Cross-Domain Augmentation Diagnosis: An Adversarial Domain-Augmented Generalization Method for Fault Diagnosis under Unseen Working Conditions

Qi Li[1, 2], Liang Chen[1,*], Lin Kong[3], Dong Wang[4], Min Xia[5], Changqing Shen1, *

[1]*School of Mechanical and Electric Engineering, Soochow University, Suzhou 215131, P.R. China.*

[2]*Department of Mechanical Engineering, Tsinghua University, Beijing 100084, P.R. China.*

[3]*Chang Guang Satellite Technology CO.LTD, Changchun, 130000, P.R. China.*

[4]*Department of Industrial Engineering and Management and with the State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai, 200000, P.R. China*

[5]*Department of Engineering, Lancaster University, Lancaster, LA1 4YW, U.K.*

Abstract

Intelligent fault diagnosis based on domain adaptation has recently been extensively researched to promote reliability of safety-critical assets under different working conditions. However, target data may be inaccessible in the model training phase, resulting in the degradation or failure of the diagnosis model. Therefore, this paper introduces a new idea called cross-domain augmentation (CDA) to achieve diagnosis under unseen working conditions, which are frequently occurred in industrial scenarios. To realize this idea, an adversarial domain-augmented generalization (ADAG) method is proposed with domain augmentation via convex combination of data and feature-label pairs. Through adversarial training on multi-source domains and the augmented domain, ADAG enables learning generalized and augmented features, which are proximal representation in the unseen domain, facilitating the generalization ability of the model. Moreover, feature extractor and domain classifier are optimized as adversaries in model training to obtain domain-invariant features, while the fault classifier is trained to identify the features. Extensive experiment studies indicate that ADAG can successfully solve the cross-domain diagnosis problem unseen working conditions. For SDUST case study, ADAG promotes the model accuracy by 1.44%; while for a more challenging Ottawa case study, it promotes the model accuracy by 5.34%. Moreover, the domain discrepancy is reduced by 4.6%.

Keywords: Domain augmentation; Fault diagnosis; Unseen working condition; Rotating machinery; Domain generalization.

1. Introduction

1.1. Background

Aiming to increase reliability of assets, advanced fault diagnosis technology has been adapted in various industrial fields such as manufacturing, aerospace and renewable energy [1–3]. These approaches can significantly enhance the safety of safety-critical engineering systems [4]. Driven by industrial data explosion owing to sensor technology and the

*Corresponding authors:
Email addresses: chenl@suda.edu.cn (L. Chen), cqshen@suda.edu.cn (C. Shen)

internet of things, intelligent fault diagnosis has developed rapidly, playing a pivotal role in Industry 4.0 [5]. Therefore, data-driven intelligent fault diagnosis methods for prognostics and health management (PHM) have attracted extensive attention from researchers [6]. The intelligent fault diagnosis mainly leverages machine learning, such as convolutional neural network (CNN), autoencoder, and deep belief network, to extract fault features from big industrial data [7,8].

However, due to the variations of the signal pattern under varying and complex working conditions, the different working conditions trigger the domain shift problem, which commonly violates the i.i.d. assumption i.e., the training data from the source domains and the testing data from target domain obey the same distribution independently. Considering the existence of domain shift, the empirical risk minimization (ERM) principle is invalid [9], whereas the diagnosis model trained in a specific domain fails to generalize the diagnosis knowledge to an unseen working condition [10]. Thus, catastrophic degradation of diagnostic performance occurs.

1.2. Fault diagnosis with domain adaptation

Recently, domain adaptation (DA) has offered a solution to the obstruction above by observing the distribution discrepancy between the source and the target domain in model training [11]. DA aims to learn a shared representation between training and testing data, and thus enables the fault classifier to identify the representation from the target domain. Generally, DA usually takes advantage of distribution alignment or adversarial learning to learn the cross-domain features, which have the capability of narrowing down the domain shift[12]. Following the distribution alignment idea, Wang et al. [12] proposed intra-class maximum mean discrepancy (MMD) with multi-scale ResNet to shorten the conditional distribution discrepancy of the vibration signal. Hu et al. [13] introduced tensor-aligned invariant subspace learning to learn a shared tensor representation for cross-domain diagnosis. Following the adversarial learning idea, Li et al. [14] mapped the knowledge from target to source working condition based on generative adversarial net. Jiao et al. [15] utilized maximum classifier discrepancy to gain class-separable and domain-invariant features. Jointly using distribution alignment and adversarial learning, Li et al. [16] combined correlation alignment (CORAL) and a gradient reversal layer (GRL) where Jiao et al. [17] introduced joint MMD (JMMD) to adversarial training. Considering statistical metric, adversarial training and maximum classifier discrepancy method [18], Lee et al. proposed a asymmetric inter-intra domain alignments approach [19]. Moreover, the meta-learning and disentangle learning [20] are also introduced in domain adaptation to boost the generalization ability. Combining meta-learning and domain adaptation, Feng et al. [21] utilized similarity-based prototypical networks to improve identification performance. Wu et al. [22] progressively disentangle the domain-invariant and domain-specific features by feature decomposition, feature separation and feature reconstruction.

In a real industrial application, however, the machines often run continuously and faulty data are commonly collected from different domains. Thus, in a conventional DA, the fault samples collected from specific domains are insufficient for feature extraction to ensure domain invariance. Hence, multi-source DA methods are proposed to fully leverage faulty data from different domains and exploit the domain-invariant features that are robust representations across varying working conditions [23,24]. Huang et al. [23] fused multi-source information and fault label information with a modified DenseNet. Li et al. [24]

developed a multiple DA with weakly supervised data from the target domain using different but related machines to enhance the diagnostic performance. Xia et al. [25] used a moment matching-based metric to reduce the discrepancy among all source domains and a target domain for fault identification, promoting the model's reliability.

Nevertheless, the multi-source DA methods still have several limitations. The above approaches work only if the distribution of target domain data is accessible during the training phase, which is unfortunately practically impossible since conducting the target data collection or manual labeling is time-consuming and tedious. The diagnostic performance based on the above approaches would inevitably degrade. Therefore, it is difficult but important to have a cross-domain diagnosis method for faults under unseen working conditions. In such a method, the faulty features across domains should be further excavated to learn the invariance of domains. A bridge could be designed for discrete domains rather than only considering the available domains as the DA methods did. This is the motivation of the study carried out in this paper.

1.3. Fault diagnosis with domain generalization

Limited work has attempted to solve this challenge through domain generalization (DG) diagnosis by normalization or metric learning [10,26]. The key point of the challenge is the model capability of generalization under unseen working conditions. The signal in the unseen working conditions may be out of the distribution in the seen domain, and data augmentation is a promising technology to enrich the data distribution. Zhuo et al. [27] developed a generative approach with auxiliary information to diagnose unseen faults. Li et al. [28] compared five different data augmentations for diagnosis, in which signal translation provided the most remarkable improvements. Pei et al. [29] utilized a Wasserstein auto-encoder with a meta-learning strategy to conduct data augmentation to address the issue of data limitation and imbalance problem. Zhang et al. [30] proposed a signal augmented semi-supervised learning scheme through a generative adversarial network for fault diagnosis toward the small sample problem. It is noted that the above augmentation only focuses on faults rather than domains. Due to the limited domain label, Matsuura et al. [31] utilized clustered pseudo label to train a domain-generalized model. Inspired by AdaIN in style transfer, Zhou et al. [32] captured multiple style information to learn mixed feature statistics, enhancing the generalizability of the trained model. The up-to-date research indicates that the existing data from different available domains may be a trigger to generate augmented data in a new domain. Hence, this work will generate fresh insight into a new idea called cross-domain augmentation (CDA) diagnosis to enhance the domain-invariant feature learning and to boost the generalization capability of the diagnosis model.

The novelty of the work can be further illustrated by comparing CDA diagnosis with existing literature methods, as shown in Table 1. For an industrial diagnosis task with varying work conditions and unseen target domain, the feasibility of different methods is summarized. The DG-based methods are more feasible for a real industrial diagnosis because they can further learn the domain-invariant features, and the CDA diagnosis can further enrich the data distribution even the target domain is unknown. In short, the proposed work can fully exploit the augmented domain technologies for diagnosis under unseen working conditions.

Table 1 Feasibility for industrial diagnosis of different methods.

Method	Feasibility for industrial scenarios			
	domain shift	multiple & variable domains	unseen faults & unseen domains	Domain augmentation
ERM	×	×	×	×
DA	√	×	×	×
Multiple DA	√	√	×	×
DG	√	√	√	×
CDA	√	√	√	√

1.4. Motivation and contribution

In a nutshell, to alleviate the existing drawbacks of DA-based diagnosis and to fulfill a more robust and reliable diagnosis, this work proposes the idea of CDA diagnosis and develops one of its potential implements, i.e., adversarial domain-augmented generalization (ADAG), for fault diagnosis of rotating machinery under unseen working conditions. In this method, three modules are integrated to fulfill multi-source feature learning. A feature extractor and a fault classifier are simultaneously trained with a domain classifier in an adversarial way to achieve fault identification from available source domains to unseen domains. Moreover, the CDA diagnosis is carried out at the instance and feature levels to boost the generalization capability. In the latent space, the augmented domain is exploited to construct a convex hull, which may generate proximal data to bridge the discrete domains. Enhanced by the CDA, adversarial training on multi-domain can learn smoother features with the domain invariance. To evaluate our method, elaborately designed experiments based on two well-known bearing vibration platforms under variable and unseen working conditions are fully explored.

The contributions of this work can be summarized as follows:

(1) Beyond basic DA diagnosis, a new idea called CDA diagnosis is first introduced for fault diagnosis under unseen working conditions, which leverages available domains to build an augmented domain.

(2) A detailed implementation for CDA diagnosis, i.e. ADAG, is developed. ADAG can generate an augmented domain by convex combination of the signal and its labels from different domains, which expands a continuous latent space.

(3) Guided by multi-level CDA, the augmented feature space derived by adversarial training can eliminate the domain shift across different source domains to achieve a robust diagnosis system.

The remainder of this paper is structured as follows. Section 2 introduces the preliminaries and the main idea of CDA diagnosis. Section 3 develops the ADAG model to strengthen the idea of CDA diagnosis. Section 4 provides an in-depth discussion on different case studies. Section 5 summarizes the work.

For ease of navigation through the manuscript, all the notations used in the paper are summarized as below.

