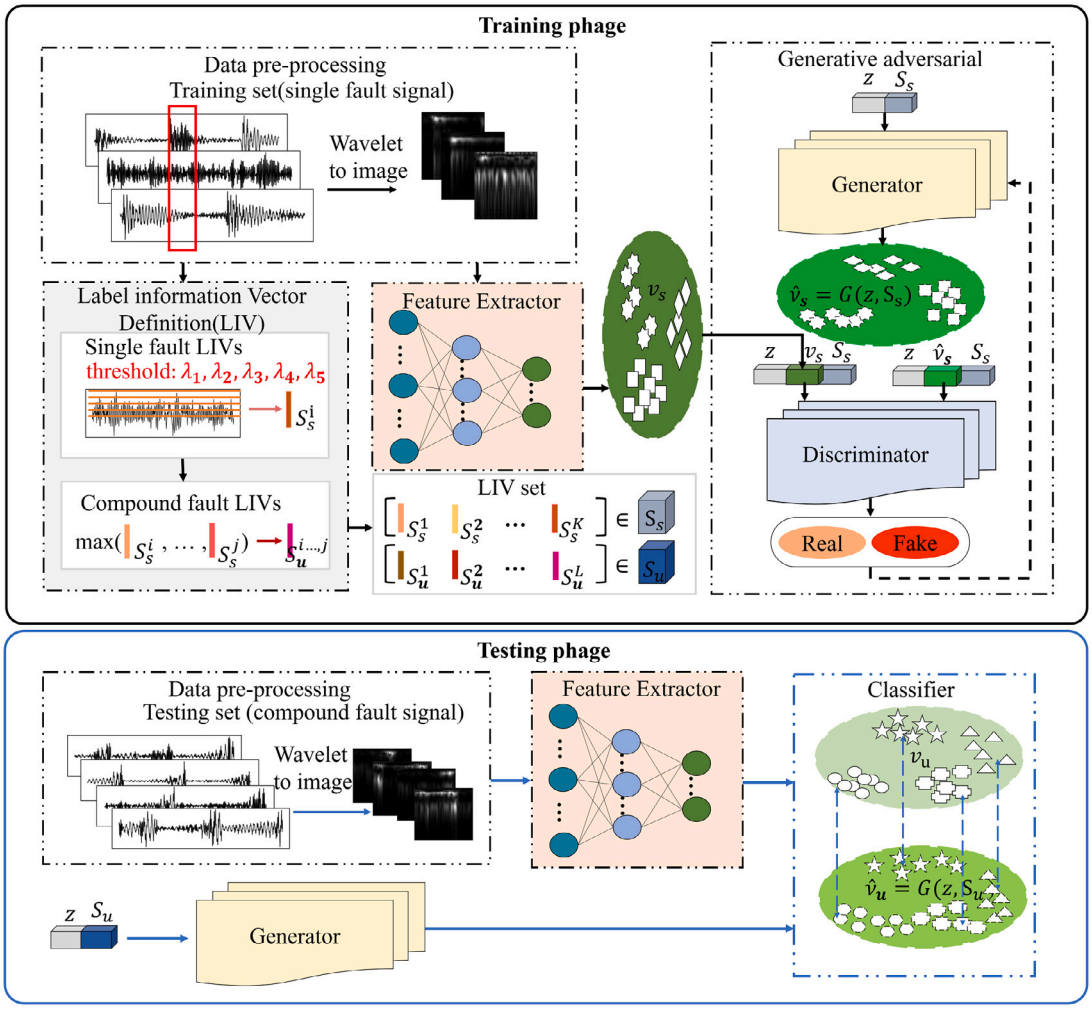


Fig. 1. The architecture of our proposed method.

where $Dis(v_u^i, \hat{v}_u^j) = \{d_j \mid d_j = \sqrt{\left(v_u^i - \hat{v}_u^j\right)^2}, j = 1, 2, \ldots, L\}$, v_u^i is the compound fault sample features extracted by the feature extractor, $\hat{v}_u^j$ is the compound fault features of different classes generated by G. $H(\cdot)$ is to determine the categories of the unknown compound fault samples, by calculating the closest distance between the features extracted from the unknown compound fault samples v_u^i and the generated compound fault feature $\hat{v}_u^j$. $\hat{c}_k$ is the label of the compound fault class.

3.2. Objective function

The training of our model includes the training of a feature extraction module and a generative adversarial module. The feature extraction module is trained to distinguish fault features from the labeled single fault samples. The loss function L_1 of the feature extraction module is

$$L_1 = -\sum_{i}^{N} y^i \log p_i + (1 - y^i) \log (1 - p_i) \tag{13}$$

where p_i is the predicted value of the feature extractor module.

The generative adversarial module obtains a generator that produces high-quality fault features and the loss function integrates L_D and L_G in Eqs. (9) and (10):

$$\begin{aligned} L_2 = {} & E_{\hat{v}\sim p_g}[D_w(G_{\theta_2}(S, z))] - E_{v\sim p_r}[D_w(v)] \\ & + \lambda E_{\tilde{v}\sim p_{\tilde{v}}}[(\|\Delta_{\tilde{v}} D_w(\tilde{v})\|_2 - 1)^2] \\ & + E(\|v - G_{\theta_2}(S, z)\|_2) \end{aligned} \tag{14}$$

The first term is the difference between the expectation of $D_w(v)$ and the expectation of $D_w(G_{\theta_2}(S, z))$ to make the real fault features $D_w(v)$ and the generated fault features $D_w(G_{\theta_2}(S, z))$ as close as possible. The second term ensures that $D_w(G_{\theta_2}(S, z))$ is close to $D_w(v)$ but not exceeding $D_w(v)$, we add a penalty term for stabilizing the gradient. The third term is the mean squared error, which reduces the distance between the real sample features and the generated sample features.

3.3. Training and application of our proposed method

Training and application of our proposed model consist of two phases: constructing the fault LIVs of single and compound faults, correspondingly training the model using the single faults samples, and testing the fault samples of compound faults. The algorithm is presented in Table 1.

Training: The vibration time-domain signal of a single fault is transformed into a 2D time–frequency image via WT. We use the feature extractor to extract high-dimensional fault features of the time–frequency image. Based on the proposed LIV definition method, the LIVs of the single faults and compound faults are devised from the vibration samples of the single faults. The generator uses the LIV of the single faults to generate fault features of the single faults. Thereafter, the generated and real fault features of the single faults are sent to the discriminator to distinguish true or false. Through adversarial training of the generator and discriminator, the generator learns the accurate mapping relationship from fault LIV to fault features.

Table 1
Algorithm of the proposed method.

Require: N, K, I, θ_1, θ_2, β_1, β_2, λ, n_{critic}, m, α
for I epochs **do**
Extract training features $\{(v_s^1, y^1), \ldots, (v_s^n, y^n)\}$
  $\nabla_\varphi L_1 \leftarrow \nabla_\varphi \sum_{i=1}^{N} \left(y^i \log p_i + (1-y^i)\log(1-p_i)\right)$
  $\varphi \leftarrow \mathrm{Adam}\left(\nabla_\varphi L_1, \theta_1, \beta_1, \beta_2, \epsilon = 10^{-4}\right)$
end for
Select C data point from the original vibration signal $S^i \leftarrow (a_1^i, a_2^i, \ldots, a_k^i, \ldots, a_C^i)$
for i to N
  for k to C
    Calculate the value of a_k^i according to formula 11
  end for
end for
$\hat{S}^{i,j,\ldots,M} \leftarrow \{a_k \mid a_k = \max(a_k^i, a_k^j, \ldots, a_k^M), a_k^i \in S^i\}$
$S_u \leftarrow \{\hat{S}^{1,2}, \hat{S}^{1,3}, \ldots, \hat{S}^{1,2,\ldots,M}\}$
while θ_2 has not converged do
  for t to n_{critic}
    for i to m:
      Sample real data $v \in P_r$, latent variable $z \in p(z)$, random number $\epsilon \in U[0,1]$.
      $\hat{v}_s^i \leftarrow G_{\theta_2}(S^i, z)$
      $\tilde{v}_s^i \leftarrow \epsilon v_s^i + (1-\epsilon)\hat{v}_s^i$
      $L_2^{(i)} \leftarrow D_w(\hat{v}_s^i) - D_w(v_s^i) + \lambda(\|\Delta_{\tilde{v}} D_w(\tilde{v}_s^i)\|_2 - 1)^2 + E(\|v_s^i - \hat{v}_s^i\|_2)$
    end for
    $w \leftarrow RMSprop(\Delta_w \frac{1}{m}\sum_{i=1}^{m} L_2^{(i)}, \alpha)$
  end for
  Sample a batch of latent variables $\{z^{(i)}\}_{i=1}^{m} \in p(z)$
  $\theta_2 \leftarrow RMSprop(\Delta_{\theta_2} \frac{1}{m}\sum_{i=1}^{m} -D_w(G_{\theta_2}(S^i, z)), \alpha)$
end while
Input compound fault features v_u, LIV S_u and noise z
$Y_u = \arg\min Dis(v_u, G_{\theta_2}(S_u, z))$

Testing: The vibration time-domain signal of compound fault is transformed to generate a 2D time–frequency image via WT, followed by extraction of fault features of compound faults. The generator uses the LIV of the compound faults to generate fault features of the testing sample. The model compares the distance between the real fault features and the generated fault features. It predicts that the fault category of the testing samples belongs to the closest generated fault feature of the compound faults.

