Counterfactual Inference for Generalized Zero-Shot Compound-Fault Diagnosis

Juan Xu, *Member, IEEE*, Hui Kong, Xu Ding, *Member, IEEE*, and Xiaohui Yuan, *Senior Member, IEEE*

***Abstract*— Learning a model heavily depends on the training examples, which are sometimes difficult to obtain if not impossible. This a typically true for fault diagnosis in machinery, particularly for compound faults. The counterfactual inference reveals the causal components inherent in the fault data in an interpretable manner, divulging critical causes from the observable phenomena. This article proposes a method to address the imbalance and interpretability issues of generalized zero-shot learning (GZSL) methods for compound-fault diagnosis using counterfactual inference. Our method uses a structural causal model (SCM) to decouple and generate fault features, which enhances the capabilities of the variational autoencoder and generative adversarial network (VAE-GAN) through a strengthened discriminator, and reveals the intrinsic causal components in fault data, distinguishing key fault causes from accompanying phenomena. This enables the classification of both single and compound faults by learning from examples of single faults, easing the dependence on the examples of compound faults. Extensive experimental results show that our method, trained solely with single-fault samples, achieves a harmonic average of 87.40% accuracy for both single and compound faults, outperforming existing state-of-the-art methods. This significantly improves both the accuracy and interpretability of compound-fault diagnosis.**

***Index Terms*— Counterfactual inference, fault diagnosis, generalized zero-shot learning (GZSL), generative adversarial network, rolling bearing.**

I. Introduction

BEARINGS are critical components in complex industrial systems. Faults reduce the equipment's lifespan, disrupt production, and cause safety accidents. Timely diagnosis ensures the normal operation of mechanical equipment [1], [2]. A compound fault of bearings is a composite of the simultaneous occurrence of multiple single faults. The characteristics of compound faults are highly coupled and difficult to identify, which presents a significant challenge in fault diagnosis.

In learning-based fault diagnosis methods, the exponential growth of compound-fault patterns makes it almost infeasible to collect examples of all kinds of compound faults for model training. However, in real industrial scenarios, systems with bearings experience both single and compound faults. This necessitates a generalized fault diagnostic model derived from single-fault examples, enabling the identification of both single and compound faults [3], [4].

Zero-shot learning (ZSL) has emerged as a promising method for addressing the fault detection problem, which tackles the issue of learning from scarce examples. The idea is to leverage prior knowledge of unseen classes (the semantic attributes of the classes) and the examples of available classes to train models to classify the unseen classes [5]. Specifically, ZSL-based fault diagnosis is trained using single-fault samples (seen classes) and supplemented with prior semantics of compound faults (unseen classes). Hence, the model can be generalized to compound-fault diagnosis tasks.

The studies of ZSL in the field of compound-fault diagnosis are unfolding, including attribute-based, embedding-based, and generative-based ZSL compound-fault diagnosis methods. The attribute-based approaches [6] facilitate the learning of the mapping between single-fault samples (seen-class) and attributes by the attribute classifier. Subsequently, the attribute classifier is utilized to classify the compound-fault samples (unseen class) based on the Bayesian attribute prediction. Embedding-based approaches [7], [8] combine samples and semantic information by embedding them in specific spaces to transfer from the single faults to compound faults. Generative-based approaches [9], [10], [11] convert ZSL into traditional supervised learning by generating the compound-fault samples from compound-fault semantics. Furthermore, confronted with a generalized zero-shot compound-fault diagnosis task, in our previous studies, we constructed the prior fault semantics using a semantic-feature embedding module, and input it into a contrast-embedding generative adversarial network to generate pseudo-compound-fault samples, which assist in training the adaptive smoothing module to classify fault samples [12].

However, compound-fault diagnosis based on generalized ZSL (GZSL) faces more complex test scenarios, requiring the model to recognize single (seen classes) and compound faults (unseen classes) under the condition of training with only single faults (seen classes), which faces the following challenges: First, the end-to-end nonlinear mapping characteristics of the

Received 20 December 2024; revised 27 March 2025; accepted 16 April 2025. Date of publication 28 April 2025; date of current version 16 May 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 52375089, in part by the Dreams Foundation of Jianghuai Advance Technology Center under Grant 2023-ZM01J003, and in part by the Open Foundation of State Key Laboratory of High-End Compressor and System Technology under Grant SKL-YSJ202307. The Associate Editor coordinating the review process was Dr. Arunava Naha. *(Corresponding author: Xiaohui Yuan.)*

Juan Xu, Hui Kong, and Xu Ding are with the Key Laboratory of Knowledge Engineering With Big Data, Ministry of Education, School of Computer and Information, Hefei University of Technology, Hefei 230601, China (e-mail: xujuan@hfut.edu.cn; 2022111040@mail.hfut.edu.cn; dingxu@hfut.edu.cn).

Xiaohui Yuan is with the Department of Computer Science and Engineering, University of North Texas, Denton, TX 76205 USA (e-mail: xiaohui.yuan@unt.edu).

Digital Object Identifier 10.1109/TIM.2025.3565070

existing model learning-based compound-fault diagnosis methods lead to a lack of interpretability in the decision-making process [13], and users are prevented from understanding the model's inference logic and decision-making basis, making it difficult to be applied in high-risk scenarios; second, for the existing GZSL methods, the models are trained only with single-fault samples (seen classes), which leads to the tendency to predict the compound faults (unseen classes) as single faults (seen classes); finally, in GZSL, the compound-fault samples are generated only from semantics and noise, which may differ from the ground-truth distribution of compound-fault samples, leading to classification failure.

To tackle these challenges, we point out the existence of causal and non-causal features in fault data and construct a structural causal model (SCM) for the compound-fault diagnosis model through theoretical analysis, which provides theoretical guidance for the subsequent construction of an interpretable diagnosis model. On this basis, we introduce counterfactual inference into compound-fault diagnosis based on ZSL and propose a generalized zero-shot compound-fault diagnosis method based on counterfactual inference. Our method consists of a feature extraction module, a semantic alignment module, and a generative adversarial module. The samples and semantics of single faults are used to train the model's decoupling and generation ability of causal/non-causal features under the guidance of the SCM. In counterfactual inference, the designed generative adversarial module generates counterfactual fault features (i.e., generated pseudo-single and compound-fault features), devised from the faults' non-causal features and the different fault semantics. In such a way, the classifiers can be trained to obtain high-quality binary classification boundaries (single or compound faults), which solves the problem that the generated pseudo-compound-fault features in the existing ZSL are biased toward the single-fault distribution in an interpretable causal inference manner. In addition, our proposed model strengthens the effect of the discriminator in the generative adversarial module through a comparator, thus realizing the generation of high-quality compound-fault samples only from the compound-fault semantics and mitigating the distributional differences between the generated and ground-truth samples. Finally, a generalized zero-shot compound-fault diagnosis task is achieved using supervised learning and traditional ZSL methods for separated single and compound faults, respectively.

The main contributions of this article are summarized as follows.

1) An SCM for compound-fault diagnosis is constructed to describe the process of counterfactual generation and inference. A novel generative adversarial module has been devised to decouple the non-causal feature subset from the fault features. It combines the compound-fault semantics to generate counterfactual fault features that can be fit to the ground-truth data distributions, which provides high-quality classification boundaries for the classifier. This balances the classification accuracies between single and compound faults and improves the interpretability and accuracy of the model.
2) In the generative adversarial module, we introduce a comparator in variational autoencoder and generative adversarial network (VAE-GAN) to enhance the discriminator's layer-wise focus on the semantics, thereby constraining the generative adversarial module to better decouple causal and non-causal fault feature subsets, and thus improve the quality of generation of counterfactual fault features, improving the accuracy of the models.