Notation	Description
x_i	Signal sample

y_i	Signal label
$P(\cdot,\cdot)$	Joint distribution
$\mathcal{L}$	Loss function
$\mathcal{X}$	Signal space
$\mathcal{F}$	Feature space
$\mathcal{Y}$	Label space
f_θ	Prediction model
$\mathcal{R}$	Risk
$\mathcal{R}_S$	Source risk
$\mathcal{R}_T$	Target risk
$d_{\mathcal{H}\Delta\mathcal{H}}(\cdot,\cdot)$	$\mathcal{H}\Delta\mathcal{H}$ distance
$\boldsymbol{\mathcal{D}}_S$	Seen domain
$\boldsymbol{\mathcal{D}}_U$	Unseen domain
N_S	Numbers of seen domain
N_U	Numbers of unseen domain
γ	Upper-bound of $\mathcal{H}\Delta\mathcal{H}$ distance between augmented and unseen domains
ε	Upper-bound of $\mathcal{H}\Delta\mathcal{H}$ distance between augmented and seen domains
$\boldsymbol{D}_S$	Dataset of seen domain
$\boldsymbol{D}_A$	Dataset of augmented domain
$Dir(\cdot)$	Dirichlet distribution
R	Gradient reversal layer
E	Feature extractor
C	Fault classifier
D	Domain classifier
$\mathcal{L}_{CP}$	P-level loss of C-branch
$\mathcal{L}_{DP}$	P-level loss of D-branch
$\mathcal{L}_{CI}$	I-level loss of C-branch
$\mathcal{L}_{DI}$	I-level loss of D-branch
$\mathcal{L}_{CF}$	F-level loss of C-branch
$\mathcal{L}_{DF}$	F-level loss of D-branch

2. Preliminaries and ideas

2.1. Intelligent fault diagnosis from ERM to DG

The basic idea of intelligent fault diagnosis based on traditional machine learning follows the ERM principle [10,11]. Given the training set $D = \{X, Y\} = \{(x_i, y_i)\}_{i=1}^n$ sampled from the distribution $P(X,Y)$ and specific loss function $\mathcal{L}: \mathcal{Y} \times \mathcal{Y} \to [0, \infty)$ to optimize the prediction model $f_\theta: \mathcal{X} \to \mathcal{Y}$ with the parameter θ. $\mathcal{X}$ is the signal space, and $\mathcal{Y}$ represents the label space. Hence, we want to learn the optimal parameters θ^* by minimizing the ERM.

$$\mathcal{R} = \frac{1}{n}\sum_{i=1}^{n}[\mathcal{L}(f_\theta(x_i), y_i)] \quad (1)$$

In the DA methods, the concept of domain $\mathcal{D} = \{\mathcal{X}, P(X)\}$ is defined where $P(X)$ is the margin distribution of data, and $X = \{x_1, \ldots, x_n\} \in \mathcal{X}$. To alleviate the issue of domain shift,

DA aims to learn a shared feature space $\mathcal{F}_{S,T}$ between source domain $\mathcal{D}_S$ and target domain $\mathcal{D}_T$ to build a reliable prediction model f. The risk bound to design the loss function is given as [9]:

$$\mathcal{R}_T \le \mathcal{R}_S + \frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_T, \mathcal{D}_S) + \lambda \tag{2}$$

where $d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_T, \mathcal{D}_S)$ is the $\mathcal{H}\Delta\mathcal{H}$ distance that indicates the discrepancy between $\mathcal{D}_S$ and $\mathcal{D}_T$ and λ is a constant denoting as the minimal risk of measuring adaptability between two domains. From the perspective of the risk bound, $\mathcal{R}_S$ could be optimized by traditional ERM, and $d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_T, \mathcal{D}_S)$ could be approximated by MMD [12] or adversarial learning. The MMD metric can be unbiased empirically estimated as [33].

$$MMD\left(\mathcal{D}_j, \mathcal{D}_k\right)^2 = \left|\left| \frac{1}{n_j} \sum_{i=1}^{n_j} \phi\left(x_{i,j}\right) - \frac{1}{n_k} \sum_{i=1}^{n_k} \phi\left(x_{i,k}\right) \right|\right|_{\mathcal{H}}^2 \tag{3}$$

where adversarial learning can be achieved in a min-max game [34].

$$\min_G \max_D V\,(D, G) = E_{x \sim p_{\text{data}}\,(x)}[\log D\,(x)] + E_{Z \sim p_z(z)}[\log(1 - D(G(z)))] \tag{4}$$

In the multiple DA methods, by leveraging the multi-source domains, the risk bound of the target domain can be further developed as:

$$\mathcal{R}_T \le \sum_{j=1}^{N_S} \alpha_j \left(\mathcal{R}_S^j + \frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}\left[\mathcal{D}_T, \mathcal{D}_S^j\right] \right) + \lambda \tag{5}$$

where N_S is the number of domains. The intuitive solution is adding a loss term to reduce the discrepancy between source domains and target domains, such as multiple MMDs [33]:

$$MMD\left(\{\mathcal{D}_S^j\}_{j=1}^{N_S}, \mathcal{D}_T\right) = \frac{1}{N_S} \sum_{1 \le j \le N_S} MMD\left(\mathcal{D}_S^j, \mathcal{D}_T\right) \tag{6}$$

In the multiple DA, however, it is assumed that $\mathcal{D}_T$ is accessible to design an appropriate loss function to guide the model training. This assumption may be violated in many industrial scenarios. Hence, this paper introduces the DG into the fault diagnosis field to tackle the problem. The idea of DG for diagnosis is shown in Fig. 1.

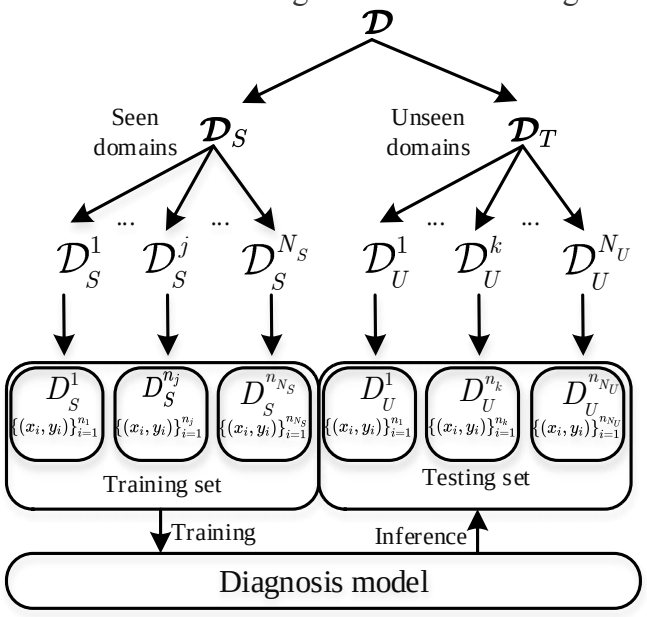

Fig. 1. Domain generalization for diagnosis.

For a domain $\boldsymbol{\mathcal{D}} = \{\boldsymbol{X}, \boldsymbol{P}(\boldsymbol{X})\}$, several seen and unseen domains can be sampled as $\boldsymbol{\mathcal{D}} = \{\boldsymbol{\mathcal{D}}_S; \boldsymbol{\mathcal{D}}_U\} = \{\mathcal{D}_S^1, \ldots, \mathcal{D}_S^j \ldots, \mathcal{D}_S^{N_S}; \mathcal{D}_U^1, \ldots, \mathcal{D}_U^k \ldots, \mathcal{D}_U^{N_U}\}$ according to different working conditions, where N_S and N_U are the numbers of seen and unseen domains. The goal of DG

is to learn a generalizable model from the source domains $\boldsymbol{\mathcal{D}}_S$ to diagnose faults to achieve a minimum error on unseen domains $\boldsymbol{\mathcal{D}}_U$.

$$\min_f \mathcal{E}_{(x,y)\in\boldsymbol{\mathcal{D}}_U}\left[\mathcal{L}(f_\theta(x),y)\right] \tag{7}$$

where $\{D_S^j\}_{j=1}^{N_S}=\left\{\left\{\left(x_{i,j},y_{i,j},d_{i,j}\right)\right\}_{i=1}^{n_j}\right\}_{j=1}^{N_S}$ is the training set with instance $x_{i,j}$, label $y_{i,j}$ and working condition label $d_{i,j}$. $\{D_U^k\}_{k=1}^{N_U}=\left\{\left\{\left(x_{i,k},y_{i,k},d_{i,k}\right)\right\}_{i=1}^{n_k}\right\}_{k=1}^{N_U}$ is the testing set which has no contribution to the model training.

An intuitive diagram to illustrate intelligent diagnosis from ERM to DG is shown in Fig. 2. ERM is the most common setting in which training and testing data follow the i.i.d assumption. DA aims to break the limitation to narrow the discrepancy between source and target domains, where multiple DA attempts to fully utilize the multi-source domains to achieve better performance. Moreover, without access to the target data, DG diagnosis aims to learn generalized knowledge from seen domains to diagnose faults in unseen domains. To sum up, the diagnosis model maybe fragile and easy to fail if only considering the ERM or multiple DA technologies in an engineering scenario. Therefore, the correlation and discrepancy between seen and unseen domains should be further explored in the DG diagnosis setting.

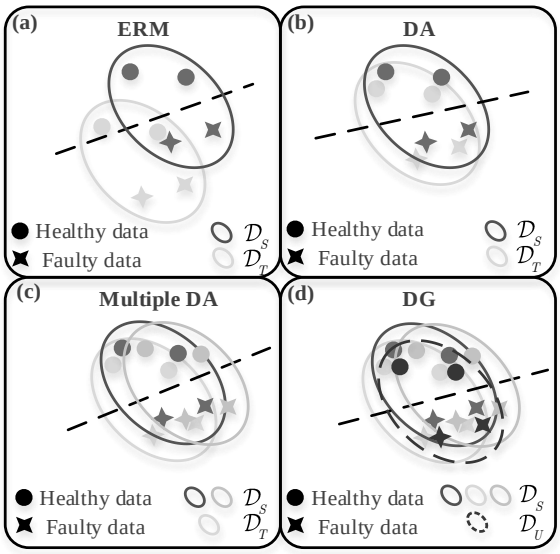

Fig. 2. Intelligent fault diagnosis methods: (a) ERM, (b) DA, (c) multiple DA, (d) DG.

2.2. Data Augmentation based on Mixup.

To boost the generalization of the diagnosis model, Mixup [36] can be a simple but remarkable technique. In essence, this technique enlarges the training set $D=\{(x_i,y_i)\}_{i=1}^{n}$through linear combination:

$$\tilde{x}_i=\alpha x_i+(1-\alpha)x_{i2} \tag{8}$$

$$\tilde{y}_i=\alpha y_i+(1-\alpha)y_{i2} \tag{9}$$

where (x_{i2},y_{i2}) are instances randomly drawn from D, and $\alpha\sim Beta(\varphi,\varphi),\varphi\in(0,\infty)$. Intuitively, through this linear interpolation, Mixup constructs additional and proximal instances and enables ERM to build a prediction function with a more robust decision boundary:

$$\mathcal{R} = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}(f_\theta(\tilde{x}_i), \tilde{y}_i) \tag{10}$$

Notably, Mixup only focuses on the sample from the same domain. In this paper, we introduce the Mixup into the DG framework to enhance its ability to learn domain-invariant features under different working conditions.

3. ADAG method based on CDA

3.1. The proposal of CDA idea for diagnosis

In this subsection, we introduce a new idea called CDA to diagnose and design an implementation by a convex combination of data and features based on the DG and Mixup.

