OPEN ACCESS
IOP Publishing | ISECAEP
Journal of Reliability Science and Engineering
J. Reliab. Sci. Eng. **2** (2026) 015302 (17pp)
https://doi.org/10.1088/3050-2454/ae4e31

A novel cross-domain open-set fault diagnosis model for distribution boxes in finishing rolling group

Yonglin Guo[1], Shihao Li[1], Tangbin Xia[2], Di Zhou[1,*] and Ershun Pan[2]

[1] School of Mechanical Engineering, Donghua University, Shanghai 200051, People's Republic of China
[2] Department of Industrial Engineering, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China

E-mail: zhoudi@dhu.edu.cn

Received 1 December 2025, revised 8 February 2026
Accepted for publication 5 March 2026
Published 26 March 2026

Abstract

As a critical link in the drive system of the finishing mill, distribution boxes play a vital role in rolling force transmission and distribution. Therefore, researching fault diagnosis for distribution boxes is essential for ensuring the security and reliability of process operation. However, when fault characteristics are indistinct and the fault type is unknown, traditional diagnostic methods may fail to provide reliable diagnosis. To address these challenges, this paper proposes a cross-domain open-set model that leverages adversarial learning to enhance feature representations under indistinct fault characteristics, and integrates distribution-based methods for identifying unknown faults. Considering the importance of feature representation for adversarial learning and unknown fault identification, we build an information-exchange feature extractor integrates depthwise separable convolution and GCN. This structure extracts depth features and addresses the channel independence limitation of traditional depthwise separable convolution. To identify unknown faults in distribution boxes, the saddle-point approximation (SPA) method is employed. This method is used during training to establish a distribution for each known class, and these distributions are utilized in testing to identify unknown samples. Unlike traditional methods for establishing distributions, the SPA method directly uses data without relying on predefined distribution assumptions or formulas. To verify the superiority of the proposed method, we conducted a comparative experiment based on historical fault data of distribution boxes. The results demonstrate that the proposed method achieves an accuracy of 94.13% and an H-score of 91.35%, which demonstrates the superior performance in classifying known faults and identifying unknown faults in distribution boxes.

Keywords: fault diagnosis, open-set, adversarial network, finishing mill, distribution boxes

* Author to whom any correspondence should be addressed.

1. Introduction

The hot rolling flow, as shown in figure 1, typically consists of four major stages: heating furnace, roughing mill, finishing mill, and cooling. Among these, the finishing rolling process in hot rolling plays a decisive role in determining the final quality of the hot-rolled steel [1]. This stage involves multiple identical finishing mills, each performing nearly identical rolling operations. Within the finishing mills, the distribution boxes play a critical role in distributing and regulating power in the drive system [2]. The relationship between the distribution box and the finishing mill is shown in figure 2. In the finishing mill group, the F1–F7 stands adopt a four-roll irreversible horizontal mill configuration arranged in a continuous layout, and the detailed structural schematic of the finishing mill is illustrated figure 3. However, due to heavy stress, vibrations, and other adverse factors, these components are particularly susceptible to different kinds of faults [3]. Considering the complex environment and coupled signals, fault diagnosis becomes a challenge [4]. To minimize downtime caused by diagnostic delays, the development of an effective fault diagnosis model for distribution boxes is critical.

With advancements in data processing and computational power, the data-driven intelligent fault diagnosis for rolling equipment has gained increasing attention [5, 6]. For the unique characteristics of rolling mill bearing vibration signals, Zhao *et al* [7] used CNN to improve diagnosis accuracy. Yu *et al* [8] incorporated multi-sensor information into the DBN, improving diagnostic ability under limited data. To address the challenge of data imbalance, Shi *et al* [9] employed a dual attention mechanism for comprehensive feature extraction from the input. Guo *et al* [10] considered both the temporal and spatial features of faults signals, constructing a model that combines CNN and Bi-LSTM. To enhance the capability and robustness of the diagnostic model, Peng *et al* [11] integrated optimization algorithms into the deep belief network. Considering further the fluctuations in data quality under different operating conditions, the accuracy of diagnoses using low-quality data may decrease. Therefore, several scholars have applied transfer learning methods to rolling equipment. To overcome the difficulties caused by variable working conditions, Li *et al* [12] developed a multisource domain adversarial graph convolutional network (GCN). Hou *et al* [13] used image data from multiple sensors to tackle the challenges of variable operating conditions. These applications of intelligent methods effectively address the diagnostic challenges in rolling equipment and achieve significant results.