The rest of the article is organized as follows. Section II describes the current work related to GZSL and counterfactual inference in ZSL. Section III specifically describes our proposed method. Section IV presents the experimental results and analyses. Section V concludes this article with a summary of the proposed method and the experimental results.

II. Related Work

A. Generalized ZSL

The field of ZSL can be divided into traditional ZSL and GZSL based on the relationship between the train and the test sets. The train set comprises solely seen-class (single faults) samples, ZSL only needs to identify unseen-class (compound faults) samples. However, GZSL is confronted with the dual challenge of identifying both seen and unseen classes. At present, generative GZSL represents the dominant direction of research. The underlying premise is to generate unseen-class samples from semantics, thereby achieving ZSL tasks by supervised learning. Generative GZSL-based methods can be divided into three classes: autoencoder-based, GAN-based, and flow-model-based.

The autoencoder-based methods aim to learn the mapping between the sample and semantic space and reconstruct the unseen-class samples through semantics to identify the classes to which the ground-truth unseen-class samples belong. Kim and Shim [14] minimized the Wasserstein distance between the ground truth and the generated feature distribution to generate unseen-class features from attributes to achieve GZSL. Shao and Li [15] introduced multichannel multimodal VAEs into Gaussian mixture distributions to explore the relationship between semantic and feature space. Han et al. [16] proposed a hybrid GZSL framework that combines a generative model with an embedding model for final GZSL classification by mapping real and synthetic samples into an embedding space.

The GAN-based methods aim to train the generator to generate unseen-class samples from unseen-class semantics and random noise by gaming a pair of generator and discriminator, thus allowing the classifier to infer the classes of unseen-class samples. Sun et al. [17] imposed additional constraints on the GAN through a semantically enhanced cross-modal model to generate fine-grained unseen-class features. Tang et al. [11] propose a structurally aligned generative adversarial network framework to improve ZSL by mitigating semantic gaps, domain bias, and hubness issues. Verma et al. [18] proposed a meta-learning-based generative model that combines model-independent meta-learning with Wasserstein GAN (WGAN), which learns a generic parameter to generate seen and unseen class samples and improves model performance. Chen et al. [19] proposed to incorporate

semantic and visual mappings into a unified generative model to refine the visual features of seen and unseen class samples, and guided the feature refinement module to learn class- and semantically relevant representations through the adaptive marginal center loss and the semantic cyclic consistency loss.

The flow-based approaches model the feature transformation process as a mapping of the reversible neural network, thus reconstructing the samples by reverse mapping and inferring the unseen-class samples. Shen et al. [20] mapped samples to semantic and non-semantic spaces via a reversible neural network based on the flow model, and directly generated unseen-class samples via an inverse neural network.

The field of fault diagnosis based on GZSL is still in its infancy. Yue et al. [21] proposed a semi-supervised hybrid triplet network to learn the similarity between data and the matching between data and semantic descriptions, to reduce the effect of domain shift. Mou et al. [22] encode the fault features and semantic vectors of fault attributes from two different modalities as latent variables by two variational autoencoders and design Barlow matrices to measure the consistency between the distribution of fault features and fault semantic vectors.

Despite the improvements in classifying both single and compound faults by achieving data distribution alignment, the model interpretability is lacking, and the classification accuracy for single and compound faults is significantly imbalanced.

B. Counterfactual Inference

Counterfactual inference is one of three principal components of causal inference. It entails modifying observable variables to infer the cause of the resulting effect, analogous to human estimation of the significance of an event by imagining the potential outcomes of actions. This approach to causal inference has gained considerable prominence in recent years within the interpretability research of machine learning.

Sauer and Geiger [23] interpret image generation as the independent roles of background, shape, and texture to generate counterfactual images to be added to the training set. Chang et al. [24] construct counterfactual samples by altering a part of the image outside the manually labeled boundary to facilitate learning of invariant features in image classification. Madumal et al. [25] analyzed the action reasons of intelligence from a counterfactual perspective and constructed an SCM on the action effects of intelligence to provide reasonable explanations for their behavior.

The combination of GZSL and counterfactual inference has recently been explored. Tai and Guo [26] constructed a causal map to describe the relationship between images and Wikipedia descriptions. The method reduced the effect of negative causality on the relationship by varying the distribution of data and combined it with comparative learning to establish a cross-modal mapping relationship for the GZSL task. Yue et al. [27] proposed a generative causal model to generate counterfactual faithful samples from sample attributes and class attributes to balance classification accuracy between seen and unseen classes.

However, there is a paucity of research exploring the potential causal structure in data from the counterfactual inference perspective in the generalized zero-shot compound-fault diagnosis task. Furthermore, there is a dearth of research proposing interpretable generalized zero-shot compound-fault diagnosis models that achieve stable generalization from single faults to both single and compound faults.

III. Methodology

In GZSL, the training set contains a single-fault dataset, and the test set contains single and compound-fault datasets. We denote the datasets of single and compound faults as

$$(x_i, y_i, a_i)|x_i \in X_S, \quad y_i \in Y_S, \ a_i \in A_S, \ i \in [0, N_S] \text{ and}$$
$$(x_j, y_j, a_j)|x_j \in X_U, \quad y_j \in Y_U, \ a_j \in A_U, \ j \in [0, N_U]$$

where x_i and x_j are original fault samples in single-fault dataset X_S and compound-fault dataset X_U, y_i and y_j are fault labels corresponding to x_i and x_j, a_i and a_j are prior fault semantics corresponding to x_i and x_j, and N_S and N_U are the total sample numbers of the single and compound-fault datasets, respectively, and $Y_S \cap Y_U = \emptyset$.

Our method has three stages. The first stage trains the model with labeled single-fault samples X_S and single-fault semantics A_S and learns a model with the decoupling feature and generating feature capability, expressed as mapping $F(x_i, a_i; \theta_1) \rightarrow y_i$, where θ_1 is the trained model parameter. In the second stage, we freeze the model parameters (trained in the first stage) to conduct counterfactual generation and inference using samples to be tested, which are classified into single faults or compound faults. In the third stage, for single faults, we train and test directly using supervised learning with the help of a classifier, and for compound faults, we train the model using single faults of the first stage to obtain a mapping function $F(x_i, a_i; \theta_2) \rightarrow y_i$, which predicts the ground-truth fault labels of the fault samples recognized as compound faults in the second stage, and the θ_2 is the trained model parameter in this stage.

A. Feature Decoupling and SCM

1) Feature Decoupling: Feature decoupling is the separation of features with certain characteristics from the original features [28]. The fault features of bearings characterize the high-dimensional information of faults that occur in specific equipment and working conditions. From the perspective of signal decomposition, the fault features have background information that is closely related to the equipment and working conditions, as well as causal information that is related to the fault classes. From the causal theory point of view, according to the common cause principle elaborated by Reichenbach [29], the following holds.

Definition 1: If two observable variables X and Y are statistically dependent, there is a variable Z that causally affects both and accounts for all dependencies by making them independent when conditioned on Z.