As shown in Eq.(5), a DG task without access to $\mathcal{D}_T$ is challenging. However, it is still feasible because DG assumes that the samples in the unseen domain can be embedded into the proximal space of source features so that the fault classifier can identify the features in unseen domains. Accordingly, we introduce the idea of CDA to build an augmented domain from available source domains through the operator $\sigma(\cdot)$, which can be an explicit kernel or an implicit neural network.

$$\widetilde{\boldsymbol{\mathcal{D}}}_S \coloneqq \left\{ \sigma(\mathcal{D}_S^j) \mid \mathcal{D}_S^j \in \boldsymbol{\mathcal{D}} \right\} \tag{11}$$

Following this idea, we build a convex hull in the domain space through a convex combination of seen domains:

$$\widetilde{\boldsymbol{\mathcal{D}}}_S \coloneqq \left\{ \sum_{j=1}^{N_S} \alpha_j \mathcal{D}_S^j \mid \mathcal{D}_S^j \in \boldsymbol{\mathcal{D}}, \sum_{j=1}^{N_S} \alpha_j = 1, t_j \in [0,1] \right\} \tag{12}$$

The augmented domain with domain labels yields a more continuous distribution. Then, a special augmented domain can be given [37]:

$$\widetilde{\mathcal{D}}_U^k = \sum_{j=1}^{N_S} \alpha_{j,k} \mathcal{D}_S^j = \underset{\alpha}{argmin}\, d_{\mathcal{H}\Delta\mathcal{H}}\{\mathcal{D}_U^k, \sum_{j=1}^{N_S} \alpha_{j,k} \mathcal{D}_S^j\} \tag{13}$$

It means that the augmented domain can be a proximal domain to the unseen domain with the lowest discrepancy. The greedy optimization algorithm can iteratively narrow the gap between $\widetilde{\mathcal{D}}_U^k$ and $\widetilde{\boldsymbol{\mathcal{D}}}_S$. Let $\gamma = d_{\mathcal{H}\Delta\mathcal{H}}\{\widetilde{\mathcal{D}}_U^k, \mathcal{D}_U^k\}$ denote the upper boundary of the $\mathcal{H}\Delta\mathcal{H}$ discrepancy between the augmented and unseen domains, and $\varepsilon = max\{d_{\mathcal{H}\Delta\mathcal{H}}\{\widetilde{\mathcal{D}}_U^k, \mathcal{D}_S^j\}\}$ denotes the upper-bound among the augmented and available source domains. The upper-bound of Eq.(5) could be updated as follows:

$$\mathcal{R}_U^k \le \sum_{i=1}^{N_S} \alpha_{j,k} R_S^j + \frac{\gamma+\varepsilon}{2} + \lambda \tag{14}$$

Although $\mathcal{D}_U^k$ is inaccessible, we can optimize the model iteratively by using samples drawn from the convex hull with its labels to minimize the discrepancy term $\gamma + \varepsilon$. For better understanding, Fig. 3 illustrates the effect of the feature combination from a seen to an unseen domain. Fig. 3 (a) shows the vanilla DG, and Fig. 3 (b) shows the convex hull built by the convex combination. In this manner, the unseen faulty data may be covered in the convex hull. As shown in Fig. 3 (c), the trained model can learn more generalized features and more robust decision boundaries.

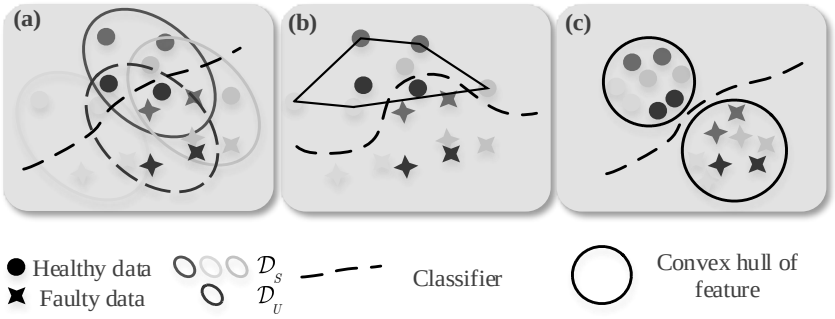

Fig. 3. Effect of CDA using convex combination: (a) vanilla DG without CDA, (b) the convex hull of the seen faulty data, and (c) generalized features and robust decision boundary after training.

3.2. Overview of ADAG method

Fig. 4 shows the proposed method called ADAG under the guideline of CDA idea for diagnosis.

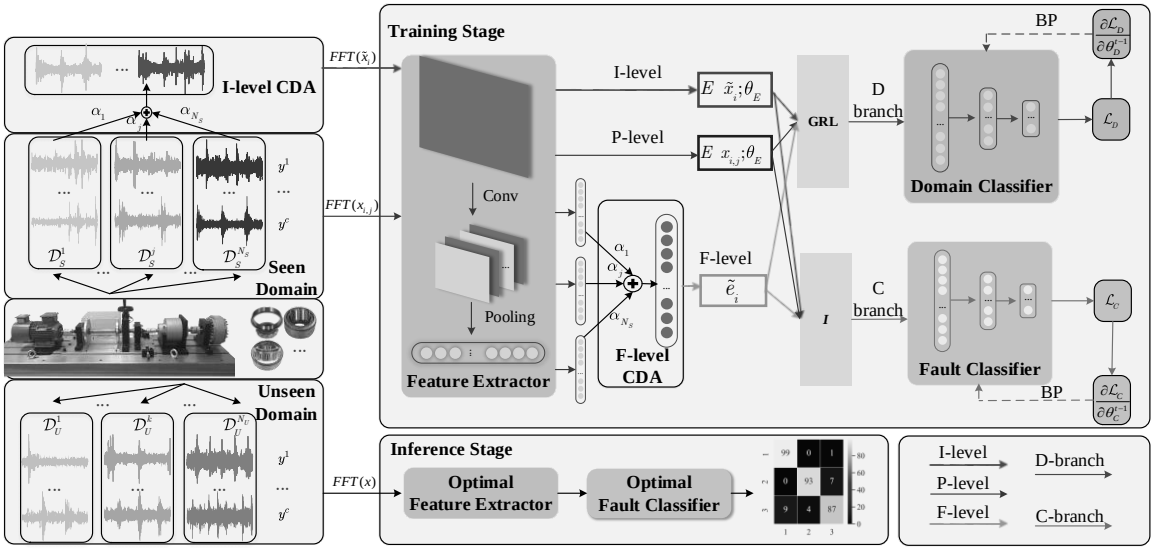

Fig. 4. Framework of Adversarial Domain-Augmented Generalization.

First, data acquisition and data preprocessing are conducted. Considering that the machine runs under variable working conditions, the vibration signal has variable data distribution. The accessible data are collected to build a seen dataset $\boldsymbol{D}_S = \{D_S^j\}_{j=1}^{Ns}$, which involves the training and validation datasets. An augmented dataset can be constructed as follows:

$$D_A = (\tilde{x}_i, \tilde{y}_i, \tilde{d}_i) = (\sum_{j=1}^{Ns} \alpha_j x_{i,j}, \sum_{j=1}^{Ns} \alpha_j y_{i,j}, \sum_{j=1}^{Ns} \alpha_j d_{i,j}) \tag{15}$$

Specifically, the augmented data with faulty and domain labels are generated through the convex combination of the seen data from different domains, where α_j is sampled from the Dirichlet distribution with the gamma function $\Gamma(\cdot)$:

$$Dir(\alpha|\beta) = \frac{\Gamma(\beta_0)}{\Gamma(\beta_1)\cdots\Gamma(\beta_{N_S})} \prod_{j=1}^{N_s} \alpha_j^{\beta_k - 1} \tag{16}$$

where $\beta_0 = \sum_{j=1}^{N_S} \beta_j$. Intuitively, β_j is set equally because the inaccessible target domains and the Dirichlet distribution have a special function that enables $\sum_{j=1}^{N_S} \alpha_j = 1$. Fig. 5 shows its distribution in three dimensions.

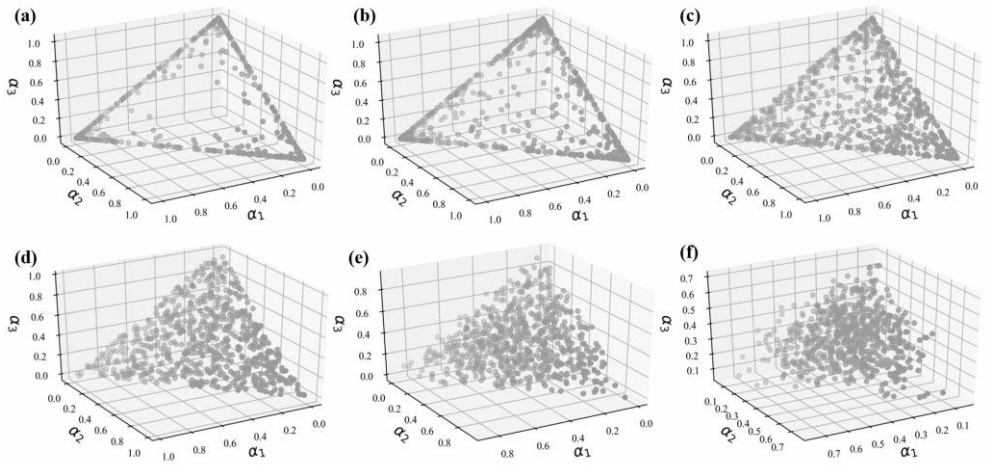

Fig. 5. Dirichlet distribution with different values of β_j: (a) 0.1, (b) 0.2, (c) 0.5, (d) 1.0, (e) 2.0, and (f) 5.0.

As shown in Fig. 5, the points are located by the three-dimensional coordinate $(\alpha_1, \alpha_2, \alpha_3)$. The points in panel (d) are evenly distributed in the triangle plane whose vertices are (1,0,0), (0,1,0), and (0,0,1). In addition, with the growth of β_j, the points move from the edge to the center.

The augmented dataset with the original signal uses Fast Fourier Transform (FFT) and the reshape operation to build two-dimensional (2-D) instance for model training. The prediction model derived in the model training phase can be divided into two parts as $f = E \circ C$, where feature extractor (E) is denoted as a mapping from data to features $E: \mathcal{X} \to \mathcal{F}$ and fault classifier (C) is denoted as $C: \mathcal{F} \to \mathcal{Y}$. In particular, a neural network is preferable to perform both E and C, due to its feasibility and capability to learn nonlinear transformation.

In practice, adversarial learning generally constructs a double-branched architecture for fault classification and domain classification, respectively. Similar to $f = E \circ C$, the domain branch is marked as $h = E \circ R \circ D$, where the domain classifier is denoted as D and R in the GRL [24,35]. The GRL could be formulated as:

$$R(x) = x; \quad dR(x)/dx = -I \tag{17}$$

where the forward transformation is identical and the behavior of the back propagation reverses the sign.

The model structure of E, C, and D is a vanilla CNN and fully connected network [38]. The model learns the mapping between the original instance and supervised label to minimize the risk of unseen domain by the loss function below.

3.3. Loss function of ADAG with multi-level CDA

(1) Prototype-level

The prototype-level (P-level) has no CDA data engaged in the loss function, and it obeys the conventional fashion of model training.