4. Experiments

4.1. Experiment description

We collected the vibration signals of the fault bearing through a lab-built experimental platform to evaluate our proposed method, as shown in Fig. 2. Our platform includes an AC variable frequency drive motor, a shaft system consisting of a front bearing and its bearing seat, a radial load-bearing, and its bearing seat, the faulty bearing and its bearing seat, a gearbox with 75 large teeth and 55 small teeth, a base, and an electrical control box. To control the bearing speed, we used a three-phase motor with a flexible coupling. An acceleration sensor was placed in the faulty bearing seat to collect the vibration signals, which were acquired using a data acquisition card. Fig. 3 depicts the bearings with different faults, acceleration sensors, and data acquisition cards in the lab-built testbed.

The working load of the bearings is 0 HP, and the speed of the bearings is 1500 rpm. We collected vibration signals with the sampling frequency of 51.2 kHz and the sampling length is 10 s. Then we segmented the vibration data using a sliding window approach with a window size of 512 and a sliding stride of 180, resulting in obtaining 2000 fault samples per class, with each sample containing 512 data points.

Fig. 4 illustrates a few examples of the vibration signal sections. These examples of signals are the health conditions of the bearings.

The examples include signals of three single faults and four compound faults. The three single faults are composed of an inner ring fault (IF) rolling body fault (BF) and outer ring fault (OF) respectively. The four compound faults include the combination of single faults: inner ring (IF) & rolling body fault (BF), outer ring (OF) & rolling body fault (BF), inner ring (IF) & outer ring (OF) and inner ring (IF) & outer ring (OF) & rolling body (BF).

4.2. Performance of different fault diagnosis tasks

We conducted experiments in two groups of compound fault diagnosis tasks as shown in Table 2. In different tasks, the number of training samples and testing samples is different. For instance, the number of training samples per class (i.e., single faults per class) in Task A1 to Task A4 is 500, 1000, 1500, and 2000, respectively, thus the total number of single fault samples from Task A1 to Task A4 is 1500, 3000, 4500, and 6000, respectively. The number of testing samples per class (i.e., compound faults per class) in Task A1 to Task A4 is 2000, correspondingly the total number of testing samples from Task A1 to Task A4 is 6000. The classification accuracy is shown in Fig. 5. As the number of training samples increases, the classification accuracy of Task A and Task B is significantly improved. The accuracy of task A and task B achieves 77.03% and 65.80%, respectively, when the model is trained with 2000 examples in each fault class.

It should be noted that most learning models require numerous compound fault samples for training. However, in zero-shot learning models, the compound fault samples are unavailable for training. Thereby it is a methodology consensus that the performance of the zero-shot learning models (not just our model) is inferior to that of the conventional learning models, which can achieve an accuracy of over 86% (Chen et al., 2023; Yu et al., 2020). To our best knowledge, in the field of computer vision, the state-of-the-art zero-shot learning models achieve the highest classification accuracy of 76% on the benchmark dataset CUB (Ding et al., 2018; Gao et al., 2020; Gull & Arif, 2022; Ye et al., 2019). In contrast, the classification accuracy of 77.03% for Task A in our experiment is satisfactory.

In our experiments, the accuracy of Task A is higher than that of Task B. This is attributed to the complexity of task B. In contrast to task A, task B is a coupling of three kinds of faults, which makes it difficult to classify.

To evaluate the performance of the model, we obtain the precision, recall, and F1-Score on TaskA4 and TaskB4 respectively, as shown in Fig. 6. In Task A4, the model performs reasonably well on three kinds of compound faults (IF&OF, IF&BF, and OF&BF), exceeding 66%

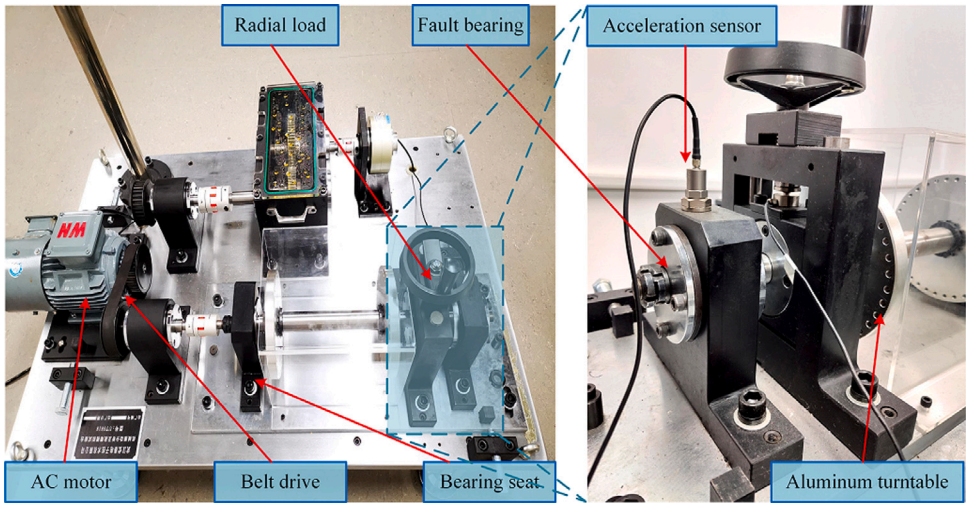


Fig. 2. Lab-built testbed.

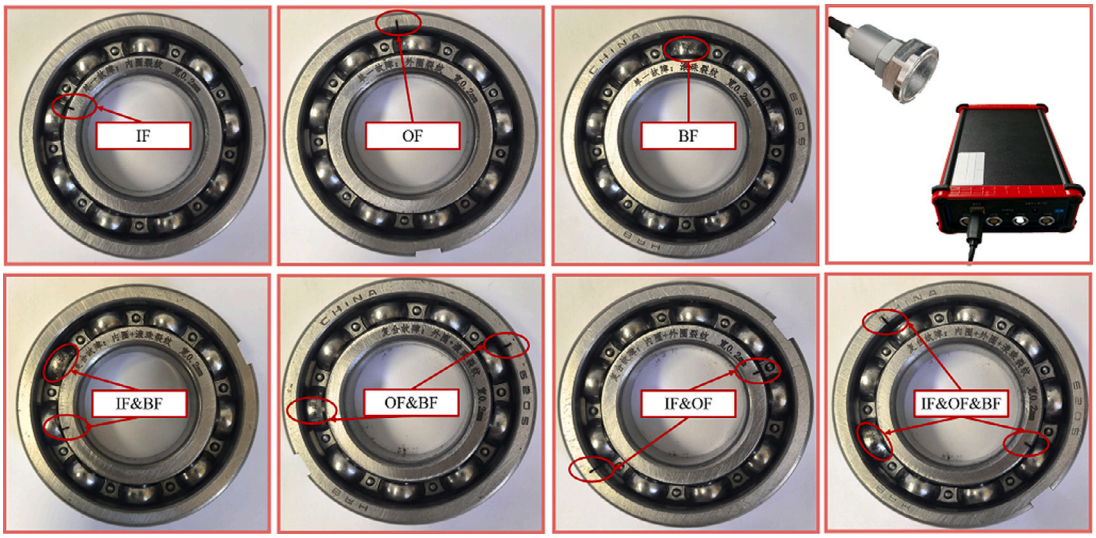


Fig. 3. Bearings with different faults, acceleration sensor, and data acquisition card in the lab-built testbed.

Table 2
Description of the diagnosis tasks.