It is reasonable to assume that there is a subset of causal features that directly determines the class of the fault signal in high dimensions. We decouple the causal feature subset A and

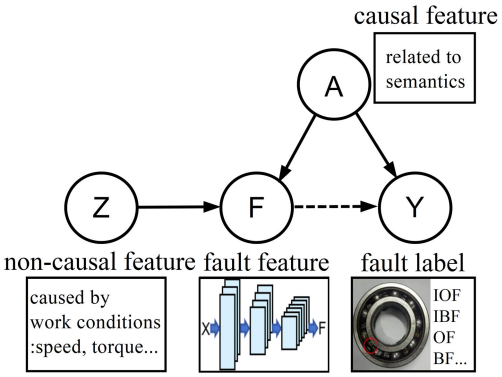


Fig. 1. SCM of the proposed method. F is the feature representation of the original fault sample X, Y denotes its corresponding class, Z denotes the non-causal feature that are not relevant to Y, and A denotes the causal feature that are truly relevant to Y.

the non-causal feature subset Z, meaning that given the fault feature F, we can obtain A and Z by sampling the marginal distributions $P(A|F)$ and $P(Z|F)$. The decoupled feature subsets are required to comply with the basic assumption that A and Z can be combined to generate F, that is, the fault feature F can be reconstructed by sampling the joint distributions $P(F|A, Z)$.

Furthermore, the decoupled causal and non-causal feature subsets are independent of each other according to the independent causal mechanism (ICM) principle [29].

Definition 2: The process of causal generation of system variables consists of autonomous modules that do not inform or influence each other. In probability, this means that the conditional distribution of each variable given its cause does not inform or influence other mechanisms.

We consider the fault features as a combination of causal and non-causal feature subsets, which do not interact with each other, meaning that when we change the value or distribution of one set of them, the other set will not change, and from this, we can conduct counterfactual generation based on the decoupled feature subsets.

2) SCM for Compound-Fault Diagnosis: The first step in counterfactual generation and inference is to decouple the causal and non-causal feature subsets from the fault features. To standardize the subsequent process, we construct a priori SCMs to describe the causal relationships among the objects in the compound-fault diagnosis, as illustrated in Fig. 1.

A spurious correlation exists between fault features F and fault labels Y. This is due to the influence of operating condition factors such as temperature and speed in F. The probability of the joint feature-label distribution may change significantly when the working conditions change or small fluctuations occur in the data. The SCM provides a theoretical descriptive framework for counterfactual generation and inference, where we separate A and Z from F by feature decoupling, and then separate single and compound faults through interpretable counterfactual inference.

B. Counterfactual Generation and Inference

As illustrated in Fig. 2, we generate counterfactual fault features F' for each class by performing counterfactuals

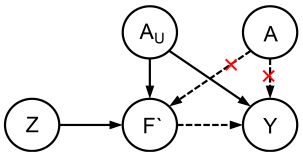


Fig. 2. Counterfactual generation step for compound-fault diagnosis. $\times$ denotes the interruption of the relationship due to modification of the causal feature subset A.

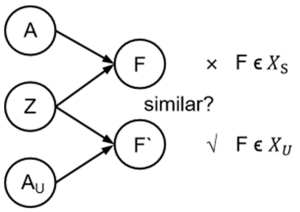


Fig. 3. Counterfactual inference step for compound-fault diagnosis.

on causal feature subsets in the to-be-tested fault samples. Subsequently, we employ correlation metrics to perform counterfactual inference and ascertain whether the samples belong to single or compound faults. The entire process is divided into two steps: counterfactual generation and counterfactual inference.

Involved in the counterfactual generation, we generate counterfactual samples through three counterfactual computation steps: 1) induction: recover the noise Z based on the endogenous variable F (original fault feature); 2) action: modify the endogenous variable A to A_U; and 3) prediction: generate the counterfactual fault feature F' as shown in Fig. 2. Specifically, we decouple the non-causal feature subset Z from the original fault features, then replace the causal feature subset A with the a priori compound-fault semantics A_U, and finally generate the counterfactual fault feature F' from A_U and Z.

Involved in counterfactual inference, we can predict fault classes based on the counterfactual consistency principle.

Definition 3: If the counterfactual fault features are similar to the to-be-tested fault features, the counterfactual semantics A_U is the underlying causal feature set $A(F)$ of the to-be-tested fault features. The to-be-tested fault samples are compound faults. Conversely, if the counterfactual fault features are not similar to the to-be-tested fault features, the counterfactual semantics A_U is different from the causal features set $A(F)$ of the to-be-tested fault, that is, the to-be-tested fault samples belong to a single fault.

According to Definitions 1 and 2, the model decouples the non-causal feature groups that are unrelated to the faults from the features. The causally related compound-fault semantics can be combined with non-causal features to generate all classes of compound-fault features (i.e., counterfactual fault features) following counterfactual generation. As shown in Fig. 3, if the generated counterfactual fault feature F' results in a correct classification, the causally related feature group $A(F)$ in the original fault feature is highly correlated with the compound-fault semantics A_U in the counterfactual fault feature. Hence, the fault is a compound one; otherwise, it is a single fault.

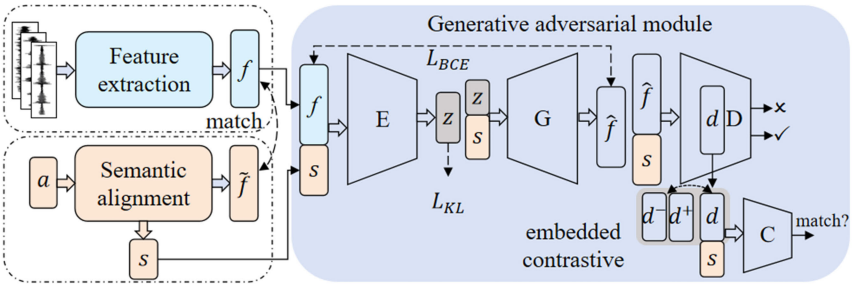


Fig. 4. Framework of the proposed CI-GZCFD.

C. Model Structure for CI-GZCFD

Our proposed generalized zero-shot compound-fault diagnosis model based on counterfactual inference comprises a feature extractor, a semantic alignment module, and a generative adversarial module, as illustrated in Fig. 4. The feature extraction module extracts fault features f from the original vibration signals, the semantic alignment module maps the prior fault semantics a to the fault feature space to obtain $\tilde{f}$, and the adjusted semantics s is obtained by modal matching, and the generative adversarial module contains an encoder E, a generator G, a discriminator D, and a comparator C. The encoder extracts the non-causal feature z from the fault features f and fault semantics s, the generator generates pseudo-fault features $\hat{f}$. The discriminator is responsible for identifying whether the fault features $f/\hat{f}$ are generated or not, and the comparator constrains the match between the hidden layer d of the discriminator and the fault semantics s.

1) Feature Extraction Module: The compound fault is defined as multiple single faults that are coupled with each other. To accurately extract compound-fault features, it is necessary to design an effective feature extraction module. We adopt the method in [9] to extract fault features from original vibration signals. As illustrated in Fig. 5, first, we extract the salient features with convolution-pooling layers. Subsequently, the salient features are weighted through spatial attention and channel attention layers. Ultimately, the features are mapped to the appropriate dimensions with fully connected layers.

In brief, we can obtain the fault features f from the input original vibration signal x with the following equation:

$$f = \mathrm{Fc}(\mathrm{Sp} \otimes [\mathrm{Ch} \otimes [\mathrm{Pool}(\mathrm{Conv}(x))]]) \tag{1}$$

where Fc denotes the fully connected layer, and Sp() and Ch() denote the spatial and channel attention layers, respectively. $a \otimes b$ denotes the feature weighting of a over b, which is equivalent to $a(b) * b$. Conv() and Pool() denote the convolution and pooling layers, respectively.