Despite their promising performance, most of these methods implicitly assume that all fault types encountered during operation are known and available during training. However, these methods rely on the assumption that all fault types are known; that is, all fault types are labeled. In real-world processes, unknown faults will inevitably emerge over time, but the model may fail to identify these new faults without labels [14]. Like distribution boxes, common faults such as gear surface cracks and gear tooth spalling are likely to occur and be labeled, while flat-headed sleeve tooth cracks are rare. Although such rare faults occur infrequently, they typically belong to low-frequency but high-risk failure modes that are difficult to collect and label in advance. When the model is applied for diagnosis, it may mistakenly classify such unknown faults as known fault types. Therefore, the capability to identify unknown fault samples that deviate from known fault categories is essential for preventing confident misclassification and enabling timely maintenance intervention. Recently, open-set diagnostic approaches have attracted significant attention in fault diagnosis field [15]. Open-set diagnosis refers to methods that can classify known faults and identify unknown ones [16]. Some studies have been conducted on this topic. Razavi-Far *et al* [17] developed a fault detection and classification model to detect new faults without requiring updates. Based on subspace learning, Tian *et al* [18] proposed a diagnosis model with self-adaptive capability for open-set. In Peng *et al* [19], negative out-of-distribution data are generated using the Soft Brownian Offset sampling method to simulate previously unknown faults. Jung [20] combined Weibull-calibrated one-class support vector machines to establish fault classes for identifying unknown faults. Zhang *et al* [21] employed a novel *k*-means feature clustering algorithm to distinguish unknown faults based on feature distance. These studies have demonstrated the effectiveness of open-set fault diagnosis in overcoming the challenges posed by unknown faults and overcoming the limitations of traditional methods.

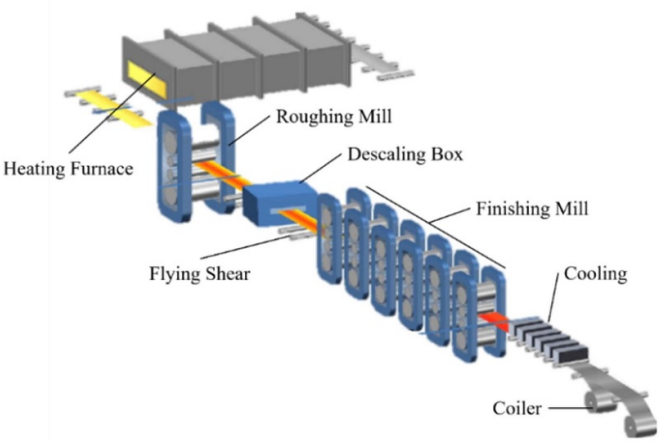


Figure 1. Overview of the hot rolling flow.

Building on these observations, recent studies have begun to explore open-set fault diagnosis methods that can handle unknown faults. However, in practical rolling processes, these challenges are often coupled with varying working conditions, making the integration of open-set techniques with domain adaptation particularly necessary. During the rolling process, improper sensor installation and environmental noise may obscure fault-related components in the acquired signals, leading to a reduced signal-to-noise ratio and weakened fault characteristic features, which ultimately degrades the data quality. Additionally, unknown faults may occur, as illustrated in figure 4. To address these challenges, integrating open-set methods into domain adversarial neural network (DANN) present a promising solution [22]. This method can effectively

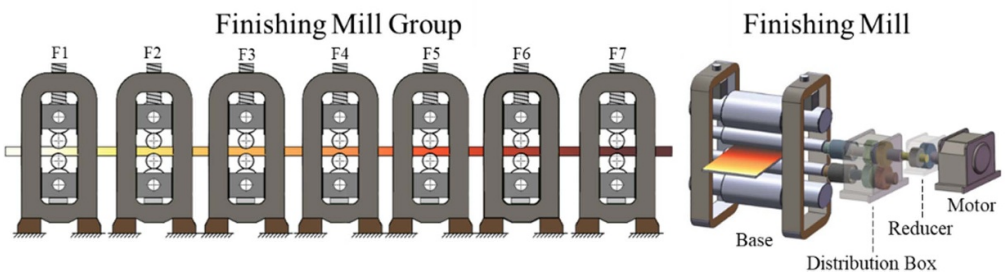


Figure 2. Arrangement of the distribution box within the finishing mill group.

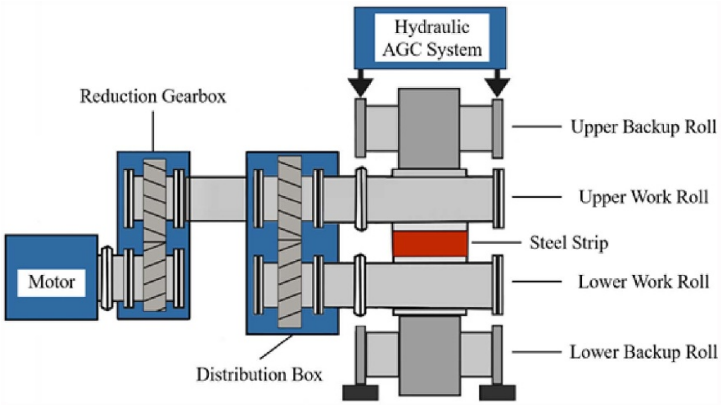


Figure 3. Finishing mill structure.