Task	Training Faults	Testing Faults	# of Training Samples (samples × class)	# of Testing Samples (samples × class)
Task A1	BF, IF, OF	IF & BF, OF & BF, IF & OF	1500 (500 × 3)	6000 (2000 × 3)
Task A2	BF, IF, OF	IF & BF, OF & BF, IF & OF	3000 (1000 × 3)	6000 (2000 × 3)
Task A3	BF, IF, OF	IF & BF, OF & BF, IF & OF	4500 (1500 × 3)	6000 (2000 × 3)
Task A4	BF, IF, OF	IF & BF, OF & BF, IF & OF	6000 (2000 × 3)	6000 (2000 × 3)
Task B1	BF, IF, OF	IF & BF, OF & BF, IF & OF, IF & OF & BF	1500 (500 × 3)	8000 (2000 × 4)
Task B2	BF, IF, OF	IF & BF, OF & BF, IF & OF, IF & OF & BF	3000 (1000 × 3)	8000 (2000 × 4)
Task B3	BF, IF, OF	IF & BF, OF & BF, IF & OF, IF & OF & BF	4500 (1500 × 3)	8000 (2000 × 4)
Task B4	BF, IF, OF	IF & BF, OF & BF, IF & OF, IF & OF & BF	6000 (2000 × 3)	8000 (2000 × 4)

for all three metrics. Among them, the precision, recall, and F1-Score on IF&OF are 83%, 95%, and 89% respectively, which has the most stable performance in Task A4. In Task B4, the testing set contains the compound fault of IF&OF&BF. It is coupled with three kinds of single faults, and its fault features are more difficult to identify. Thus the performance of the model drops on Task B4, and all three metrics on four kinds of compound faults (IF&OF, IF&BF, OF&BF, and IF&OF&BF) are above 51%. Among them, the precision, recall, and F1-Score on OF&BF are 81%, 78%, and 75% respectively, which has the most stable performance in Task B4.

4.3. Ablation study

To provide further insight into our model, we conduct ablation studies to evaluate the effect of different feature extractors and generative modules.

(1) Analysis of feature extractor: We evaluate the impact of using Principal Component Analysis (PCA) and Autoencoders (AE) as feature extractors. The proposed model is unchanged except for the feature extraction module. Note that we only use single fault samples to train FE in the absence of compound fault samples, and use the pre-trained

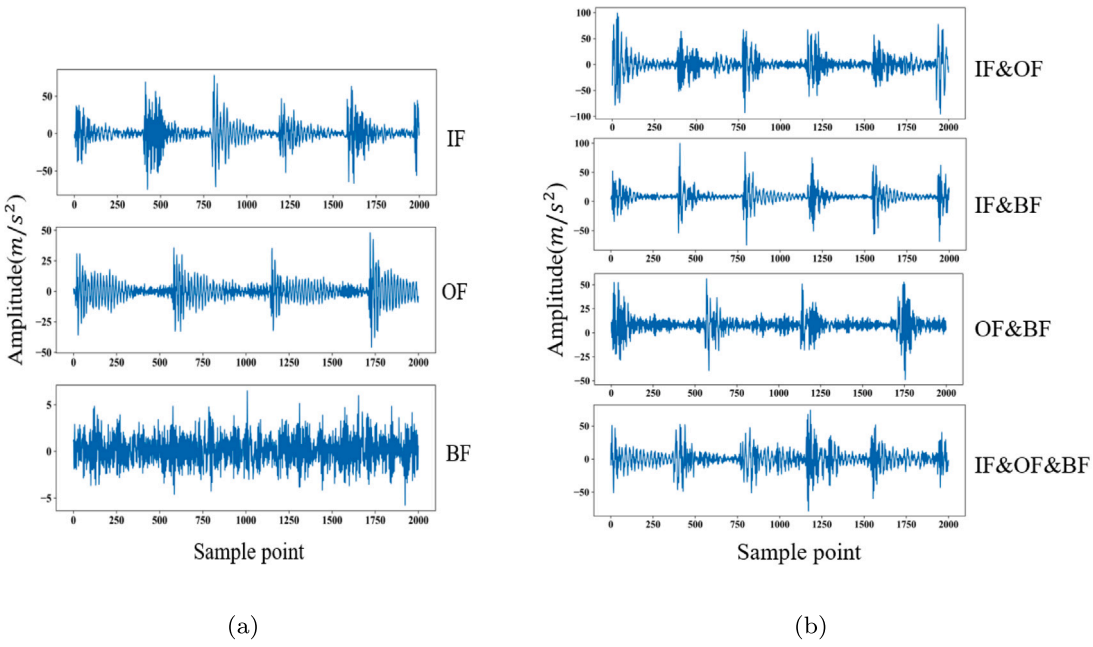


Fig. 4. Vibration signal of seven health conditions: (a) single fault signals; (b) compound fault signals.

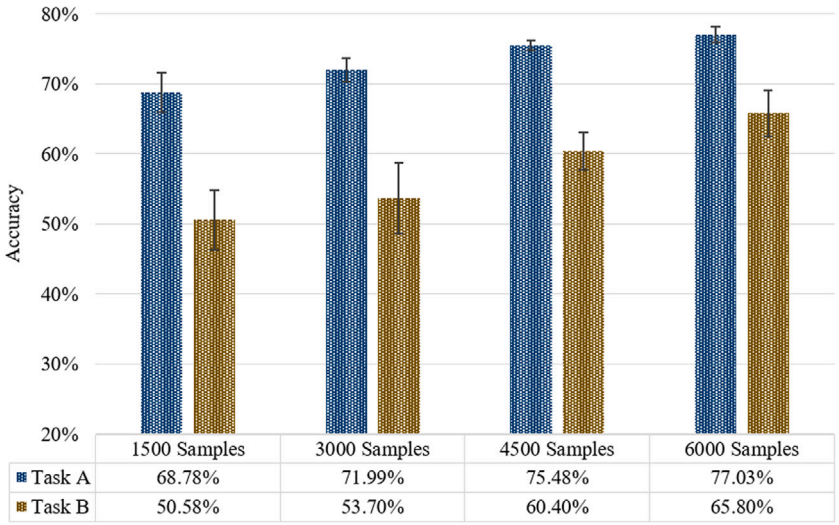


Fig. 5. The fault classification accuracy on different diagnosis tasks.

FE to extract the compound fault features. Visualization of the t-SNE for fault features extracted by PCA, AE, and ours is presented in Fig. 7, where four different colors represent different fault classes. It can be seen that the features of the four classes extracted by PCA and AE are mostly overlapped, while the four features extracted by our FE are quite separated. This implies that our FE has a better distinguishing ability at the class level, which is conducive to the final classification.

Fig. 8 illustrates the compound fault classification accuracy using different feature extractors on tasks A4 and B4. As the number of iterations increases, for task A4, when the number of iterations exceeds 2000, the classification accuracy using our feature extractor is much higher (77.03%) than PCA (38.63%) and AE (71.01%). For task B4, when the number of iterations exceeds 2000, the classification accuracy using our feature extractor is much higher (65.80%) than PCA (30.68%) and AE (52.76%).

(2) Analysis of the generative module: We evaluate the influence of different compared generative models on the accuracy of fault classification. Specifically, the proposed model is unchanged except for the generative adversarial module. GAN, VAE, VAE_GAN, and our model are used to generate compound fault features for classification. Note that the GAN in our model is different from the comparison GAN herein. The comparison GAN and our model have the same structure of generator, while our model improved the loss functions, and the discriminator structure is also modified, removing the sigmoid in the last layer.

The 3D visualization of the t-SNE of compound fault features generated by different generative models on task A4 is shown in Fig. 9, where the three different colors represent different fault classes. Hence, the features produced by our generator are more discriminative, aggregative, and structure consistent than the three other generative models.