In the training and testing phases, the extracted fault features and fault labels are computed using the cross-entropy loss function as follows:

$$L_c = \sum_{i}^{n} y_i \log p_i + (1 - y_i) \log(1 - p_i) \tag{2}$$

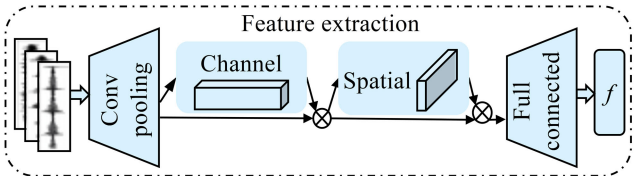


Fig. 5. Feature extraction module.

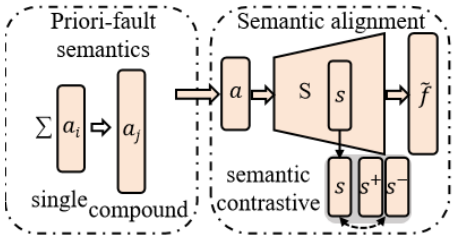


Fig. 6. Semantic alignment module.

where y_i is the label of the original vibration signal, p_i is the predicted probability of feature extractor, and n is the number of fault classes.

2) Semantic Alignment Module: In ZSL, the key to achieving generalization from single to compound faults is the semantics, which describe both the single and compound faults. Our method employs the semantic alignment module [9] to construct single and compound-fault semantics. As illustrated in Fig. 6, the 28-D time-frequency statistical features of the vibration signals are aligned with the fault feature distributions to adjust the priori fault semantics (28-D statistical features).

We construct the prior fault semantics a_i of single faults with 28-D statistical characteristics of vibration signals, where $a_i \in A_S$. The prior fault semantics of compound faults are defined based on the single-fault semantics

$$A_U = \left\{ \sum_{i=1}^{k} \lambda_i a_i | \forall a_i \in A_S, \sum_{i=1}^{k} \lambda_i = 1 \right\} \quad \forall a_j \in A_U \tag{3}$$

where the k is the number of single-fault classes.

We map the prior fault semantics a to the subspace through fully-connected layers, and we map the adjusted semantics s in the subspace to the fault feature space through fully-connected layers for the Euclidean distance metric. The comparative loss ensures class discrimination of the adjusted semantics s_i, and the mse loss ensures alignment of the prior fault semantics

with the fault features. The loss of the semantic alignment module is as follows:

$$L_S = -\log \frac{\exp\left(s_i^T s^+ / \tau_s\right)}{\exp\left(s_i^T s^+ / \tau_s\right) + \sum_{m=1}^{M} \exp\left(s_i^T s_m^- / \tau_s\right)} - E\|\widetilde{f} - f\|_2 \tag{4}$$

where τ_s is the temperature coefficient of comparative loss. s_i is the adjusted semantics in the subspace, s^+ and s_m^- are the unique positive sample and all M negative samples, respectively. The $\widetilde{f}$ is the projection of s_i in the fault feature space, and f is the fault feature.

3) Generative Adversarial Module: In the second stage of our method, the generative adversarial module generates counterfactual fault features, and the discriminator D must identify counterfactual fault features that are combinations of causal and non-causal feature subsets. The constraints of D on the generator at the input, conditional on the semantics of the faults, are weak, and the constraints must be increased to make the model resistant to distributional disturbances induced by the counterfactual fault features.

Since the labels of the counterfactual fault features are highly correlated with the causal feature subset describing the class information, we add a new constraint to the hidden layer of D. We use the comparator C to measure the correlation between the fault features and the fault semantics and constrain the hidden layer features to cluster by class through contrastive loss. We encode the fault features and the fault semantics as latent features by the encoder of VAE-GAN and reconstruct the fault features from them by the decoder, with the loss represented as $L_{\text{VAE-GAN}}$

$$\begin{aligned} L_{\text{VAE-GAN}} = {} & \text{KL}(E(f,s)\|p(z|s)) \\ & - E_{E(f,s)}\left[\log G(z,s)\right] \\ & + E\left[D(f,s)\right] - E\left[D\left(\hat{f},s\right)\right] \\ & - \lambda E\left[\left(\|\nabla D\left(\overline{f},s\right)\|_2 - 1\right)^2\right]. \end{aligned} \tag{5}$$

This loss function has four terms: a KL divergence, a generator loss, a discriminator loss, and a gradient penalty. KL() denotes the Kullback–Leibler divergence, which measures the difference between the latent variable distribution of the encoder output and the prior distribution (following the Gaussian distribution) for decoupling fault features. $E_{E(f,s)}[\log G(z,s)]$ is the generator loss, which encourages the generator to generate realistic samples. $E[D(f,s)] - E[D(\hat{f},s)]$ is the discriminator loss, which encourages the discriminator to be able to distinguish between generated and ground-truth samples. The last term denotes the gradient penalty, which ensures 1-Lipschitz continuity and the smoothness of output. $\overline{f}$ is random sample of fault feature f and generated fault feature $\hat{f}$.

The loss of the comparator added in the discriminator of VAE-GAN is denoted as L_{dc}

$$L_{\text{dc}} = -\log \frac{\exp\left(d_i^T d^+ / \tau_d\right)}{\exp\left(d_i^T d^+ / \tau_d\right) + \sum_{m=1}^{M} \exp\left(d_i^T d_m^- / \tau_d\right)} - \log \frac{\exp\left(C\left(d_i, s^+\right)/\tau_d\right)}{\sum_{n=1}^{N} \exp(C(d_i, s_n)/\tau_d)}. \tag{6}$$

The first term is a supervised loss that computes an $M+1$-way classification loss for each batch so that the unique positive sample d^+ and all other M negative samples d_m^- are far away from each other. This allows for the aggregation of the instances of the same class and the separation of the dissimilar ones. d_i^T denotes the transposed hidden layer feature of D. The second term is comparator loss, which computes the correlation between D's hidden layer feature d_i and the unique positive semantics s^+. The semantics with the same class as d_i in the first term are regarded as positive semantics s^+, and the rest are negative semantics. This loss function optimizes the correlation between the D's hidden layer and the corresponding semantics. τ_d is the temperature coefficient.

D. Model Processing Procedure

The proposed method includes three stages, as shown in Fig. 7. The first stage trains the decoupling and generation ability of the model with single-fault samples and a priori single-fault semantics. We input single-fault samples x and prior single-fault semantics a into the model, the feature extractor extracts fault features f from the fault samples, the semantic alignment module aligns the prior single-fault semantics with the fault features f to get the adjusted single-fault semantics s, and in the generative adversarial module, the encoder E decouples the non-causal feature subset z from the single-fault features and single-fault semantics. It combines the single-fault semantics into the generator to generate single-fault features, by which the process allows the model to decouple and generate fault features.

The second stage involves counterfactual generation and counterfactual inference. In the counterfactual generation process, given an arbitrary class of to-be-tested fault samples, which can be single or compound faults, we input the to-be-tested fault samples x into the model, and the feature extractor extracts the fault features f from x. In the generative adversarial module, the encoder E extracts the non-causal feature subset z from the fault features, and the priori compound-fault semantics a_j are adjusted by the semantics alignment module to obtain s_j, which is combined with the non-causal feature subset to generate the counterfactual fault feature f' via the generator G. In the counterfactual inference process, we predict the class of the to-be-tested fault samples by the classifier and measure the similarity between the ground-truth fault features and the counterfactual fault features. According to the counterfactual consistency principle (Definition 3), if the counterfactual fault feature is similar to the to-be-tested fault sample, it is inferred to be the compound fault; otherwise, it is a single fault, which implements the binary classification of single and compound faults, as shown in Fig. 8.