After module initialization, ADAG performs forward propagation of seen data. The pre-processed seen data include the source and augmented datasets. The source data is transformed into E with its parameters θ_E as $E\left(x_{i,j};\theta_E\right)$. In order to learn the shared representations that are domain unrelated but fault related, the fault classifier and domain classifier can predict the domain label and fault label with their parameters as $D\left(R\left(E\left(x_{i,j};\theta_E\right)\right);\theta_D\right)$ and $C\left(E\left(x_{i,j};\theta_E\right);\theta_C\right)$. In Fig. 4, the block I is the identical transformation and the block *GRL* is the gradient reversal layer. For clarity, D-branch is named for the loss acting on the D and E, while C-branch is named for the loss acting on the C and E. The loss function of D-branch and C-branch can be derived as follows:

$$\mathcal{L}_{DP} = \frac{1}{N_S \times n_j}\sum_{j=1}^{N_S}\sum_{i=1}^{n_j}\left[\mathcal{L}\left(D\left(R\left(E\left(x_{i,j};\theta_E\right)\right);\theta_D\right),d_{i,j}\right)\right] \tag{18}$$

$$\mathcal{L}_{CP} = \frac{1}{N_S \times n_j}\sum_{j=1}^{N_S}\sum_{i=1}^{n_j}\left[\mathcal{L}\left(C\left(E\left(x_{i,j};\theta_E\right);\theta_C\right),y_{i,j}\right)\right] \tag{19}$$

where $\mathcal{L}(\cdot,\cdot)$ means cross-entropy loss with softmax of one instance:

$$\mathcal{L}(\hat{y},y) = -\sum_{m=1}^{M}\hat{y}_m \ln y_m = -\sum_{m=1}^{M}\frac{\exp(o_m)}{\sum_{l=1}^{M}\exp(o_l)}\ln y_m \tag{20}$$

The o denotes the output nodes of classifier and y_m denotes the true labels.

(2) Instance-level

The Instance-level (I-level) CDA involves the I-level loss, which leverages the pre-processing signal to build a convex hull through a convex combination of seen domains as given in Eq. (15). In the augmented domain, $\tilde{x}_i, \tilde{y}_i, \tilde{d}_i$ denote augmented instance, augmented faulty label and augmented domain label, correspondingly.

Benefiting from the augmented dataset D_A at the I-level, the features extracted by E are marked as $E(\tilde{x}_i;\theta_E)$. The fault classifier and domain classifier can predict the augmented domain label and augmented fault label with their parameters as $D\left(R\left(E(\tilde{x}_i;\theta_E)\right);\theta_D\right)$ and $C(E(\tilde{x}_i;\theta_E);\theta_C)$. Then we can compute the loss between the I-level augmented data and its augmented domain label as follows:

$$\mathcal{L}_{DI} = \frac{1}{n_j}\sum_{i=1}^{n_j}\left[\mathcal{L}\left(D\left(R\left(E(\tilde{x}_i;\theta_E)\right);\theta_D\right),\tilde{d}_i\right)\right] \tag{21}$$

The loss of fault classifier can be computed as:

$$\mathcal{L}_{CI} = \frac{1}{n_j}\sum_{i=1}^{n_j}[\mathcal{L}(C(E(\tilde{x}_i;\theta_E);\theta_C),\tilde{y}_i)] \tag{22}$$

(3) Feature-level

The Feature-level (F-level) CDA involves the F-level loss, which leverages the learned features $E\left(x_{i,j};\theta_E\right)$ of P-level from different domains. The features at P-level build a convex hull through a convex combination of the seen domains features. Specifically, the augmented feature and its label can be linearly combined to explore the robust representation in the feature space as:

$$(\tilde{e}_i,\tilde{y}_i,\tilde{d}_i) = \left(\sum_{j=1}^{N_S}\alpha_j E\left(x_{i,j};\theta_E\right),\sum_{j=1}^{N_S}\alpha_j y_{i,j},\sum_{j=1}^{N_S}\alpha_j d_{i,j}\right) \tag{23}$$

Thus, the fault classifier and domain classifier can predict the augmented domain labels and augmented fault labels with their parameters as $D(R(\tilde{e}_i);\theta_D)$ and $C(\tilde{e}_i;\theta_C)$. Therefore, the relevant loss function is formulated as follows:

$$\mathcal{L}_{DF} = \frac{1}{n_j}\sum_{i=1}^{n_j}\left[\mathcal{L}\left(D(R(\tilde{e}_i);\theta_D),\tilde{d}_i\right)\right] \tag{24}$$

$$\mathcal{L}_{CF} = \frac{1}{n_j}\sum_{i=1}^{n_j}[\mathcal{L}(C(\tilde{e}_i;\theta_C),\tilde{y}_i)] \tag{25}$$

In summary, there are six losses contributing in two branches on three transformation levels. The total loss function for D-branch can be an accumulation of the three terms of D-branch as:

$$\mathcal{L}_D = \mathcal{L}_{DP} + \mathcal{L}_{DI} + \mathcal{L}_{DF} \tag{26}$$

Symmetrically, based on the features from the three levels, the loss function to train the fault classifier can be formulated as follows:

$$\mathcal{L}_C = \mathcal{L}_{CP} + \mathcal{L}_{CI} + \mathcal{L}_{CF} \tag{27}$$

Consequently, the total loss function of ADAG can be formulated as follows:

$$\mathcal{L}_O = \alpha_C\mathcal{L}_C + \alpha_D\mathcal{L}_D \tag{28}$$

where α_C and α_D are the trade-off parameters. Table 2 presents a summary of the loss functions.

Table 2 Six loss functions of ADAG on two branches at three levels.

	P-level	I-level	F-level
C-branch	$\mathcal{L}_{CP}$	$\mathcal{L}_{CI}$	$\mathcal{L}_{CF}$
D-branch	$\mathcal{L}_{DP}$	$\mathcal{L}_{DI}$	$\mathcal{L}_{DF}$

The augmentation of the *I*-level and *F*-level benefits from the convex combination of the data-label pairs and feature-label pairs, respectively. Theoretically, the above augmentation can enhance the performance of feature extraction for cross-domain diagnosis.

3.4. Optimization

Gradient back propagation (BP) through stochastic gradient descent is used to optimize the model. In each iteration, the parameters of neural networks are trained to satisfy the following optimization constrains:

$$\widehat{\theta_E} = arg\left\{\min_{\theta_E}\alpha_C\mathcal{L}_C\left(\theta_E,\widehat{\theta_C}\right),\max_{\theta_E}\alpha_D\mathcal{L}_D\left(\theta_E,\widehat{\theta_D}\right)\right\} \tag{29}$$

$$\widehat{\theta_C} = \underset{\theta_C}{argmin}\ \alpha_C\mathcal{L}_C\left(\widehat{\theta_E},\theta_C\right) \tag{30}$$

$$\widehat{\theta_D} = \underset{\theta_D}{argmin}\ \alpha_D\mathcal{L}_D\left(\widehat{\theta_E},\theta_D\right) \tag{31}$$

In this way, the gradients of $\mathcal{L}_D$ are minimized with θ_D but maximized with θ_E by GRL, where $\mathcal{L}_C$ is minimized with θ_E and θ_C. The fault classifier instructs the feature extractor to learn fault-related features. Simultaneously, the domain classifier instructs the feature extractor to learn domain-invariant features through adversarial training. The details of the parameter updating using the stochastic gradient descent are formulated as follows:

$$\theta_E^t \leftarrow \theta_E^{t-1} - \gamma\left(\alpha_C\frac{\partial\mathcal{L}_C}{\partial\theta_E^{t-1}} - \alpha_D\frac{\partial\mathcal{L}_D}{\partial\theta_E^{t-1}}\right) \tag{32}$$

$$\theta_D^t \leftarrow \theta_D^{t-1} - \gamma\,\alpha_D\frac{\partial\mathcal{L}_D}{\partial\theta_D^{t-1}} \tag{33}$$

$$\theta_C^t \leftarrow \theta_C^{t-1} - \gamma\,\alpha_C\frac{\partial\mathcal{L}_C}{\partial\theta_C^{t-1}} \tag{34}$$

where γ is a learning rate in the optimization algorithm.

After model training, we can obtain the optimal parameters θ_E^* and θ_C^* to establish a generalized diagnosis model, which can be used to diagnose the fault in the unseen domain with an unseen dataset $\boldsymbol{D}_U = \{D_U^k\}_{k=1}^{N_U}$.

In brief, the ADAG method can be summarized as Algorithm 1.

Algorithm 1 : ADAG

 # Training stage

Input: Multiple seen dataset $\left\{\left\{\left(x_{i,j}, y_{i,j}, d_{i,j}\right)\right\}_{i=1}^{n_j}\right\}_{j=1}^{N_S}$ from seen domains.

Initialization: *E, C, D* with initialized parameters and other pre-setting hyper-parameters.

1: **for** *epoch* = 1 to *epochs* **do**
2: Randomly sample seen data from the dataset.
3: Generate augmented dataset by Eq.(15).
4: Forward propagation to generate augmented features by Eq. (23).
5: Forward propagation to calculate total loss function Eq.(28).
6: Backward propagation to update *E, C, D* by Eqs. (32) to (34).
7: **end for**

Return: The optimal *E, C, D*.

 # Testing stage

Input: Unseen dataset $\boldsymbol{D}_U$.

Model: ADAG with optimal *E, C, D*.

Output: Diagnosis result of $\boldsymbol{D}_U$ by optimal *E* and *C*.

4. Experimental validation

4.1. Experiment setup

The reliable condition monitoring of rolling bearings is highly demanded for rotating machines because the state of rolling bearings directly determines the health and remaining lifetime of rotating machines [39]. Therefore, this study fully exploits the augmented domain technologies for diagnosis under unseen working conditions and two case studies are setup. The two experiments are based on the test rig at SDUST [40,41] and Ottawa [42] for bearing fault diagnosis.

To evaluate the diagnosis model under unseen conditions, we have designed two progressive experimental case studies. In the first case, we consider constant domain shift in the unseen domain where the experiment rig operated with multiple conditions of constant speed. In the second case, we consider a more challenging situation where the domain shift is variable, i.e. the experiment rig is operated under unseen and variable speeds. In this way, we can verify that the ADAG method based on CDA is fully applicable to complex industrial applications.

4.1.1 SDUST dataset

Fig. 6 shows the experimental platform of SDUST, which includes a motor, a shaft coupling, a rotor, a testing bearing, a gearbox, and a break.

The bearing type is N205EU. The sampled data include four health conditions, namely, normal (Nor), inner ring fault (I), rolling element fault (B), and outer ring fault (O). Each fault type that includes three sizes of 0.2, 0.4, and 0.6mm. Four different working conditions are set under different speeds of 1000, 1500, 2000, and 2500r/min. Accordingly, four domain data are collected by the vibration sensor, and each domain has data with 10 different types of health conditions (abbreviated as Nor, I02, I04, I06, B02, B04, B06, O02, O04, and O06, respectively). The data in each health condition have 100 samples. Each

sample has 2048 data points that fully cover the fault information.