Further to intuitively evaluate the compound fault features generated by different generative models, we generated 100 compound fault features using different generative models and calculated the center point of each class of compound fault on task A4. We used Euclidean distance to metric the distance between center points of different fault

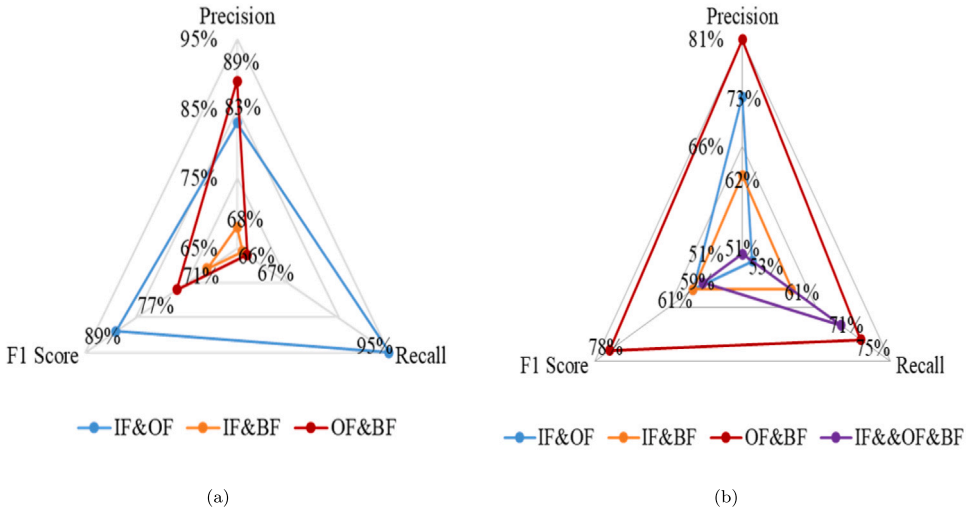


Fig. 6. Precision, recall and F1-Score results of different tasks: (a) Task A4; (b) Task B4.

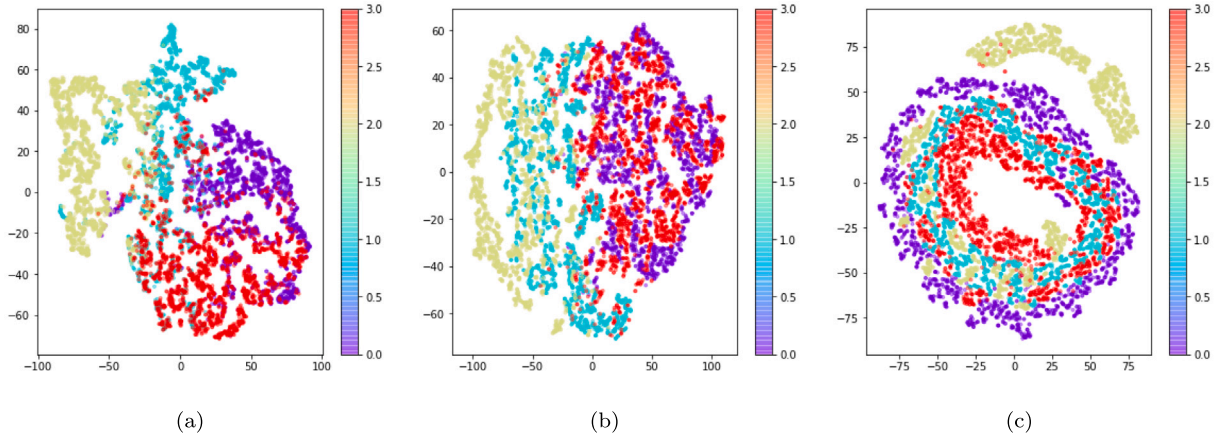


Fig. 7. Visualization of the compound fault features on task B4 using different feature extractors: (a) Ours; (b) AE; (c) PCA.

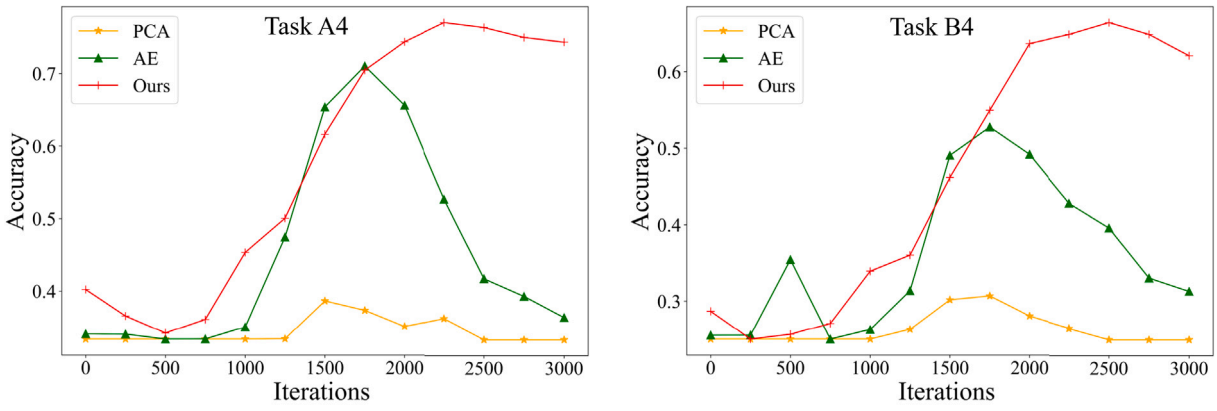


Fig. 8. Compound fault classification accuracy using different feature extractors with the number of iterations.

classes, as shown in Fig. 10. The larger the distance between different classes of fault features, the smaller the overlap of different fault features.

Compared with the three other models, our model has the maximum distance between IF&OF and IF&BF (12.79), IF&OF, and OF&BF (22.25). With respect to the distance between IF&BF and 0F&BF, our model with a distance of 16.52 is close to the best model VAE_GAN with a distance of 20.18. Overall the compound fault features generated by our method have a lower overlap than other methods.

Further, we also use Euclidean distance to metric the distance per class between center points of generated compound fault features and the real compound fault features extracted from the compound fault

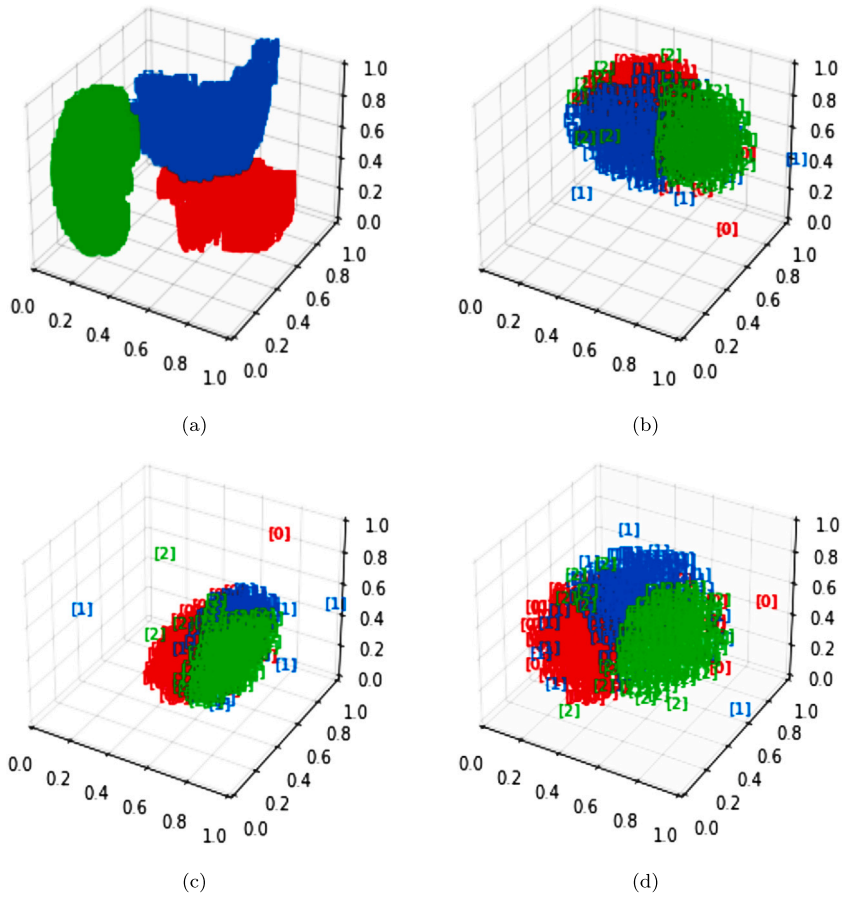
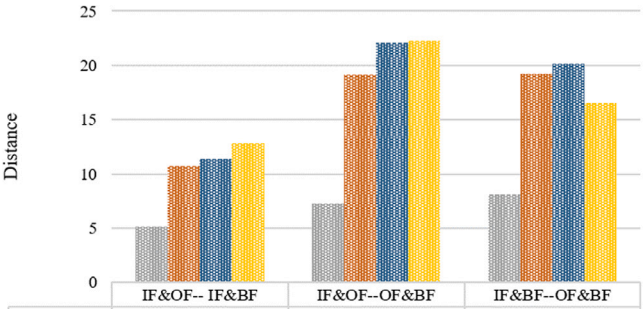
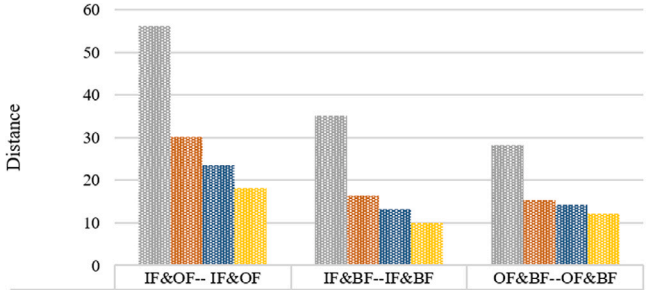


Fig. 9. Visualization of compound fault features generated by different generative models on task A4: (a) Our method; (b) VAE; (c) GAN; (d) VAE-GAN.