The third stage consists of single-fault classification and compound-fault classification. For single-fault to-be-tested samples X_S, we use traditional supervised training, where the classifier is trained using single-fault samples from the training set to infer single-fault samples from the test set. For compound-fault test samples X_U, we use the traditional ZSL method in the training phase. We input single-fault samples into the proposed model and obtain single-fault features f_S and single-fault semantics s_i by the feature extraction module

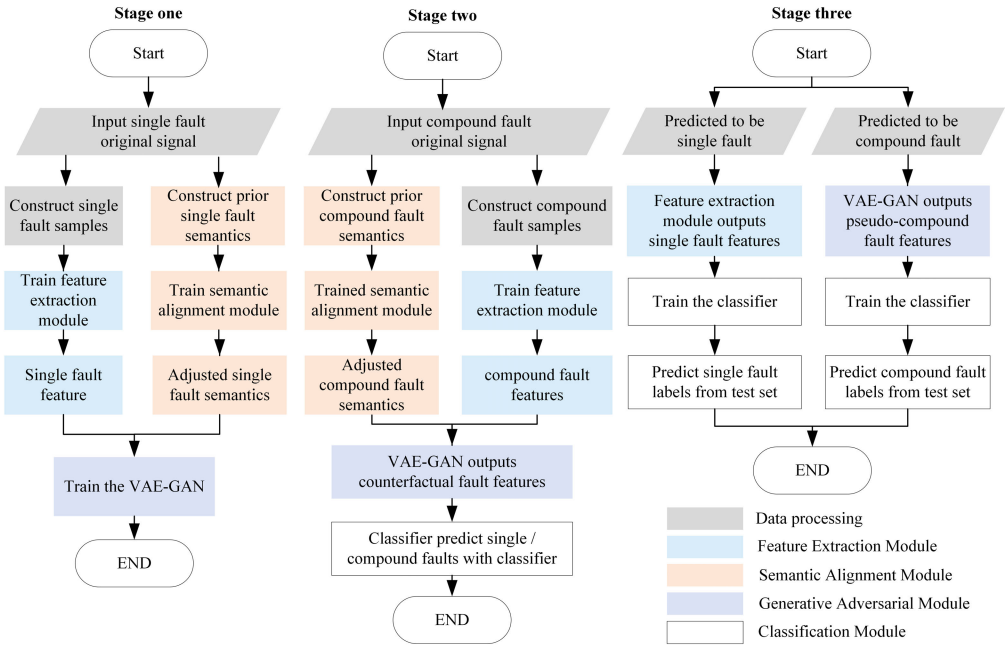


Fig. 7. Three-stage flowchart of the proposed CI-GZCFD.

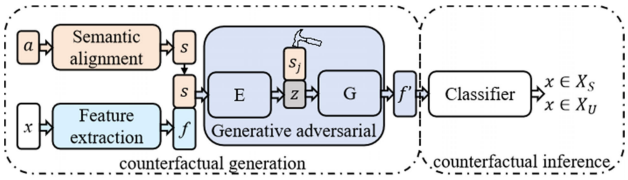


Fig. 8. Test flow in the second stage.

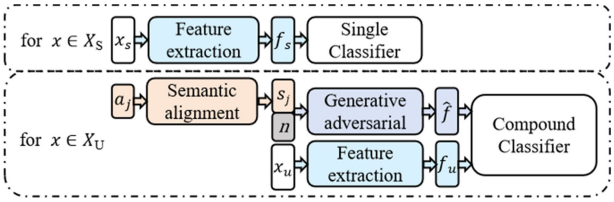


Fig. 9. Test flow in the third stage.

and semantic alignment module. Input single-fault semantics and random noise n into the generator to generate pseudo-single-fault features by generative adversarial training; In the testing phase, the compound-fault samples X_U and a priori compound-fault semantics a_j are input into the trained model, and the to-be-tested compound-fault features f_U, adjusted compound-fault semantics s_j, and pseudo-compound-fault features $\hat{f}_U$ are obtained from the feature extraction module, semantic alignment module, and generator, respectively. The pseudo-compound-fault features are used to train the classifier, which is able to identify the to-be-tested compound-fault features as illustrated in Fig. 9.

IV. Experimental Results and Discussion

To assess the proposed model, an experimental platform was constructed (as shown in Fig. 10), and a series of experiments were conducted to acquire vibration signals from bearings.

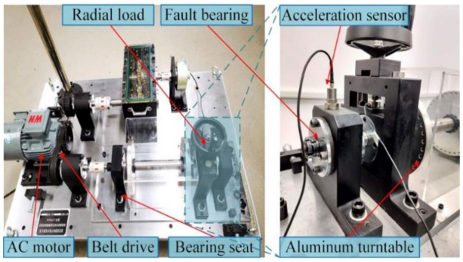


Fig. 10. Self-built experiment platform for signal acquisition.

The experimental platform comprises an ac variable-frequency motor and a shaft system comprising front-end bearings, bearing housings, radial load bearings, and fault bearings. The bearing type is a deep groove ball bearing 6205. Outer and inner ring faults are cut by a machine with a width of 0.5 mm, and roller faults are formed by pitting corrosion. Accelerometers mounted on the bearing housings, the sampling frequency is 51.2 kHz, the load is 0 HP, the rotational speed is 1500 r/min, and the sampling time for all classes of faults lasts 10 s. The dataset comprises data from healthy bearings, three classes of single faults, and four classes of compound faults. The healthy bearing is denoted as H, and the three classes of single faults are the inner-ring fault (IF), the outer-ring fault (OF), and the roller fault (BF), and the four classes of compound faults are the inner-ring& OF (IOF), inner-ring&roller fault (IBF), outer-ring&roller fault (OBF), and inner-ring&outer-ring&roller fault (IOBF). A total of 1000 samples were obtained for each class of faults, with the step size at 500 and the window length at 2048. Min-Max normalization is performed before the fault samples are fed into the model, mapping the original vibration data between [0, 1] by linear variation.

In our experiments, accuracy is used for evaluating the performance of single and compound faults. In addition, the

TABLE I
EXPERIMENT TASK SETTING. THE TESTING SET SIZE IS 400 FOR ALL TASKS

Task	Training class	Training samples per class	Testing class
Task A	H,IF, OF,BF	100	H,IF, OF,BF, IOF,IBF, OBF,IOBF
Task B		200	
Task C		300	
Task D		400	

harmonic average accuracy (denoted with A_h) is used to provide a balanced view of the overall performance for both types of faults

$$A_h = 2A_s A_u/(A_s + A_u) \tag{7}$$

where A_s and A_u denote the classification accuracy of single and compound faults, respectively. The accuracy is calculated as follows:

$$\frac{1}{nm}\sum_{i=0}^{n}\sum_{j=0}^{m}\mathbb{I}\left(\max\left(y_{0,j},\ldots,y_{i,j},\ldots,y_{n,j}\right)=y_i\right) \tag{8}$$

where n is the number of classes, m is the number of samples in each class, $y_{i,j}$ is the probability that the jth sample is predicted to be the class i. The max() outputs the class label that has the greatest probability. $\mathbb{I}$ denotes the indicator function that returns 1 when the output of the max function equals the class label y_i, otherwise returns 0.

A. Impact of Training Set Size

Table I summarizes the four experimental groups, where the number of training samples for each class varies from 100 to 400, and the number of test samples for each class is 400. In the training phase, the classes include four single faults. In the testing phase, the examples are from eight classes, including both single and compound faults. The training and testing sets are balanced. To mitigate the impact of random fluctuation, we conduct five repetitions for each experimental group and report the average accuracy and standard deviation.