The sensors must maintain a minimum level of sampling rate in order to accurately detect impact signals produced by the rolling element in the bearing striking a local fault in the inner or outer race during operation, e.g., the ball pass frequency of inner race (BPFI) can be formulated as:

$$BPFI = \frac{nf_s}{2}\left\{1 + \frac{d}{D}cos\,\alpha\right\} \tag{35}$$

where D, d, α, n and f_s are the pitch diameter, ball diameter, contact angle between the ball and the cage, number of rolling elements and rotating frequency of bearing.

For SDUST dataset, the fault characteristic frequency is about 8.375 f_s according to Eq. (35). Since the rotating frequency of the test shaft is 8.3Hz~42Hz, the fault characteristic frequency range could be 69.51Hz~347.6Hz, and the lower limit of the sampling rate is about 695Hz. Therefore, the sampling rate 25.6kHz in the experiment is enough to identify the corresponding faults.

The experiment sets four generalization tasks for four domains: T1000, T1500, T2000, and T2500 as shown in Table 2. For instance, the T1500 task means that the model is trained from seen domains under different speeds of 1000, 2000, and 2500r/min, thereby generalizing the knowledge to the unseen target domain under the speed of 1500 r/min.

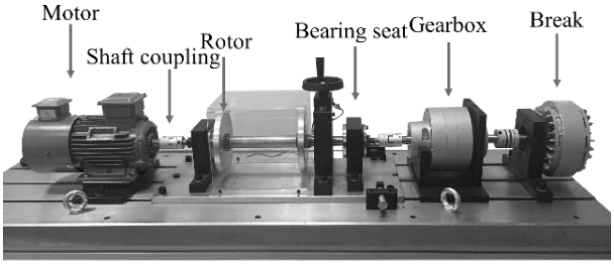

Fig. 6. Experimental platform of SDUST dataset.

Table 3 Generalization tasks of SDUST dataset.

Generalization tasks	Seen domains	Unseen domain
T1000	1500r/min, 2000r/min, 2500r/min	1000r/min
T1500	1000r/min, 2000r/min, 2500r/min	1500r/min
T2000	1000r/min, 1500r/min, 2500r/min	2000r/min
T2500	1000r/min, 1500r/min, 2000r/min	2500r/min

In summary, we construct a seen dataset $\left\{\left\{\left(x_{i,j}, y_{i,j}, d_{i,j}\right)\right\}_{i=1}^{n_j}\right\}_{j=1}^{N_S}$, where $N_S = 3$ and $n_j = 1000$ in each working condition with 10 different health conditions $y_{i,j}$. Similarly, the validation dataset is also constructed by seen domains to perform model selection, and the data in the validation dataset have no overlap with the training set. Finally, the sample number in the unseen domain is 2000 for testing.

4.1.2 Ottawa dataset

Further in-depth researches were carried out with another well-known Ottawa dataset under time-varying conditions. The experiment was conducted to collect vibration signal on a mechanical-failure simulator (MFS-PK5M) with 200kHz sampling rate. According to Eq. (35), the fault characteristic frequency is about 5.43 f_s. Since the rotating frequency of bearing is 14Hz~30Hz, the fault characteristic frequency range could be 76Hz~162.9Hz,

and the lower limit of the sampling rate is about 325.8Hz. Therefore, the sampling rate 200kHz in the experiment is enough to identify the corresponding faults. Each sample has 8192 data points that fully cover the fault information.

Fig. 7 shows the time-varying speed fault diagnosis test bench of the Ottawa dataset [42]. The speeds of different working conditions named as SA to SD, are shown in Table 4. Similar to SDUST dataset, four generalization tasks exist for TA, TB, TC, and TD. For instance, the TA task means that the model is trained from seen domains under working condition of SB, SC, and SD, thereby generalizing the knowledge to the unseen target domain under working condition of SA.

In the Ottawa dataset, we construct a seen dataset $\left\{\left\{\left(x_{i,j}, y_{i,j}, d_{i,j}\right)\right\}_{i=1}^{n_j}\right\}_{j=1}^{N_S}$, where $N_S = 3$ and $n_j = 300$ under each working condition with three different health conditions, i.e., healthy, Inner ring fault and Outer ring fault.

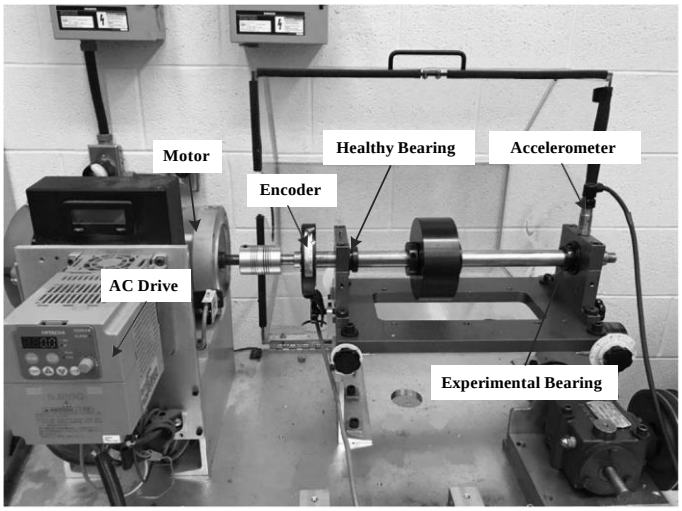

Fig. 7. Fault diagnosis test bench of the Ottawa dataset under time-varying speeds.

Table 4 Different speed conditions of the Ottawa dataset.

Bearing health condition	SA Speed increase	SB Speed decrease	SC Speed increase then decrease	SD Speed decrease then increase
Health	14.1-23.8	28.9-13.7	14.7 -25.3-21.0	24.2-14.8-20.6
Inner race fault	12.5-27.8	24.3-9.9	15.1-24.4-18.7	25.3-14.8-19.4
Outer race fault	14.8-27.1	24.9-9.8	14.0-21.7-14.5	26.0 -18.9-24.5

4.2. Compared methods

To evaluate the effectiveness of the CDA idea and the ADAG method, we define a set of compared methods with some typical or up-to-date technologies. As shown in Table 5, all the methods use the same pre-processing and network backbone for a fair comparison. The compared methods are divided into two series, i.e. M1-M6 series are competitive related methods while A, AD, AC, AF and AI are proposed ADAG and some variant studies.

Table 5 Compared methods.

Methods	Description

M1	ERM.
M2	ERM with MMD
M3	ERM with JMMD
M4	ERM with CORAL
M5	ADIG [10]
M6	IEDGNet [26]
A	The prototype of ADAG
AD	Remove $\mathcal{L}_{CI}, \mathcal{L}_{CF}$
AC	Remove $\mathcal{L}_{DI}, \mathcal{L}_{DF}$
AF	Remove $\mathcal{L}_{CI}, \mathcal{L}_{DI}$
AI	Remove $\mathcal{L}_{CF}, \mathcal{L}_{DF}$

M1 means the ERM method using multi-domain data based on the general cross-entropy loss. M2-M4 follow the same setting in [17,23,43] by adding a distance metric or distribution alignment as a loss term, such as MMD, JMMD, and CORAL. M5 [10] is a start-of-the-art method DG that uses adversarial training with normalization strategies and a strategy of multi-task training. M6 [26] is a competitive method for cross-domain diagnosis under unseen domain through triplet loss and data augmentation with Gaussian noise.

The second series shows the variety of the proposed method. A is a prototype of the proposed ADAG. AD removes $\mathcal{L}_{CI} + \mathcal{L}_{CF}$ to examine the augmentation of the *C*-branch, whereas AC removes $\mathcal{L}_{DI} + \mathcal{L}_{DF}$ to verify the *D*-branch augmentation. Similarly, AF removes $\mathcal{L}_{CI}$ and $\mathcal{L}_{DF}$ to examine the effectiveness of *I*-level augmentation, whereas AI removes $\mathcal{L}_{CF}$, $\mathcal{L}_{DF}$ to verify the effectiveness of *F*-level augmentation.

4.3. Hyper-parameter Settings

The hyper-parameter selection, part by referring to [10,44] and part by trial and error, is presented in Table 6 as the preferred hyper-parameters.

Table 6 Hyper-parameter setting.

Hyper-parameter	value	Hyper-parameter	value
Learning rate	0.0001	$\alpha_C; \alpha_D$	0.5; 1
Batch size	128	Weight decay	0.0001
$\{\beta_j\}_{j=1}^{N_S}$	0.9	Epoch	200

Particularly, we carried out a thorough study on the selection for parameter value of $\{\beta_j\}_{j=1}^{N_S}$ in the Dirichlet distribution. As shown in Fig. 8, the influence of different values for β on accuracy are plotted, where the overall accuracy is stable as the β varies between 0.8 to 1.2. However, this doesn't mean we can set a random value to β, and by referring to Fig. 5, the weight of augmented data can be evenly distributed in the triangle plane when β approaching 1.0. Hence, in our experiments $\{\beta_j\}_{j=1}^{N_S}$ is set to 0.9 as preferred.

Remark 1: the parameter $\{\boldsymbol{\beta_j}\}_{\boldsymbol{j=1}}^{\boldsymbol{N_S}}$ is a crucial parameter to be fine-tuned since it has a close relationship with the quality of the augmented data in the unseen domain. For a CDA task, it can be set equally for simplification in a Dirichlet distribution.

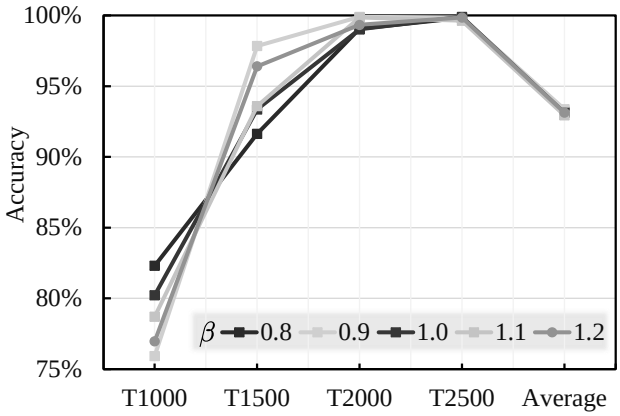

Fig. 8. Influence of different β values on the accuracy.

4.4. Result and discussion

To verify the power of the idea CDA and the superiority of the ADAG method, we carry out two experiments in this section to show the performance.

4.4.1 SDUST dataset

In this experiment, the diagnosis results of the unseen domain in each task are averaged by five trials to eliminate contingency, showing accuracy and relevant standard deviation. The highest value in each column is in bold, and the second-highest value is underlined. All the results are given in Table 7.

Table 7 Diagnosis results of SDUST dataset (%).