	IF&OF-- IF&BF	IF&OF--OF&BF	IF&BF--OF&BF
GAN	5.18	7.28	8.11
CVAE	10.71	19.16	19.26
VAE_GAN	11.41	22.05	**20.18**
Ours	**12.79**	**22.25**	16.52

Fig. 10. The distance between each class of the compound fault features generated by different generative models on task A4.



	IF&OF-- IF&OF	IF&BF--IF&BF	OF&BF--OF&BF
GAN	56.14	35.22	28.27
CVAE	30.21	16.41	15.44
VAE_GAN	23.41	13.2	14.27
Ours	**18.11**	**9.88**	**12.05**

Fig. 11. The distance per class between center points of generated compound fault features and the real compound fault features.

samples, as shown in Fig. 11. It can be seen that our method has the minimum distance between the two. This demonstrates the superior performance of our method.

Fig. 12 illustrates the compound fault classification accuracy using different generative models on tasks A4 and B4. As the number of iterations increases, especially when the number of iterations exceeds 1500, our generative module has superior classification accuracy than other generative models, with respect to task A4 or task B4.

4.4. Performance of different distance metrics

To understand the impact of distance metrics on our classification, we evaluate four different metrics in addition to Euclidean distance. Fig. 13 depicts the fault classification accuracy of using these distance metrics. The average fault classification accuracy with different distance metric methods exceeds 73% and 62% on tasks A4 and B4, respectively. This affirmed that the model had excellent generalization ability with respect to different distance metrics.

Among them, the classification accuracy using the Euclidean distance is greater than the other four distance metric methods. After

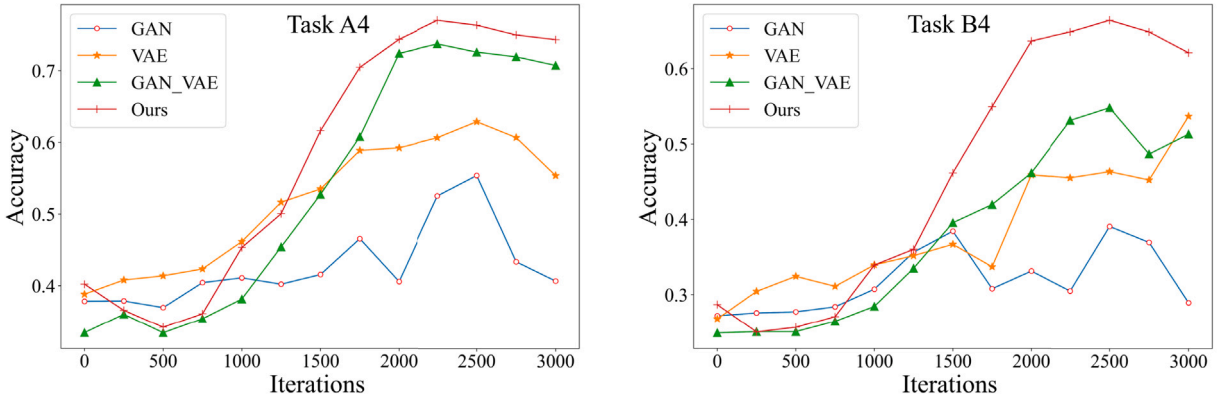
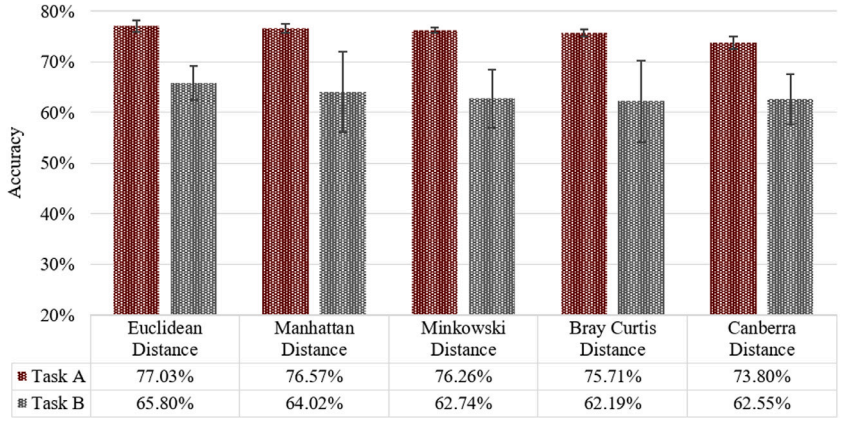
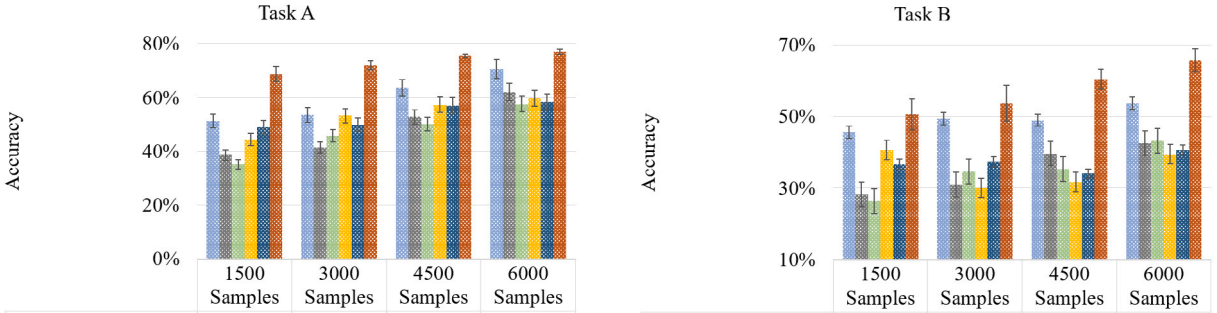


Fig. 12. The compound fault classification accuracy using different generative module models with the number of iterations.



	Euclidean Distance	Manhattan Distance	Minkowski Distance	Bray Curtis Distance	Canberra Distance
Task A	77.03%	76.57%	76.26%	75.71%	73.80%
Task B	65.80%	64.02%	62.74%	62.19%	62.55%

Fig. 13. Classification accuracy using different distance metrics.



	1500 Samples	3000 Samples	4500 Samples	6000 Samples
Feature_Attribute	51.33%	53.63%	63.57%	70.62%
Word_Embedding	38.58%	41.35%	52.87%	62.07%
CNN_Attributc	35.17%	45.88%	50.23%	57.57%
Threshold_Attribute	44.38%	53.30%	57.35%	59.75%
LDS	48.97%	49.90%	57.13%	58.38%
Ours	**68.78%**	**71.99%**	**75.48%**	**77.07%**

	1500 Samples	3000 Samples	4500 Samples	6000 Samples
Feature_Attribute	45.59%	49.43%	48.96%	53.72%
Word_Embedding	28.28%	31.01%	39.65%	42.50%
CNN_Attributc	26.38%	34.60%	35.34%	43.18%
Threshold_Attribute	40.64%	30.01%	31.68%	39.48%
LDS	36.73%	37.36%	34.03%	40.60%
Ours	**50.58%**	**53.70%**	**60.40%**	**65.80%**

Fig. 14. Classification accuracy using different LIV definition methods.

analysis, the Manhattan distance is more likely to generate a higher distance value than the Euclidean distance because it is not the shortest possible path, while Minkowski distances do not take into account the distribution of the individual components (expectation, variance, etc.). On the other hand, the Bray-Curtis distance, which is an asymmetric index, is more suitable for analyzing community species data. Furthermore, Canberra distance assumes that the variables are independent of each other, and does not take into account the correlation between the variables. Therefore, the fault classification accuracy of these four distance metric methods is inferior to Euclidean distances.