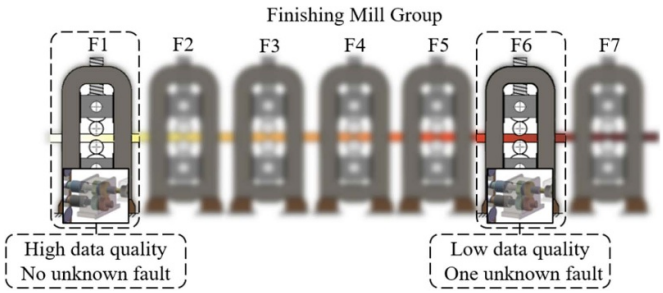


Figure 4. Fault data scenario of rolling mill group.

address scenarios involving indistinct fault characteristics as well as unknown faults under different working conditions. The advantages of this method, compared to direct diagnosis methods and standard DANN, are shown in figure 5. And some scholars have focused on researching this method. For instance, Wang *et al* [23] proposed a method that combines weighted domain adaptation with dual classifiers to recognize unknown faults. Zhu *et al* [24] built multiple complementary classifiers to identify both known and unknown fault categories in the target domain. Zhao and Shen [25] developed an outlier detection module within the domain network to classify both known and unknown faults. Chen *et al* [26] introduced a weighted learning strategy to adaptively differentiate between samples from known and unknown classes. Further improving the interpretability of the model, some scholars use data for distribution modeling to identify and eliminate unknown faults. Zhang *et al* [27] modeled distributions using statistical extreme value theory to identify unknown faults. Wang *et al* [28] based on the Mahalanobis distance to model distributions for unknown classes to distinguish unknown faults. Yu *et al* [29] combined DANN with Weibull distribution for cross-domain open-set fault diagnosis. Zhang *et al* [30] improved the OpenMax function for open-set recognition. Gao *et al* [31] based the source domain of DANN to establish distributions for open-set sample recognition. However, these methods rely on predefined distribution function formulas to establish distributions. The straightforward use of these formulas generally depends on parameter estimates, which may not capture the intricate distribution characteristics of the data accurately, especially in the tail regions or for complex distributions. This inaccuracy in establishing distributions ultimately impairs the identification of unknown faults.

For this motivation, this research proposes a model that combines DANN with saddle-point approximation (SPA) to address the above-mentioned problem. We have developed a new feature extractor within the DANN framework to obtain more effective feature representations of the input data. During the training phase, the SPA method is employed to model the distribution of each class, while in the testing phase, the distribution is used to assess whether a sample is unknown. Furthermore, we introduce an adaptive threshold method to overcome the limitations of traditional hard thresholds, which rely on domain knowledge