Method	T1000	T1500	T2000	T2500	Average
M1	69.73±1.43	94.65±2.72	96.28±2.47	83.19±3.89	85.96±2.63
M2	59.85±9.48	83.14±1.28	89.51±3.21	73.62±4.67	76.53±4.66
M3	65.80±3.05	94.95±2.90	96.28±1.98	84.98±6.26	85.50±3.55
M4	66.51±2.20	94.81±3.27	96.76±1.80	83.96±3.60	85.51±2.72
M5	71.87±2.87	93.58±3.12	98.26±1.14	90.91±3.22	88.66±2.59
M6	49.98±0.15	96.18±4.15	97.75±2.99	84.40±6.48	82.08±3.44
A	74.03±4.17	**96.60±2.64**	99.06±0.42	89.37±2.94	89.77±2.54
AD	73.75±1.91	95.19±1.52	98.35±0.84	90.42±0.76	89.43±1.26
AC	73.27±3.64	94.48±2.86	97.27±0.74	86.06±4.49	87.77±2.93
AF	74.09±2.40	96.24±2.51	**99.18±0.57**	**90.89±1.87**	**90.10±1.84**
AI	**74.80±2.54**	95.16±2.41	97.61±1.49	89.55±3.58	89.28±2.51

The following are some of our findings and highlighted remarks.

1) The methods based on ERM and distribution alignment are insufficient or useless for cross-domain diagnosis with unseen working conditions in target domain. As shown in Table 7, we take the results of M1 as a baseline. Compared with ERM, multiple MMD methods have no improvement toward the unseen domain without the target data. M2-M4, which uses metrics to reduce domain discrepancy, has no improvement. Although multiple DA following M2-M4 can achieve satisfactory diagnosis under the DA assumption, once the target domain data are inaccessible, the performance will deteriorate. We can conclude that minimizing the distribution discrepancy may have some effects, but it is far from sufficient for industrial application.

2) DG-based methods are fit for CDA tasks, where domain augmentation is important

and effective. Compared with M1, M5 based on DG diagnosis achieves significantly enhanced performance. Although M6 obtains an unusually small accuracy in T1000 due to the large and unseen domain for metric learning, M6 shows great improvement in T1500. Notably, for these methods, the data distribution under unseen condition is unknown, while ADAG with CDA enriches the data distribution, which further boosts the generalization ability. This finding can guide us to design the CDA based methods with extra techniques such as domain augmentation, metric learning, or normalization to learn generalized features by fully using the available data.

Remark 2: the augmented domain generated by CDA enables the model fitting not only the source data distribution but also the data under unseen working condition, which is a core concern to DG-based diagnosis, and this is one of the main findings and contributions of this paper.

3) We conduct a thorough study on the related variants of ADAG by ablation experiments. The results in Table 7 indicate that the components in ADAG, i.e. *C*-branch, *D*-branch and *I*-level, *F*-level, are effective since their performances are better than baseline. The comparison among A, AD, and AC shows that *D*- and *C*-branch augmentation can benefit learning generalized features. Likewise, the observations on A, AF, and AI show close results. Through feature-level CDA, AF has informative high-level representation to construct an augmented domain, whereas instance-level loss uses only low-level samples.

Remark 3: in the ADAG framework, *C*-branch and *D*-branch are equally important and are indispensable. This conclusion is consistent with the properties of adversarial learning. The *F*-level CDA, however, is more important than the *I*-level because the high-level features can be built to explore the robust representation in the feature space. This is a key to the innovation of CDA diagnosis as depicted in Fig. 4.

4) As shown in Table 7, the results of ADAG for CDA tasks of T1500 and T2000 achieve remarkable improvement. By specifically focusing on T1500 and T2000, as the CDA builds the convex hull of features, the fault data in the unseen domain can be covered in the feature sets. Hence, the augmented domain has some vicinal samples or features to approximate the data in the unseen domain.

Remark 4: The results indicate that data collection in a wide-ranging domain is preferred. However, the industrial in-site data sensing is unable to have quality-checked data, and a wide-ranging data sampling is impossible. The power of CDA diagnosis roots in the supplement with augmented domain to bridge the domain shifts.

To verify the above comment, we carried out new DG tasks as shown in Table 8 with re-collected data from speeds of 500 and 3000r/min to boost the result of T1000 and T2500. The results are shown in Table 9. With the new collected domain data, all comparison methods improve their performance. In T1000*, the ADAG outperforms other methods, where in T2500*, ADAG achieves relatively high performance. Comparison of the results in T1000 and T1000* reveals that more domain data can boost model generalization ability. With limited domain data, the ADAG can reduce dependence on domain data and still obtain the best performance, which shows the superiority of the proposed method.

Table 8 New DG tasks with re-collected domain data.

Generalization task	Seen domains	Unseen domain
T1000*	500r/min, 1500r/min, 2000r/min	1000r/min

T2500*	1500r/min, 2000r/min, 3000r/min	2500r/min

Table 9 Diagnosis results with re-collected domain data.

Method	T1000*	T2500*
M1	88.24±1.45	98.09±1.68
M2	83.50±5.63	93.22±4.87
M3	87.20±2.33	99.51±0.29
M4	88.97±1.36	99.49±0.55
M5	84.33±6.97	98.24±1.34
M6	87.74±4.93	97.52±3.22
A	92.16±2.17	98.42±1.82

4.4.2 Ottawa dataset.

In the experiment on the Ottawa dataset, we also conduct five trials to show the superiorities of our method and to outline some new findings.

1) ADAG is still effective for CDA diagnosis even in time-varying conditions**.** As shown in Table 10, under time-varying conditions, the distribution sharply changed, but ADAG can still achieve the best accuracy comparing with M1-M6. It shows that the concept of CDA with convex combination is feasible and effective, revealing the augmented domain benefit the model learning the unseen pattern. The results of M1-M4 show that the multiple DA method has no obvious effect on the generalization ability of the time-varying diagnosis tasks. Only reducing the domain discrepancy of the seen domain often has little effect on unseen domain owing to the larger domain discrepancy of time-varying signals. M5 and M6 get little improvement for the design of normalization and metric learning.

2) Be cautious of instance-level loss. We further find that only using I-level loss may degenerate the performance, a similar conclusion as indicated above in Remark 3. When domain is dramatic shift such as time-varying condition, because the linear combination of the samples may confuse the model to learn the nonlinear feature, which is weak to adapt this situation. Comparison of A, AF and AI reveals that the standard deviation of variant AI is large, indicating that AI is unstable. Given the large difference in instance-level distribution, the augmentation effect at the instance level does not perform well, so constructing a feature convex hull in the feature space is more effective.

Table 10 Diagnosis results of the Ottawa dataset (%).

Method	TA	TB	TC	TD	Average
M1	46.21±9.63	92.04±4.88	92.85±4.42	97.35±2.25	82.11±5.23
M2	57.25±12.35	85.81±3.94	94.15±3.76	91.15±2.79	82.09±5.71
M3	43.58±12.30	91.15±7.24	89.52±7.41	93.77±4.80	79.51±7.94
M4	53.10±6.01	91.92±4.48	94.69±4.18	97.50±1.73	84.30±4.10
M5	60.00±16.28	88.52±5.14	92.38±4.76	96.48±1.38	84.35±6.89
M6	62.11±22.3	96.2±4.04	94.95±5.36	85.82±10.58	84.77±10.57
A	66.25±0.69	95.71±2.55	91.33±14.53	**99.13±1.14**	88.11±4.73
AD	**66.69±2.19**	93.23±6.40	98.67±2.14	97.10±3.96	88.92±3.67
AC	65.79±0.88	**97.40±1.28**	**99.25±0.65**	98.00±2.76	**90.11±1.39**
AF	66.15±0.97	91.13±10.82	88.62±14.15	94.13±9.21	85.01±8.79
AI	66.31±1.58	73.58±24.64	96.29±3.26	91.31±14.69	81.87±11.04

Remark 5: I level and F-level joint-augmentation is preferred to construct a feature convex hull in the feature space.

To investigate the domain discrepancy of learned features between the seen domains and the unseen domain, we use multiple MMD error in Eq.(6) as a criterion to verify our method. Table 11 exhibits the results of the estimated domain discrepancy via the MMD on the Ottawa dataset. As shown in Eq.(14), the domain discrepancy between the seen and unseen domains is critical for the DG performance. Comparing the results between M2 and A, the metric estimated by MMD of ADAG is 4.6% lower than that of M2, although M2 is directly trained by MMD. Notably, the proposed ADAG obtains the lowest multiple MMD error without accessing the target domain distribution, proving the augmented domain built by ADAG can cover the representation of the unseen domain. Through iteratively optimization with the augmented domain, the generalization ability can be increased.

Table 11 Estimated domain discrepancy via the MMD on the Ottawa dataset.

Method	TA	TB	TC	TD	Average
M1	**0.4734**	0.5458	0.4557	0.4526	0.4819
M2	0.7340	0.3991	**0.3180**	0.3951	0.4616
M3	0.5586	0.5462	0.5573	0.5621	0.5561
M4	0.6344	0.6615	0.4309	0.5384	0.5663
M5	0.8865	0.5224	0.5023	0.4693	0.5951
M6	0.6038	0.4980	0.4494	0.5179	0.5173
A	0.7145	**0.3479**	0.3999	**0.2986**	**0.4402**
AD	0.6869	0.4324	0.4660	0.4945	0.5200
AC	0.6782	0.4827	0.4357	0.4234	0.5050
AF	0.7457	0.4599	0.4408	0.4444	0.5227
AI	0.5777	0.5523	0.4749	0.3683	0.4933

Data quality is critical for our method to be used in industrial cases, where the data quality can be degraded by noise. To investigate this issue, we carry out a study of the impact of noisy data on the performance of our method. On the basis of a TD task, the comparative results of ERM, M5 and A are shown in Fig. 9. Noisy data with different signal noise ratio (SNR) are added to the signal to test the performance of the model under different noise conditions. As the SNR decreases, the impact of noise increases. As shown in the figure, the accuracy of the three methods decreases to certain degrees. Compared with the ERM and M5 methods, the proposed method can still maintain a higher accuracy even at a lower SNR, which shows the reliability and robustness of our method in industrial noisy environment.

$$SNR(\text{dB}) = 10log_{10}\left(\frac{P_{\text{signal}}}{P_{\text{noise}}}\right) \tag{36}$$

Remark 6: Data quality for diagnosis tasks determines the upper limit of accuracy that the model can achieve. Different data quality and data source can degenerate the model performance because the signal distribution have essentially discrepancy, which has been verified. Moreover, we have given evidence that signal range can have certain impact on the model accuracy as Remark 4 indicated. Therefore, data quality is the first issue to be considered in establishing a reliable diagnosis model.

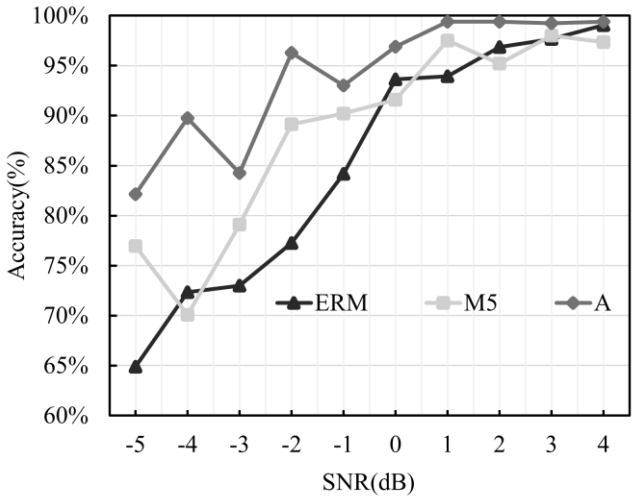

Fig. 9. Test accuracy under different Gaussian noise in a TD task.