4.5. Performance of using different label information vector

We conducted experiments using different label information in our method, including Feature_Attribute (Chen et al., 2019), Word_Embedding (Mikolov et al., 2013), CNN_Attribute (Chen, 2018), Threshold_Attribute (Xu et al., 2022b), and LSD (Xing et al., 2022). Descriptions for different LIVs are outlined in Table 3. We only change the construction of LIV of our method and report the accuracy of tasks A and B. The experimental results are shown in Fig. 14.

Table 3
LIV methods.

Method	Description
Word Embedding (Mikolov et al., 2013)	Apply the word vector model to train a fault corpus and convert the labeled words of the fault into fault LIV. We use the Skip-gram model to convert the labeled words of each fault class into a vector, which is used as the LIV for each class.
Feature Attribute (Chen et al., 2019)	Twenty-four attributes in the time and frequency domains are extracted, which are normalized to the range of 0 and 1. Each attribute is binarized using a threshold of 0.5 and used as the LIV.
CNN Attribute (Chen, 2018)	Design 192-dimensional vectors as the LIV, which is divided into three intervals representing different attributes, and the intervals of different attributes are corresponding to different single fault features. These 64-dimensional single fault features are extracted from various single fault samples by CNN. When there is no response to a single fault, the corresponding attribute interval is set to 0. LIV operation steps are described in Chen (2018).
Threshold Attribute (Xu et al., 2022b)	Firstly, R vibration signal data points of every single fault are selected, followed by the setting of a threshold value. If the data point is greater than the threshold value, then the value of dimension is 1, otherwise 0; Thus we obtain the single fault LIV. Finally, compound fault LIV is constructed by logic OR operation on the single faults LIV.
LDS (Xing et al., 2022)	Based on the theoretical relationship of features and the LIVs relationship among the compound and single faults. By using the single fault feature to train N softmax regression models to obtain weight matrices. All the weight matrices are set to construct the LIV of different faults. More details about LDS can be found in Xing et al. (2022).

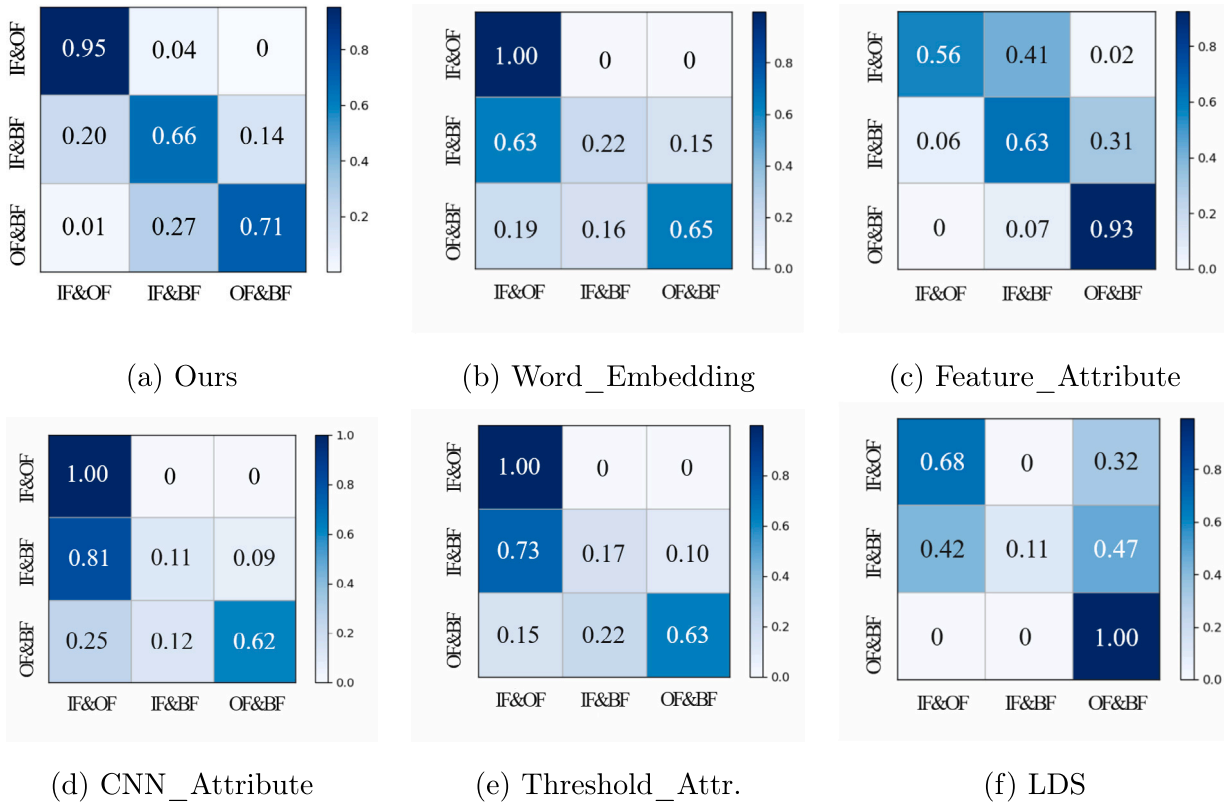


Fig. 15. Confusion matrix results using different LIV definition methods on task A4.

For task A, our method achieves the highest classification accuracy of 77.03%. The classification accuracy of the other five methods is 70.62%, 62.07%, 57.57%, 59.75%, and 58.38%, respectively. Notably, compared with the other five LIV definition methods, the accuracy of our method is significantly improved by 9% with respect to the second-best.

For task B, except for the feature attribute LIV method on task B4 (53.72%) the classification accuracy of the other four LIV methods is less than 50%, while the accuracy of our method is significantly improved. After analysis, we found that the BF features are not obvious and are easily concealed by other fault features, leading to misclassification. In contrast, our LIV method can effectively eliminate the aforementioned problem with a classification accuracy is 65.80% on task B4.

Figs. 15 and 16 depict the confusion matrix for fault classification results of task A4 and task B4 using different LIV definition methods. The abscissa and ordinate of the confusion matrix represent the predicted and actual fault labels, respectively. Each unit represents prediction accuracy. Our LIV method performed superior in each fault class, whereas other LIV methods result in indistinguishable faults. For instance, in Fig. 15, the probabilities of misclassification of IF&BF to IF&OF and OF&BF are 0.63 and 0.15 using Word_embedding, respectively.

Fig. 16 illustrates the testing set of compound faults coupled with three single faults: inner ring, outer ring, and rolling body fault. The weak features of rolling body faults are covered by other fault features, which is difficult for the models to distinguish between IF&OF and IF&OF&BF. For instance, the classification accuracy of IF&OF&BF faults using Word_embedding, CNN_Attribute, Threshold_Attribute, and LDS is 0, while our method could further mitigate this problem with the classification accuracy of 65.80%.