Fig. 11 depicts the harmonic average accuracy of the four experimental groups. As the number of training examples increases, the classification accuracy of the model increases gradually. When the number of training examples is 400 (i.e., Task D), the harmonic average accuracy A_h reaches 87.4%. The standard deviation remains similar and small in all four cases, which implies the robustness of our method.

As shown in Fig. 11, we obtained an F-statistic value of 73.50 and a P-value of 2.03E-15 by conducting one-way ANOVA. The large F-statistic value indicates that the change in A_h between Tasks A through D is much larger than the change in A_h under the same task. The small P-value indicates significant differences in model performance for Tasks A through D.

In addition, we calculated 95% confidence intervals for A_h in the task, which reflects interval estimates of the overall experimental results based on our results. The confidence interval variation reflects the estimation precision of the overall experimental results, and the narrower the interval is, the more precise the estimation is. For example, the confidence interval variation for Task A is 77.96 minus 75.10. The confidence intervals for Tasks A through D are relatively narrow, which indicates that the existing experimental results are reliable in estimating the accuracy of our proposed method.

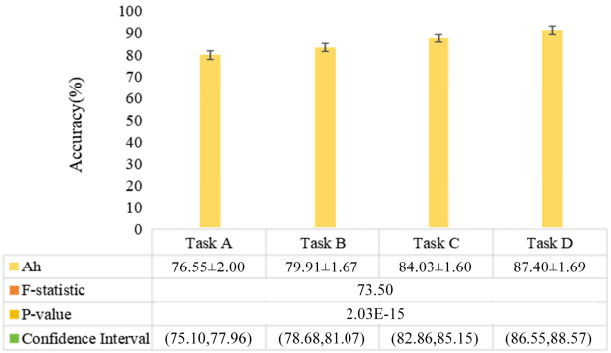


Fig. 11. Classification accuracy with different numbers of training samples.

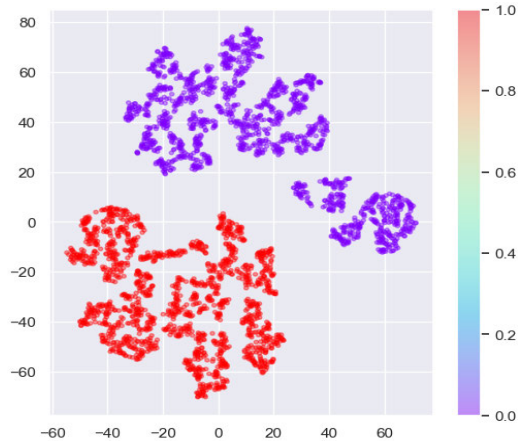


Fig. 12. Visualization results of the generated counterfactual fault features.

B. Performance Analysis of Counterfactual Inference

To evaluate the performance of the second stage of counterfactual generation and inference, we use the feature maps of the counterfactual generation for qualitative analysis and the results of the counterfactual inference for quantitative analysis. First, the quality of the counterfactual-generated features is critical and directly determines whether the samples to be tested are classified as single or compound faults. The results based on the feature visualization are shown in Fig. 12, where there is a clear demarcation between counterfactual-generated features for single and compound faults that very much facilitates the classifier to perform the classification.

In addition, we compute two accuracies in the second stage of the counterfactual inference process: Binary_S and Binary_U. Binary_S is the accuracy for single faults, and Binary_U is the accuracy for compound faults, which are calculated as follows:

$$\frac{1}{m}\sum_{j=0}^{m}\mathbb{I}\left(\max\left(y_{0,j},y_{1,j}\right)=y_{*,j}\right) \tag{9}$$

where m is the number of samples. $y_{0,j}$ and $y_{1,j}$ denote the probability that the jth sample is predicted to be 0 and 1,

TABLE II
ACCURACY OF COUNTERFACTUAL INFERENCE

	Task A	Task B	Task C	Task D
Binary_S	0.91	0.93	0.95	0.96
Binary_U	0.87	0.90	0.93	0.96

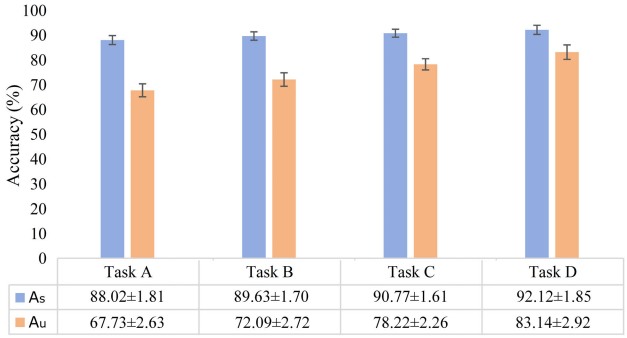


Fig. 13. Average accuracy with different numbers of training samples.

respectively. The max() outputs the class label that corresponds to the greater probability. $\mathbb{I}()$ denotes the indicator function. In computing Binary_S, it returns 1 when the output of the max function equals $y_{0,j}$, otherwise returns 0. When computing Binary_U, it returns 1 when the output of the max function equals $y_{1,j}$, otherwise returns 0. As shown in Table II, Binary_S varies from 0.91 to 0.96 and Binary_U varies from 0.87 to 0.96 for tasks changing from A to D. The relatively high accuracies indicate that ground-truth compound faults are not mostly misclassified as single faults, which is attributed to counterfactual generation and inference.

C. Performance of Single and Compound Faults

Fig. 13 depicts the average accuracy of the four experimental groups in Table I. The prediction accuracy of single-fault samples depends on the appropriate single and compound faults classification of the model. As shown in this figure, the accuracy of the single-fault classification is around 90% and the performance increases with more training examples. This indicates that the model is capable of effectively separating the single and compound faults. In addition, the proposed model achieves 83.14% accuracy for the classification of compound-fault samples and has a greater performance increment rate compared to the performance of classifying single faults. As we increase the number of training examples, the difference in performance of classifying single and compound faults reduces. The difference between A_s and A_u of Task A is 20.29%, and it reduces to 8.98% of Task D. It is evident that the proposed model effectively differentiates single- and compound-fault samples and achieves a high accuracy under their respective superior methods. This alleviates the requirement of the model on the classification of to-be-tested fault samples and improves its generality.

In the third stage, the generation quality of different compound-fault classes determines the classification ability of the traditional ZSL method, for this reason, we visualize the generated compound-fault features. As shown in Fig. 14, the different classes of compound faults have large class margins.

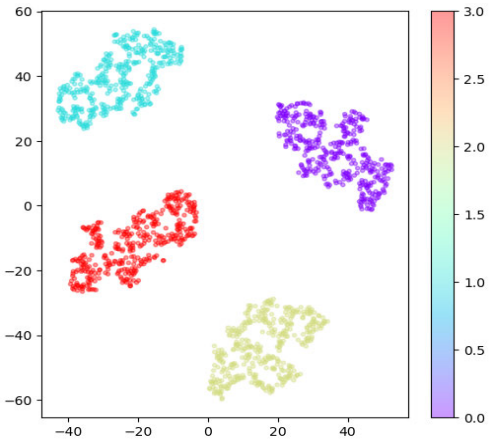


Fig. 14. Visualization results of generated compound-fault features.