4.5. Feature Visualization

DG based method aims to learn the domain-invariant features while our CDA diagnosis tries to generate the augmented domain to cover the unseen distribution. Therefore, this part use feature visualization to validate the above conclusions.

To show the distribution of fault features from the seen and unseen domains, Fig. 10 presents the 2-D features from the second layer of fault classifier by using T-SNE [45]. For an intuitive understanding, the feature vectors of SDUST dataset from four health conditions under task T1500 are plotted, and they are marked differently with 25 data points. Four colors of data points represent four domains. Specifically, the dark-green points are the features extracted from the unseen domains, whereas the points with the other colors are from available source domains.

The DG based method aims to learn the generalized features clustered together across different domains and even in the unseen domains. As shown in Fig. 10, the features learned through M1-M3 fail to capture the generalized representation in the I04 fault because of the large domain discrepancy among the seen and unseen domains. The M4 method may learn more robust features than M1-M3 but fails to cluster well in the seen domains. Owing to the idea of CDA diagnosis, the proposed ADAG can learn more domain-invariant features, thereby benefiting the classifier training and further reconfirming remark 2 in the first experiment.

For a clear comparison, Fig. 11 also shows the MMD error of each method. Considering the unseen working condition, the model training cannot access the data distribution in the unseen domain. The compared methods have larger and similar MMD errors on average. Generally, the multi-source cross-entropy guided by ERM is not competitive in DG.

Remark 7: as vicinal features are built by the feature extractor with available domains, the distribution of the augmented features attempts to cover the unseen data to enhance the model training. In this manner, the unseen distribution may lie within the convex hull of the features of the source domains.

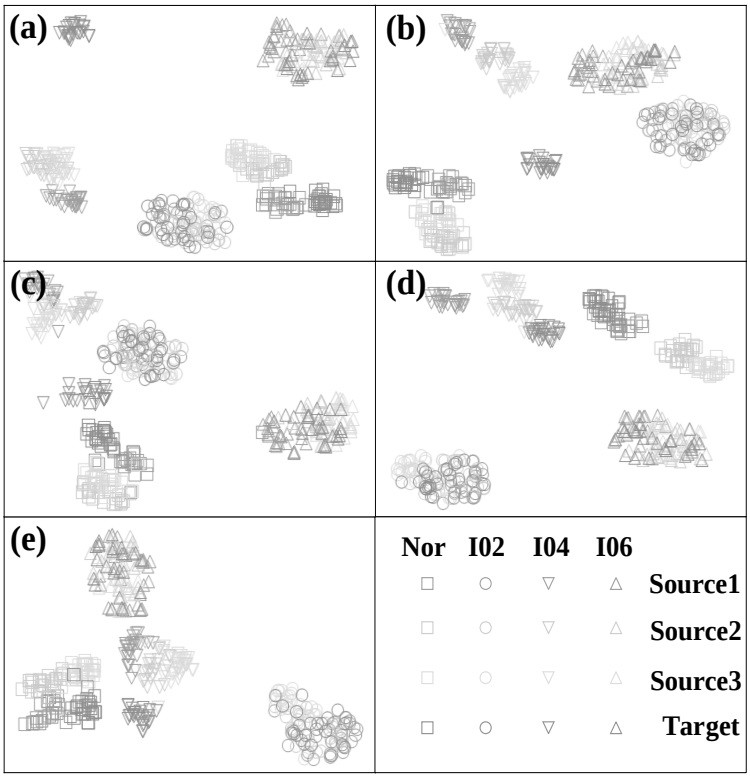

Fig. 10. Results of feature-dimension reduction via T-SNE under unseen target working condition: (a) M1, (b) M2, (c) M3, (d) M4, (e) A.

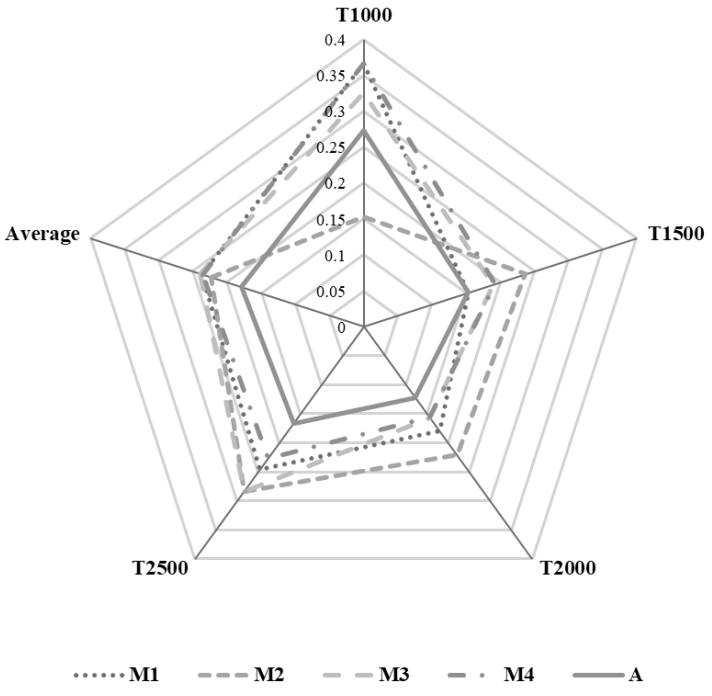

Fig. 11. Comparison of MMD errors in different tasks.

Furthermore, the effectiveness of augmented features is verified. Fig. 12 depicts the augmented features with the dark blue color of the proposed method from different tasks. As shown in panel (a), some of the features are not clustered well, which can explain why task T1000 in Table 7 is the most challenging. However, as shown in panels (b) and (c), although the unseen domain is inaccessible, the learned features are domain-invariant and fault-related. The clear decision boundary of each fault can be found to facilitate diagnosis.

For instance, the points of fault B04 under the unseen domain in panel (b) can be covered by the augmented features but cannot be covered by the source domains. In this way, the idea of CDA diagnosis remarkably boosts the generalization ability.

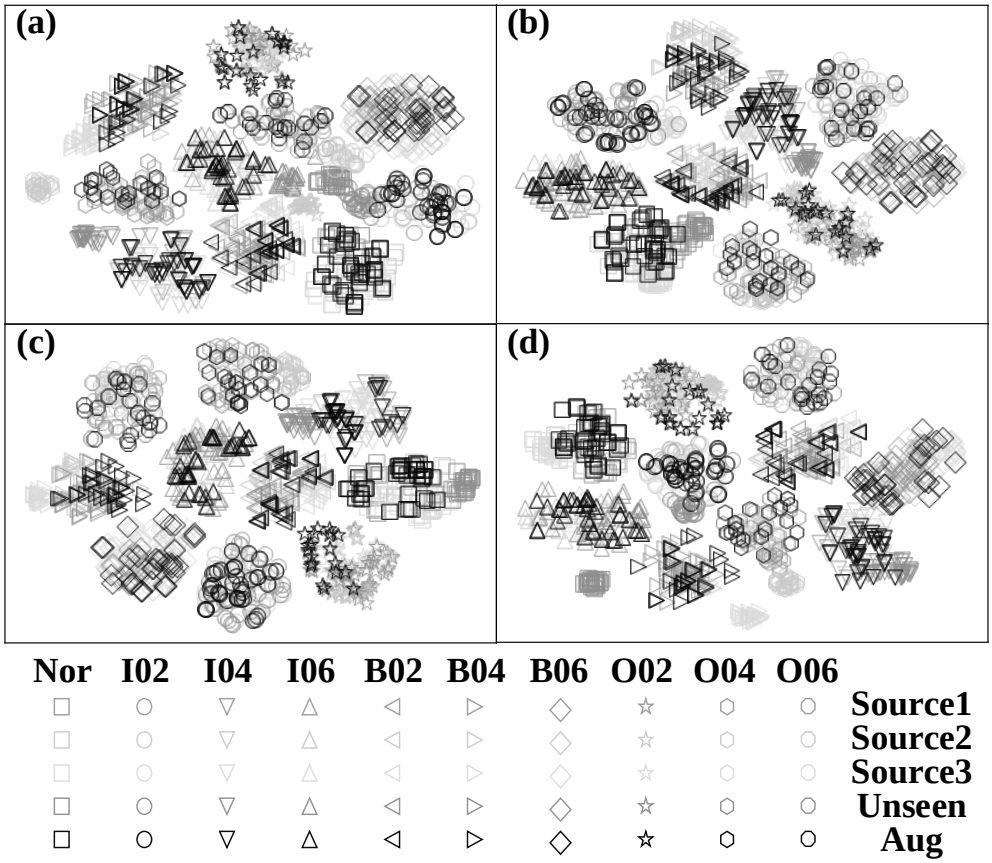

Fig. 12. Features distribution with augmented domain under unseen working condition: (a) T1000, (b) T1500, (c) T2000, and (d) T2500.

5. Summary and Future Work

Reliability of safety-critical assets under unseen working conditions is a major concern to conduct the health management. In this paper, a novel idea called CDA diagnosis based on DG and the associated ADAG are proposed to realize a robust and reliable fault diagnosis model under unseen working conditions. The CDA aims at building an augmented domain from available source domains, improving model training to diagnose the fault under unseen working conditions. As an implementation of CDA diagnosis, ADAG leverages convex combinations of features and instances to build an augmented domain. Using adversarial learning between feature extractor and domain classifier by multi-source domains and the augmented domain, the domain discrepancy could be

narrowed among available domains and the unseen domains. This process remarkably boosts the generalization ability of the diagnosis model. Finally, extensive experiments with the SDUST dataset and the Ottawa dataset prove the best performance of ADAG among the comparison methods, shedding the light on the prospect of the CDA diagnosis to manage the safety-critical assets.

In the future, some challenging tasks need to be further explored, e.g., the computational cost may have redundancy owing to the convex combination. Therefore, constructing augmented features through specific source features is valuable. Moreover, according to Eq.(11), the idea of CDA for diagnosis can be realized in other methods rather than limiting to ADAG. In the perspective of extended applications of our method, since the distribution shift of vibration signal collected from different machines always exists, our method is promising for other mechanical devices such as robots or pumps, and other industrial safety-critical assets.

Concerning the requirements of a real industrial fault diagnosis, the reliability and robustness of the method can be further improved. It is inspired to carry out continuous research on the impact of data quality, signal sensing consistency, etc. Potentially, we are conducting research on the implementation of our method to a SCARA robot, where the voltage and current signals are the only inputs to ADAG model, and the vibration signals are completely omitted. It is greatly encouraged by the industry.

Acknowledgments

This work is financially supported by the National Innovation and Development Project of Industrial Internet (No.TC190H3WR), in part by National Natural Science Foundation of China (No. 52272440 and 51875375), and the Research Project of State Key Laboratory of Mechanical System and Vibration, Shanghai Jiaotong University (No. MSV202104).