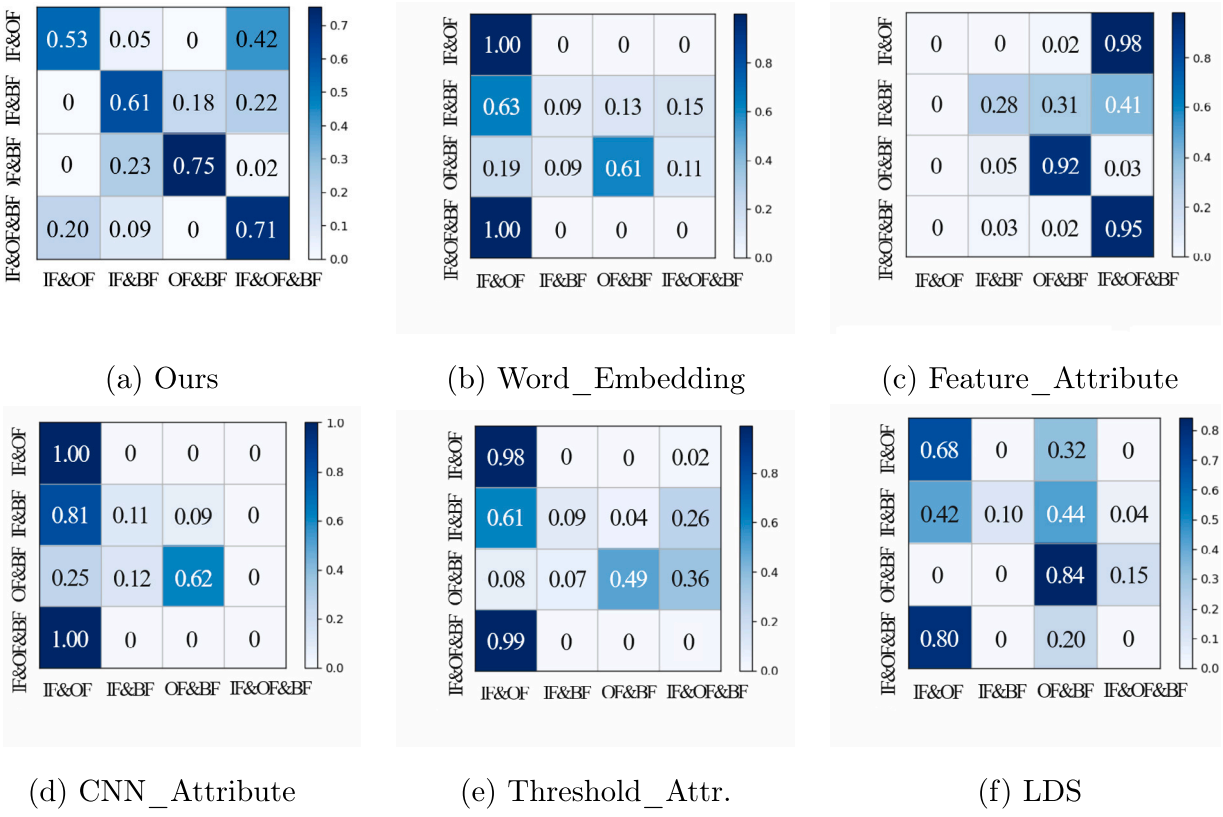


Fig. 16. Confusion matrix results using different LIV definition methods on task B4.

Table 4
Comparison methods.

Method	Description
CADA_VAE (Schonfeld et al., 2019)	The CADA-VAE model learns a latent embedding of fault features and class embedding via aligned VAEs optimized with cross-alignment and distribution alignment objectives and subsequently trains a classifier on latent features of seen and unseen classes.
VZSL (Wang et al., 2018)	The VZSL model uses the potential spatial distribution of the single faults as the priority for VAE. We learn a set of attribute-specific latent space distributions and infer attributes of the compound faults to achieve the classification.
DeViSE (Frome et al., 2013)	The DeVise model embeds the fault feature space and the label information space into the deep fault label information embedding space and uses the class prototype learned in the embedding space to classify using the nearest neighbor method.
CDL (Jiang et al., 2018)	The CDL model uses class prototypes to align fault semantic structures, where the discriminative information lying in the fault feature space is utilized to improve the less discriminative label information space. ZSL recognition can be performed in different spaces by the simple nearest neighbor approach using the learned class prototypes.
LDS_IFD (Xing et al., 2022)	Label description space (LDS) is established to construct the relationship between different fault modes. LDS is embedded between the feature space (FS) and the health condition label space (HCLS). We use a linear supervised autoencoder (LSAE) to construct the projection between FS and LDS. LDS-IFD identifies compound faults by similarity evaluation in LDS or FS.

4.6. Performance of different zero-shot learning models

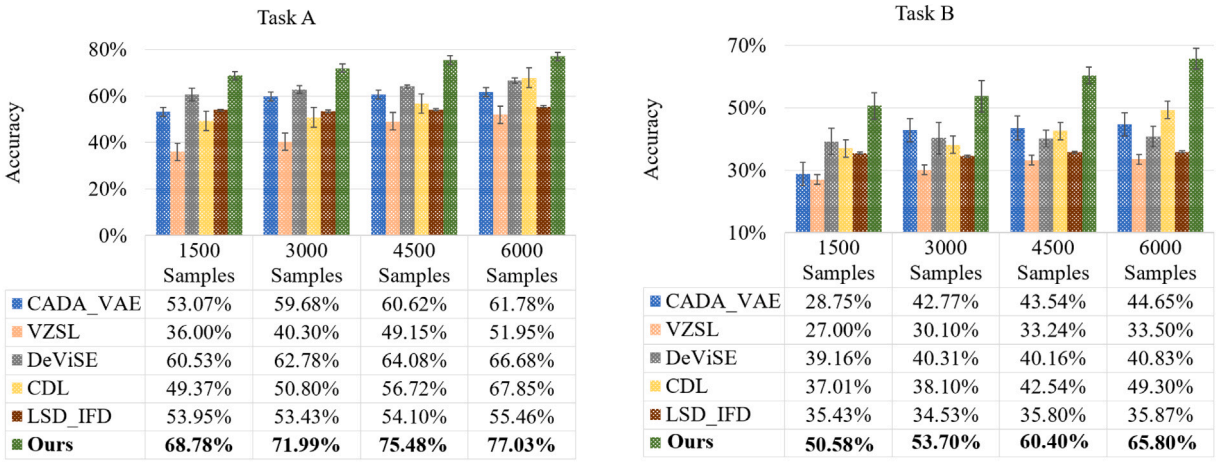
We evaluate our model against the state-of-the-art ZSL models including CADA_VAE (Schonfeld et al., 2019), VZSL (Wang et al., 2018), DeViSE (Frome et al., 2013), CDL (Jiang et al., 2018), and LSD_IFD (Xing et al., 2022). Details of the models are presented in Table 4. We implement these models as our baseline. For a fair comparison, all the models are trained using the labeled single fault samples and LIVs we proposed, then these models are used to identify the unlabeled compound fault samples. The fault classification results of different models are presented in Fig. 17. Notably, our model is superior to the other five methods in all cases.

In task A, for the methods, CADA_VAE, VZSL, DeViSE, CDL, and LSD_IFD, the highest average accuracy is from 55.46% to 67.85%, while our method can reach 77.03%. The improvement with respect to the second-best (i.e., CDL) is 9.18%. This demonstrates that our proposed method effectively distinguishes the compound faults in task A.

In task B, the improvement of our model is more obvious, with respect to the second-best (i.e., CDL) is 16.5%. Specifically, the average accuracy of VZSL and LSD_IFD is below 40%, thereby these two methods are not applicable herein. In addition, the highest average accuracy of the CADA_VAE, DeViSE, and CDL methods are dissatisfactory, from 40.83% to 49.30%. The main reason is that the weak features of BF are covered by other fault features, which makes three comparison methods fail to distinguish between IF&OF and IF&OF&BF. In contrast, the highest average accuracy of our proposed method can reach 65.80%, which effectively alleviates the recognition problem of the weak feature of BF.

5. Conclusion

To tackle the problem of compound fault diagnosis with no or extremely scarce examples, in this paper, we propose a label information vector generative zero-shot model for the diagnosis of compound faults.



	Task A 1500 Samples	Task A 3000 Samples	Task A 4500 Samples	Task A 6000 Samples
CADA_VAE	53.07%	59.68%	60.62%	61.78%
VZSL	36.00%	40.30%	49.15%	51.95%
DeViSE	60.53%	62.78%	64.08%	66.68%
CDL	49.37%	50.80%	56.72%	67.85%
LSD_IFD	53.95%	53.43%	54.10%	55.46%
Ours	**68.78%**	**71.99%**	**75.48%**	**77.03%**

	Task B 1500 Samples	Task B 3000 Samples	Task B 4500 Samples	Task B 6000 Samples
CADA_VAE	28.75%	42.77%	43.54%	44.65%
VZSL	27.00%	30.10%	33.24%	33.50%
DeViSE	39.16%	40.31%	40.16%	40.83%
CDL	37.01%	38.10%	42.54%	49.30%
LSD_IFD	35.43%	34.53%	35.80%	35.87%
Ours	**50.58%**	**53.70%**	**60.40%**	**65.80%**

Fig. 17. Comparison of the overall accuracy of our model with the state-of-the-art ZSL models in different tasks.

Our method is trained with only single fault samples and is capable of identifying compound faults. We devise a unified LIV definition method for expressing single and compound fault class information. The model learns the mapping relationship between the fault features and the fault LIVs through adversarial training. Then the trained generator generates compound fault features using the compound fault LIVs. The model identifies the class of compound faults by leveraging the similarity measurement between the real compound fault features and the generated compound fault features. We conduct several experiments using the vibration datasets acquired from our lab-built platform. Extensive comparison results demonstrate that the proposed method achieves a superior classification accuracy over the state-of-the-art methods despite the lack of compound fault examples.