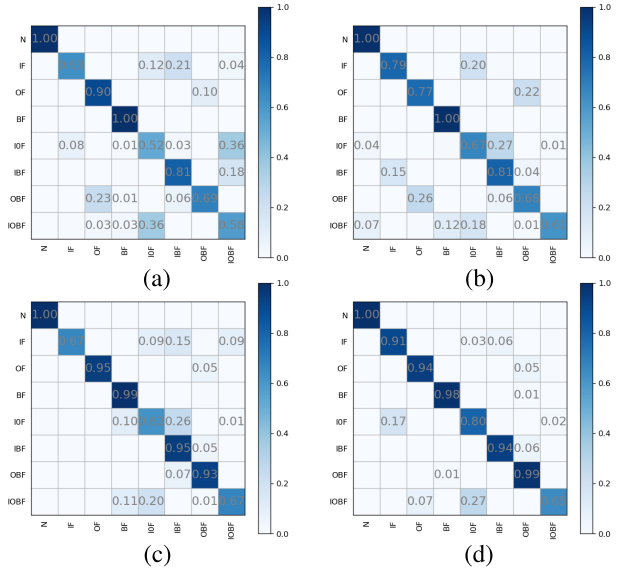
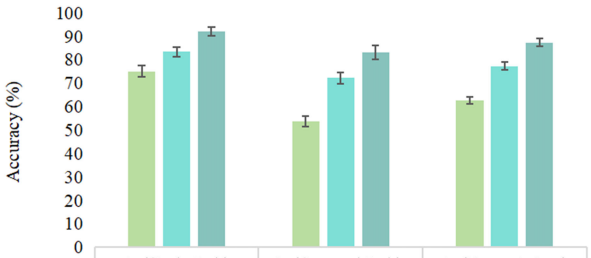


Fig. 15. Confusion matrix of the four tasks. (a) Task A. (b) Task B. (c) Task C. (d) Task D.

Fig. 15 illustrates the confusion matrix of the classification results of the four tasks to show the test results in detail. The vertical axis of this matrix represents the ground-truth classes, and the horizontal axis represents the predicted classes. Fig. 16(d) depicts that the model performs well in the prediction of single-fault samples, and there exists only a weak probability of identifying single faults as compound faults. However, the model performs slightly worse on the prediction of the compound-fault samples; there exists a probability of 0.17 to identify the IOF of the compound faults as the IF of the single faults, which is due to the prediction loss in the second stage. There exists a probability of 0.27 to identify the IOBF as IOF, which is due to the prediction loss in the third stage, it may be that the existing model is unable to avoid the IOBF from being confused with the other compound faults, as it is the most complex compound fault coupled by three single faults.

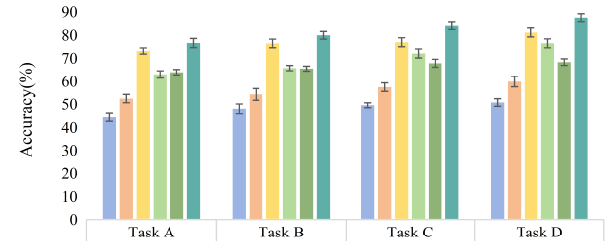
D. Ablation Experiments

We conducted a comparative analysis of the benchmarks VAE-GAN, VAE-GAN + L_{dc}, and our proposed CI-GZCFD



	A_s (Single Fault)	A_u (Compond Fault)	A_h (Harmonic Ave.)
Baseline	75.10±2.40	53.65±2.21	65.54±1.50
Baseline+Ldc	83.35±2.05	72.15±2.26	77.32±1.57
Ours	**92.12±1.85**	**83.14±2.92**	**87.40±1.69**

Fig. 16. Average accuracy of our method including different components.



	Task A	Task B	Task C	Task D
ZSML	44.41±1.75	47.97±2.07	49.50±1.06	50.71±1.66
LDS-IFD	52.39±1.81	54.21±2.56	57.38±1.88	59.80±2.29
CGASNet	73.12±1.32	76.38±1.86	76.90±1.95	81.15±1.97
PREE	62.93±1.54	65.67±1.15	72.04±1.95	76.42±1.93
CE-GZSL	63.84±1.20	65.43±1.09	67.81±1.72	68.29±1.43
OURS	**76.55±2.00**	**79.91±1.67**	**84.03±1.60**	**87.40±1.69**

Fig. 17. Average accuracy of the compared methods.

on task D. The L_{dc} corresponds to (6), which is an enhancement of the supervision of the discriminator D based on the VAE-GAN, and on which counterfactual inference is introduced to form the CI-GZCFD model. The accuracy of our experiments is presented in Fig. 16. The baseline A_h is 62.69%, and the single-fault classification accuracy is considerably higher than that of the compound faults. This is a consequence of the intrinsic domain drift issue inherent to GZSL. The A_h of VAE-GAN + L_{dc} is enhanced by 14.76% in comparison to that of VAE-GAN, which can be attributed to the powerful supervisory effect of the discriminator D on the decouple and generation process. The proposed CI-GZCFD demonstrates a substantial enhancement in comparison to both of them. The improvement in prediction accuracy for the compound faults is particularly noteworthy, which can be attributed to the distinction between single- and compound-fault samples under the influence of counterfactual generation and inference. This results in the optimal performance of the proposed model in the generalized zero-shot compound-fault diagnosis, with A_h reaching 87.40%.

E. Comparison Study

We compare our method with five state-of-the-art GZSL models, including ZSML [18], LDS-IFD [8], CGASNet [12], PREE [19], and CE-GZSL [16]. To ensure a fair comparison, different models are assigned the same training and testing sets in each experimental group. The same features and semantics are employed to train and test the models. To obtain a representative result, each experiment is repeated five times, and the average value is calculated.

Fig. 17 illustrates the average accuracy of different methods. In all four experimental groups, the test accuracy of each model demonstrated an increase in line with the number of training samples. ZSML exhibits the lowest prediction accuracy. This is because ZSML relies on a large number of meta-tasks to tune the model parameters, but only four single-fault samples are available for training, which severely limits performance. While LDS-IFD shows slight improvement, its efficacy remains constrained by the limited number of training samples, which restricts the dimensionality of the fault semantics. The CE-GZSL model demonstrates satisfactory performance on four tasks, largely due to its utilization of instance-level comparative supervision and class comparative supervision in the embedding space. This method effectively mitigates the impact of domain drift. The accuracy of PREE is 76.42%, which is attributed to the improved feature extraction method and the quality of the feature representation by adjusting the intra-class consistency and inter-class variation. The accuracy of CGASNet is 81.15%, which is second only to our proposed model. This is because the proposed method employs contrast generation and an adaptive smoothing network to generate fault features, and the class confidence is efficiently estimated by the fault identification module, thus alleviating the class bias problem in GZSL. In contrast, the proposed model demonstrates superior performance compared to the other five models across all four groups of experiments. For instance, it exhibits a 10.98% improvement in prediction accuracy compared to the suboptimal model in Task D.

V. Conclusion

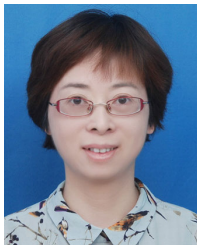
This article addresses the issues of model non-interpretability and imbalance in classification accuracy between single and compound faults by proposing a generalized zero-shot compound-fault diagnosis model based on counterfactual inference. The model effectively identifies both single and compound faults using only single-fault samples for training, by decoupling causal and non-causal features and generating counterfactual fault features. This method narrows the accuracy gap between different fault classes, achieving a harmonic average accuracy of 87.4% and reducing the performance gap from 20.29% to 8.98%. The proposed method outperforms state-of-the-art methods, offering higher and more balanced accuracies and significantly enhancing model generalizability.