References

[1] Han X, Wang Z, Xie M, He Y, Li Y, Wang W. Remaining useful life prediction and predictive maintenance strategies for multi-state manufacturing systems considering functional dependence. Reliab Eng Syst Saf 2021;210:107560. https://doi.org/ 10.1016/ j.ress.2021.107560.

[2] Zhang C, Hu D, Yang T. Anomaly detection and diagnosis for wind turbines using long short-term memory-based stacked denoising autoencoders and XGBoost. Reliab Eng Syst Saf 2022;222. https://doi.org/10.1016/j.ress.2022.108445.

[3] da Costa PR de O, Akçay A, Zhang Y, Kaymak U. Remaining useful lifetime prediction via deep domain adaptation. Reliab Eng Syst Saf 2020;195:106682. https://doi.org/10.1016/j.ress.2019.106682.

[4] Kordestani M, Saif M, Orchard ME, Razavi-Far R, Khorasani K. Failure Prognosis and Applications - A Survey of Recent Literature. IEEE Trans Reliab 2021;70:728–48. https://doi.org/10.1109/TR.2019.2930195.

[5] Xia M, Shao H, Williams D, Lu S, Shu L, de Silva CW. Intelligent fault diagnosis of machinery using digital twin-assisted deep transfer learning. Reliab Eng Syst Saf 2021;215:107938. https://doi.org/10.1016/j.ress.2021.107938.

[6] Harary H. Measurement Science Roadmap for prognostics and health management for smart manufacturing systems. Nist 2015;53:1689–99.

[7] Guo X, Chen L, Shen C. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. Meas J Int Meas Confed 2016;93:490–502. https://doi.org/10.1016/j.measurement.2016.07.054.

[8] Chen Z, Li W. Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network. IEEE Trans Instrum Meas 2017;66:1693–702. https://doi.org/10.1109/TIM.2017.2669947.

[9] Ben-David S, Blitzer J, Crammer K, Kulesza A, Pereira F, Vaughan JW. A theory of learning from different domains. Mach Learn 2010;79:151–75. https://doi.org/ 10.1007/s10994-009-5152-4.

[10] Chen L, Li Q, Shen C, Zhu J, Wang D, Xia M. Adversarial domain-invariant generalization: a generic domain-regressive framework for bearing fault diagnosis under unseen conditions. IEEE Trans Ind Informatics 2022;18:1790–800. https://doi.org/10.1109/TII.2021.3078712.

[11] Panigrahi S, Nanda A, Swarnkar T. A Survey on Transfer Learning. Smart Innov Syst Technol 2021;194:781–9. https://doi.org/10.1007/978-981-15-5971-6_83.

[12] Wang X, Shen C, Xia M, Wang D, Zhu J, Zhu Z. Multi-scale deep intra-class transfer learning for bearing fault diagnosis. Reliab Eng Syst Saf 2020;202:107050. https://doi.org/10.1016/j.ress.2020.107050.

[13] Hu C, Wang Y, Gu J. Cross-domain intelligent fault classification of bearings based on tensor-aligned invariant subspace learning and two-dimensional convolutional neural networks. Knowledge-Based Syst 2020;209:106214. https://doi.org/10.1016/ j.knosys.2020.106214.

[14] Li Q, Shen C, Chen L, Zhu Z. Knowledge mapping-based adversarial domain adaptation: A novel fault diagnosis method with high generalizability under variable working conditions. Mech Syst Signal Process 2021;147. https://doi.org/10.1016/ j.ymssp.2020.107095.

[15] Jiao J, Zhao M, Lin J. Unsupervised Adversarial Adaptation Network for Intelligent Fault Diagnosis. IEEE Trans Ind Electron 2020;67:9904–13. https:// doi.org/ 10.1109/TIE.2019.2956366.

[16] Li R, Li S, Xu K, Lu J, Teng G, Du J. Deep domain adaptation with adversarial idea and coral alignment for transfer fault diagnosis of rolling bearing. Meas Sci Technol 2021;32. https://doi.org/10.1088/1361-6501/abe163.

[17] Jiao J, Zhao M, Lin J, Liang K. Residual joint adaptation adversarial network for intelligent transfer fault diagnosis. Mech Syst Signal Process 2020;145:106962. https://doi.org/10.1016/j.ymssp.2020.106962.

[18] Saito K, Watanabe K, Ushiku Y, Harada T. Maximum Classifier Discrepancy for Unsupervised Domain Adaptation. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 2018:3723–32. https://doi.org/10.1109/CVPR.2018.00392.

[19] Lee J, Kim M, Ko JU, Jung JH, Sun KH, Youn BD. Asymmetric inter-intra domain alignments (AIIDA) method for intelligent fault diagnosis of rotating machinery. Reliab Eng Syst Saf 2022;218:108186. https://doi.org/10.1016/j.ress.2021.108186.
[20] Wu A, Liu R, Han Y, Zhu L, Yang Y. Vector-Decomposed Disentanglement for Domain-Invariant Object Detection. Proc. IEEE Int. Conf. Comput. Vis., vol. 2, 2021, p. 9322–31. https://doi.org/10.1109/ICCV48922.2021.00921.
[21] Feng Y, Chen J, Yang Z, Song X, Chang Y, He S, et al. Similarity-based meta-learning network with adversarial domain adaptation for cross-domain fault identification. Knowledge-Based Syst 2021;217:106829. https://doi.org/10.1016/j. knosys. 2021. 106829.
[22] Wu A, Han Y, Zhu L, Yang Y. Instance-Invariant Domain Adaptive Object Detection via Progressive Disentanglement. IEEE Trans Pattern Anal Mach Intell 2021. https://doi.org/10.1109/TPAMI.2021.3060446.
[23] Huang Z, Lei Z, Wen G, Huang X, Zhou H, Yan R, et al. A multi-source dense adaptation adversarial network for fault diagnosis of machinery. IEEE Trans Ind Electron 2021;46:6298–307. https://doi.org/10.1109/tie.2021.3086707.
[24] Li X, Zhang W, Ding Q, Li X. Diagnosing Rotating Machines with Weakly Supervised Data Using Deep Transfer Learning. IEEE Trans Ind Informatics 2020;16:1688–97. https://doi.org/10.1109/TII.2019.2927590.
[25] Xia Y, Shen C, Wang D, Shen Y, Huang W, Zhu Z. Moment matching-based intraclass multisource domain adaptation network for bearing fault diagnosis. Mech Syst Signal Process 2022;168:108697. https://doi.org/10.1016/j.ymssp. 2021. 108697.
[26] Han T, Li YF, Qian M. A Hybrid Generalization Network for Intelligent Fault Diagnosis of Rotating Machinery under Unseen Working Conditions. IEEE Trans Instrum Meas 2021;70. https://doi.org/10.1109/TIM.2021.3088489.
[27] Zhuo Y, Ge Z. Auxiliary Information-Guided Industrial Data Augmentation for Any-Shot Fault Learning and Diagnosis. IEEE Trans Ind Informatics 2021;17:7535–45. https://doi.org/10.1109/TII.2021.3053106.
[28] Li X, Zhang W, Ding Q, Sun JQ. Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation. J Intell Manuf 2020;31:433–52. https://doi.org/10.1007/s10845-018-1456-1.
[29] Pei Z, Jiang H, Li X, Zhang J, Liu S. Data augmentation for rolling bearing fault diagnosis using an enhanced few-shot Wasserstein auto-encoder with meta-learning. Meas Sci Technol 2021;32. https://doi.org/10.1088/1361-6501/abe5e3.
[30] Zhang T, Chen J, Pan T, Zhou Z. Towards Intelligent Fault Diagnosis under Small Sample Condition via A Signals Augmented Semi-supervised Learning Framework. IEEE Int. Conf. Ind. Informatics, vol. 2020- July, 2020, p. 669–72. https://doi.org/10.1109/INDIN45582.2020.9442224.
[31] Matsuura T, Harada T. Domain generalization using a mixture of multiple latent domains. AAAI, 2020, p. 11749–56. https://doi.org/10.1609/aaai.v34i07.6846.
[32] Zhou K, Yang Y, Qiao Y, Xiang T. Domain Generalization with MixStyle. ICLR, 2021, p. 1–15.
[33] Li H, Pan SJ, Wang S, Kot AC. Domain Generalization with Adversarial Feature Learning. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 2018:5400–9. https://doi.org/10.1109/CVPR.2018.00566.
[34] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. Adv. Neural Inf. Process. Syst., vol. 3, 2014, p. 2672–80.
[35] Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, et al. Domain-adversarial training of neural networks. J Mach Learn Res 2016;17:2096–2030.
[36] Zhang H, Cisse M, Dauphin YN, Lopez-Paz D. MixUp: Beyond empirical risk minimization. 6th Int Conf Learn Represent ICLR 2018 - Conf Track Proc 2018:1–13.
[37] Albuquerque I, Monteiro J, Darvishi M, Falk TH, Mitliagkas I. Generalizing to unseen domains via distribution matching. ArXiv Prepr 2019:1–15.
[38] Chen L, Li Q, Shen C, Zhu J, Wang D, Xia M. Adversarial Domain-Invariant Generalization: A Generic Domain-Regressive Framework for Bearing Fault Diagnosis under Unseen Conditions.

IEEE Trans Ind Informatics 2022;18:1790–800. https://doi.org/10.1109/TII.2021.3078712.
[39] Zhang L, Zhang F, Qin Z, Han Q, Wang T, Chu F. Piezoelectric energy harvester for rolling bearings with capability of self-powered condition monitoring. Energy 2022;238:121770. https://doi.org/10.1016/j.energy.2021.121770.
[40] Jia S, Wang J, Han B, Zhang G, Wang X, He J. A novel transfer learning method for fault diagnosis using maximum classifier discrepancy with marginal probability distribution adaptation. IEEE Access 2020;8:71475–85. https://doi.org/10.1109/ ACCESS.2020.2987933.
[41] Han B, Zhang X, Wang J, An Z, Jia S, Zhang G. Hybrid distance-guided adversarial network for intelligent fault diagnosis under different working conditions. Meas J Int Meas Confed 2021;176:109197. https://doi.org/10.1016/j.measurement.2021. 109197.
[42] Huang H, Baddour N. Bearing vibration data collected under time-varying rotational speed conditions. Data Br 2018; 21:1745–9. https://doi.org/10.1016/j.dib.2018.11. 019.
[43] Zheng H, Wang R, Yang Y, Li Y, Xu M. Intelligent Fault Identification Based on Multisource Domain Generalization Towards Actual Diagnosis Scenario. IEEE Trans Ind Electron 2020;67:1293–304. https://doi.org/10.1109/TIE.2019.2898619.
[44] Zhao S, Wang G, Zhang S, Gu Y, Li Y, Song Z, et al. Multi-source distilling domain adaptation. AAAI 2020 - 34th AAAI Conf. Artif. Intell., 2020, p. 1295–12983. https://doi.org/10.1609/aaai.v34i07.6997.
[45] Maaten L van der, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res 2008;9:2579–605. https://doi.org/10.1007/s10479-011-0841-3.