This paper focuses on the traditional zero-shot learning classification, that is, only multi-classification on compound fault samples. In the future, we will explore the generalized zero-shot learning method, that is, multi-classification in single fault and compound fault samples. This will help to accurately identify and classify various faults, promoting the application of an intelligent fault diagnosis system. Meanwhile, in this study, the severity of different faults is constant due to the limitations of vibration sample collection. In future work, it is interesting to identify compound fault severities in the testing stage while different severities are applied in the training stages.

CRediT authorship contribution statement

Juan Xu: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Kang Li:** Software, Writing – original draft. **Yuqi Fan:** Resources, Formal analysis. **Xiaohui Yuan:** Conceptualization, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request

References

Chatti, N., Ould-Bouamama, B., Gehin, A.-L., & Merzouki, R. (2014). Signed bond graph for multiple faults diagnosis. *Engineering Applications of Artificial Intelligence*, *36*, 134–147.

Chen, W. (2018). Application of deep learning in rolling bearing fault diagnosis. *Southwest Jiaotong University*, (621), 1–63.

Chen, C., Liu, C., Wang, T., Zhang, A., Wu, W., & Cheng, L. (2023). Compound fault diagnosis for industrial robots based on dual-transformer networks. *Journal of Manufacturing Systems*, *66*, 163–178.

Chen, J., Pan, T., Zhou, Z., & He, S. (2019). An adversarial learning framework for zero-shot fault recognition of mechanical systems. *Vol. 1*, In *2019 IEEE 17th international conference on industrial informatics* (pp. 1275–1278). IEEE.

Demirel, B., Cinbis, R. G., & Ikizler-Cinbis, N. (2019). Learning visually consistent label embeddings for zero-shot learning. In *2019 IEEE international conference on image processing* (pp. 3656–3660). IEEE.

Deng, W., Li, Z., Li, X., Chen, H., & Zhao, H. (2022). Compound fault diagnosis using optimized MCKD and sparse representation for rolling bearings. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1–9.

Dibaj, A., Ettefagh, M. M., Hassannejad, R., & Ehghaghi, M. B. (2021). A hybrid fine-tuned VMD and CNN scheme for untrained compound fault diagnosis of rotating machinery with unequal-severity faults. *Expert Systems with Applications*, *167*, Article 114094.

Ding, Z., Shao, M., & Fu, Y. (2018). Generative zero-shot learning via low-rank embedded semantic dictionary. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(12), 2861–2874.

Feng, L., & Zhao, C. (2020). Fault description based attribute transfer for zero-sample industrial fault diagnosis. *IEEE Transactions on Industrial Informatics*, *17*(3), 1852–1862.

Frome, A., Corrado, G. S., Shlens, J., Bengio, S., Dean, J., Ranzato, M., & Mikolov, T. (2013). Devise: A deep visual-semantic embedding model. *Advances in Neural Information Processing Systems*, *26*.

Gao, R., Hou, X., Qin, J., Chen, J., Liu, L., Zhu, F., Zhang, Z., & Shao, L. (2020). Zero-vae-gan: Generating unseen features for generalized and transductive zero-shot learning. *IEEE Transactions on Image Processing*, *29*, 3665–3680.

Gao, S., Shi, S., & Zhang, Y. (2022). Rolling bearing compound fault diagnosis based on parameter optimization MCKD and convolutional neural network. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1–8.

Gull, M., & Arif, O. (2022). Generalized zero-shot learning using identifiable variational autoencoders. *Expert Systems with Applications*, *191*, Article 116268.

Guo, Q., Li, Y., Song, Y., Wang, D., & Chen, W. (2019). Intelligent fault diagnosis method based on full 1-D convolutional generative adversarial network. *IEEE Transactions on Industrial Informatics*, *16*(3), 2044–2053.

Huang, R., Li, J., Li, W., & Cui, L. (2019). Deep ensemble capsule network for intelligent compound fault diagnosis using multisensory data. *IEEE Transactions on Instrumentation and Measurement*, *69*(5), 2304–2314.

Jiang, H., Wang, R., Shan, S., & Chen, X. (2018). Learning class prototypes via structure alignment for zero-shot recognition. In *Proceedings of the European conference on computer vision* (pp. 118–134).

Lampert, C. H., Nickisch, H., & Harmeling, S. (2013). Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(3), 453–465.

Li, N., Huang, W., Guo, W., Gao, G., & Zhu, Z. (2019). Multiple enhanced sparse decomposition for gearbox compound fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, *69*(3), 770–781.

Li, W., Zhang, S., & Rakheja, S. (2015). Feature denoising and nearest–farthest distance preserving projection for machine fault diagnosis. *IEEE Transactions on Industrial Informatics*, *12*(1), 393–404.

Liang, P., Deng, C., Wu, J., Yang, Z., Zhu, J., & Zhang, Z. (2019). Compound fault diagnosis of gearboxes via multi-label convolutional neural network and wavelet transform. *Computers in Industry*, *113*, Article 103132.

Mhamdi, L., Dhouibi, H., Liouane, N., & Simeu-Abazi, Z. (2013). Multiple fault diagnosis using mathematical models. In *2013 9th asian control conference* (pp. 1–6). IEEE.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *26*.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing* (pp. 1532–1543).

Piacentino, A., & Talamo, M. (2013). Innovative thermoeconomic diagnosis of multiple faults in air conditioning units: Methodological improvements and increased reliability of results. *International Journal of Refrigeration*, *36*(8), 2343–2365.

Rahman, S., Khan, S. H., & Porikli, F. (2020). Zero-shot object detection: joint recognition and localization of novel concepts. *International Journal of Computer Vision*, *128*(12), 2979–2999.

Schonfeld, E., Ebrahimi, S., Sinha, S., Darrell, T., & Akata, Z. (2019). Generalized zero- and few-shot learning via aligned variational autoencoders. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8247–8255).

Shigeto, Y., Suzuki, I., Hara, K., Shimbo, M., & Matsumoto, Y. (2015). Ridge regression, hubness, and zero-shot learning. In *Joint European conference on machine learning and knowledge discovery in databases* (pp. 135–151). Springer.

Tang, G., Wang, Y., Huang, Y., & Wang, H. (2020). Multiple time-frequency curve classification for tacho-less and resampling-less compound bearing fault detection under time-varying speed conditions. *IEEE Sensors Journal*, *21*(4), 5091–5101.

Ubar, R., Raik, J., Kostin, S., & Kousaar, J. (2012). Multiple fault diagnosis with BDD based boolean differential equations. In *2012 13th biennial baltic electronics conference* (pp. 77–80). IEEE.

Wang, W., Pu, Y., Verma, V., Fan, K., Zhang, Y., Chen, C., Rai, P., & Carin, L. (2018). Zero-shot learning via class-conditioned deep generative models. *Vol. 32*, In *Proceedings of the AAAI conference on artificial intelligence*. (1).

Xing, S., Lei, Y., Wang, S., Lu, N., & Li, N. (2022). A label description space embedded model for zero-shot intelligent diagnosis of mechanical compound faults. *Mechanical Systems and Signal Processing*, *162*, Article 108036.

Xu, J., & Li, K. (2021). Generative zero-shot learning compound fault diagnosis of bearings. In *2021 international conference on sensing, measurement data analytics in the era of artificial intelligence* (pp. 1–7).

Xu, B., Zeng, Z., Lian, C., & Ding, Z. (2022). Generative mixup networks for zero-shot learning. *IEEE Transactions on Neural Networks and Learning Systems*.

Xu, J., Zhou, L., Zhao, W., Fan, Y., Ding, X., & Yuan, X. (2022). Zero-shot learning for compound fault diagnosis of bearings. *Expert Systems with Applications*, *190*, Article 116197.

Ye, Z., Lyu, F., Li, L., Fu, Q., Ren, J., & Hu, F. (2019). SR-GAN: Semantic rectifying generative adversarial network for zero-shot learning. In *2019 IEEE international conference on multimedia and expo* (pp. 85–90). IEEE.

Yu, K., Ma, H., Lin, T. R., & Li, X. (2020). A consistency regularization based semi-supervised learning approach for intelligent fault diagnosis of rolling bearing. *Measurement*, *165*, Article 107987.

Zhao, M., Zhong, S., Fu, X., Tang, B., & Pecht, M. (2019). Deep residual shrinkage networks for fault diagnosis. *IEEE Transactions on Industrial Informatics*, *16*(7), 4681–4690.

Zhuo, Y., & Ge, Z. (2021). Auxiliary information-guided industrial data augmentation for any-shot fault learning and diagnosis. *IEEE Transactions on Industrial Informatics*, *17*(11), 7535–7545.