The proposed method includes three stages of training and testing, which could be simplified for a lighter computational cost. In our future studies, we will explore a simple model structure to reduce the computational cost of the model to improve its efficiency. In addition, we will work on the causal discovery technique to mine the real causal structure inside the data, address the needs of industrial intelligent fault diagnosis, and explore explainable artificial intelligence.

References

[1] Y. Tang, C. Zhang, J. Wu, Y. Xie, W. Shen, and J. Wu, "Deep learning-based bearing fault diagnosis using a trusted multiscale quadratic attention-embedded convolutional neural network," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–15, 2024.

[2] S. Tian, D. Zhen, H. Li, G. Feng, H. Zhang, and F. Gu, "Adaptive resonance demodulation semantic-induced zero-shot compound fault diagnosis for railway bearings," *Measurement*, vol. 235, Aug. 2024, Art. no. 115040.
[3] C. Yan, X. Chang, M. Luo, H. Liu, X. Zhang, and Q. Zheng, "Semantics-guided contrastive network for zero-shot object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, pp. 1530–1544, 2021.
[4] J. Zhang, Q. Zhang, X. He, G. Sun, and D. Zhou, "Compound-fault diagnosis of rotating machinery: A fused imbalance learning method," *IEEE Trans. Control Syst. Technol.*, vol. 29, no. 4, pp. 1462–1474, Jul. 2021.
[5] G. Kwon and G. A. Regib, "A gating model for bias calibration in generalized zero-shot learning," *IEEE Trans. Image Process.*, early access, Mar. 1, 2022, doi: 10.1109/TIP.2022.3153138.
[6] L. Feng and C. Zhao, "Fault description based attribute transfer for zero-sample industrial fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1852–1862, Mar. 2021.
[7] J. Xu, S. Liang, X. Ding, and R. Yan, "A zero-shot fault semantics learning model for compound fault diagnosis," *Expert Syst. Appl.*, vol. 221, Jul. 2023, Art. no. 119642.
[8] S. Xing, Y. Lei, S. Wang, N. Lu, and N. Li, "A label description space embedded model for zero-shot intelligent diagnosis of mechanical compound faults," *Mech. Syst. Signal Process.*, vol. 162, Jan. 2022, Art. no. 108036.
[9] J. Xu, K. Li, Y. Fan, and X. Yuan, "A label information vector generative zero-shot model for the diagnosis of compound faults," *Expert Syst. Appl.*, vol. 233, Dec. 2023, Art. no. 120875.
[10] C. Yan et al., "ZeroNAS: Differentiable generative adversarial networks search for zero-shot learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 9733–9740, Dec. 2022.
[11] C. Tang, Z. He, Y. Li, and J. Lv, "Zero-shot learning via structure-aligned generative adversarial network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6749–6762, Nov. 2022.
[12] J. Xu, H. Zhang, W. Chen, Y. Fan, and X. Ding, "CGASNet: A generalized zero-shot learning compound fault diagnosis approach for bearings," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–11, 2024.
[13] Q. Guo, G. Li, and J. Lin, "A domain generalization network exploiting causal representations and non-causal representations for three-phase converter fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–13, 2024.
[14] J. Kim and B. Shim, "Generalized zero-shot learning using conditional Wasserstein autoencoder," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 3413–3417.
[15] J. Shao and X. Li, "Generalized zero-shot learning with multi-channel Gaussian mixture VAE," *IEEE Signal Process. Lett.*, vol. 27, pp. 456–460, 2020.
[16] Z. Han, Z. Fu, S. Chen, and J. Yang, "Contrastive embedding for generalized zero-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2371–2381.
[17] H. Sun, J. Wei, Y. Yang, and X. Xu, "Semantic enhanced cross-modal GAN for zero-shot learning," in *Proc. ACM Multimedia Asia*, Dec. 2021, pp. 1–7.
[18] V. K. Verma, D. Brahma, and P. Rai, "Meta-learning for generalized zero-shot learning," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 6062–6069.
[19] S. Chen et al., "FREE: Feature refinement for generalized zero-shot learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 122–131.
[20] Y. Shen, J. Qin, L. Huang, L. Liu, F. Zhu, and L. Shao, "Invertible zero-shot recognition flows," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 614–631.
[21] J. Yue, J. Zhao, and C. Zhao, "Similarity makes difference: SSHTN for generalized zero-shot industrial fault diagnosis by leveraging auxiliary set," *IEEE Trans. Ind. Informat.*, vol. 20, no. 5, pp. 7598–7607, May 2024.
[22] M. Mou, X. Zhao, K. Liu, and Y. Hui, "Variational autoencoder based on distributional semantic embedding and cross-modal reconstruction for generalized zero-shot fault diagnosis of industrial processes," *Process Saf. Environ. Protection*, vol. 177, pp. 1154–1167, Sep. 2023.
[23] A. Sauer and A. Geiger, "Counterfactual generative networks," in *Proc. 2021 Int. Conf. Learn. Represent. (ICLR)*, 2021, pp. 1–25.
[24] C.-H. Chang, G. A. Adam, and A. Goldenberg, "Towards robust classification model by counterfactual and invariant data generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15207–15216.
[25] P. Madumal, T. Miller, L. Sonenberg, and F. Vetere, "Explainable reinforcement learning through a causal lens," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, Apr. 2020, pp. 2493–2500.
[26] J. Tai and Y. Guo, "Unsupervised zero-shot learning for achieve cross-modal alignment with counterfactuals," in *Proc. ECAI*, 2023, pp. 2290–2297.
[27] Z. Yue, T. Wang, Q. Sun, X. Hua, and H. Zhang, "Counterfactual zero-shot and open-set visual recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15399–15409.
[28] M. Yang, F. Liu, Z. Chen, X. Shen, J. Hao, and J. Wang, "CausalVAE: Disentangled representation learning via neural structural causal models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9588–9597.
[29] B. Schölkopf et al., "Towards causal representation learning," *Proc. IEEE*, vol. 109, no. 5, pp. 612–634, Feb. 2021.



Juan Xu (Member, IEEE) received the Ph.D. degree from the School of Computer and Information, Hefei University of Technology, Hefei, China, in 2012.

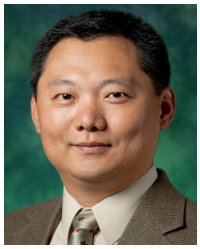
She is currently an Associate Professor at Hefei University of Technology. Her research interests include Industrial IoT and intelligent fault diagnosis, prognostics and health management, and predictive maintenance.

Hui Kong is currently pursuing the M.S. degree in information and communication engineering with Hefei University of Technology, Hefei, China.

His current research interests include deep learning and transfer learning methods for intelligent fault diagnostics and prognostics.

Xu Ding (Member, IEEE) received the B.S. and Ph.D. degrees from the School of Computer and Information, Hefei University of Technology, Hefei, China, in 2006 and 2015, respectively.

He is currently an Associate Research Fellow with the Institute of Industry and Equipment Technology, Hefei University of Technology. His research field mainly lies in wireless communications and wireless sensor networks.



Xiaohui Yuan (Senior Member, IEEE) is an Associate Professor at the University of North Texas, Denton, TX, USA, and the Director of the Computer Vision and Intelligent Systems Laboratory. His research interests include artificial intelligence, computer vision, and machine learning.

Dr. Yuan was a recipient of the Ralph E. Powe Professor Award in 2008 and U.S. Air Force Visiting Professor Award in 2011, 2012, and 2013. He serves as an associate editor, an editorial board member, and a guest editor for several journals, and an organizing member for many international conferences.