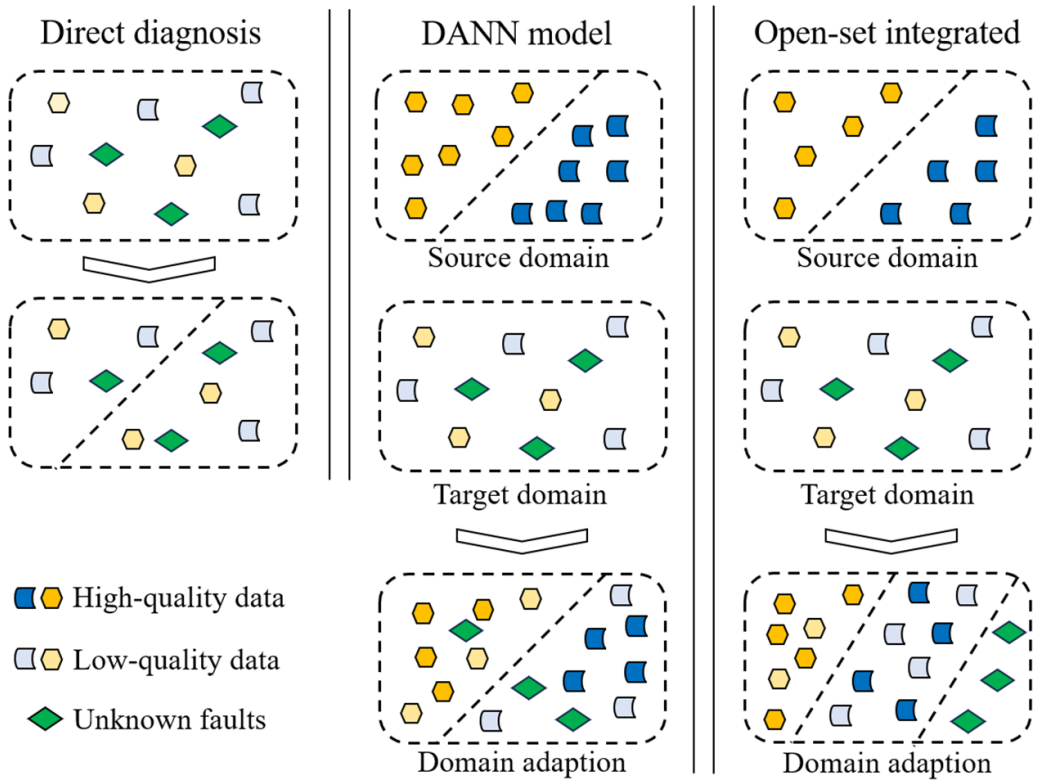


Figure 5. Traditional diagnosis model and the proposed model.

and expert experience. The proposed method is evaluated with historical fault data from the distribution boxes, and the results demonstrate its effectiveness in engineering applications.

The main contributions of this research are as follows.

(1) Considering that data collected from distribution boxes in finishing mills may exhibit weak fault characteristics and that certain fault types are unavailable during training, this study develops a cross-domain open-set fault diagnosis model to enhance diagnostic reliability under realistic industrial conditions.

(2) To improve feature representation under such challenging conditions, a novel feature extractor is designed by integrating depthwise separable convolution with graph neural networks. This structure not only extracts local depth features but also enables effective information exchange among channels, alleviating the limitation of channel independence in traditional depthwise separable convolution.

(3) To identify unknown fault samples without relying on predefined distribution assumptions, the proposed model incorporates the SPA method to construct class-wise feature distributions directly from data. Furthermore, an adaptive threshold strategy is introduced to account for distribution differences among fault classes, improving the robustness of unknown fault identification.

The rest content is organized as follows: section 2 reviews related work. Section 3 describes the proposed model. Section 4 presents the experimental results, and section 5 concludes the research.

2. Preparations

2.1. Problem definition

Considering different working conditions and open-set scenarios in engineering applications, data with clear fault characteristics are used as the source domain, while data exhibiting weak or indistinct fault characteristics and unknown faults are treated as the target domain. Both the source and target domains data are collected from different distribution boxes within the same group. The source domain contains K labeled classes, and the target domain has $K+1$ classes, adding one unlabeled class to the labeled ones. The source data $S=\{(x_1,y_1),(x_2,y_2),\ldots,(x_{n_s},y_{n_s})\}$ consist of n_s labeled samples used for training, while the target dataset $T=\{x_1,\ x_2,\ldots,x_{n_t},x_{\text{unknown}}\}$ consist of n_t unlabeled samples and additional class of unknown samples used for testing. The domain label of source samples is 1 and the domain label of each target samples is 0. The capability evaluation of the proposed method is to accurately classify the K known classes to their corresponding classes and identify the unknown samples as belonging to the $K+1$ classes.

2.2. DANN

The DANN typically consists of a feature extractor G_{f}, a label classifier G_{l} and a domain classifier G_{d}. The feature extractor G_{f} extracts deep features from the input data [32]. The label classifier G_{l} classifies the input data in the source

domain during training. The domain discriminator G_{d} predicts whether a given feature belongs to the source or target domain [33]. Additionally, a gradient reversal layer (GRL) is placed between the feature extractor and domain classifier during training. This layer flips the gradients during backpropagation, encouraging the feature extractor to learn domain-invariant features that are similar across the source and target domains.

The training process of DANN mainly involves optimizing both the label classifier G_{l} and the domain classifier G_{d}. The G_{l} is trained to classify the source domain categories, while the G_{d} aims to differentiate between features from the source and target domains. Finally, a feedback mechanism is used to balance domain alignment and classification. Below, we define the relevant parameters and losses used in the optimization process,

$$E(\theta_f,\theta_l,\theta_d)=\sum_{i=1,\ldots,N;d_i=0} L_l^i(\theta_f,\theta_l)-\lambda\sum_{i=1,\ldots,N} L_d^i(\theta_f,\theta_d), \tag{1}$$

where θ_f, θ_y, θ_{d} are the parameters of the feature extractor G_f $(\cdot;\theta_f)$, label classifier G_l $(\cdot;\theta_l)$, and domain classifier G_d $(\cdot;\theta_d)$, respectively. L_l $(\cdot,\cdot)$ represents the loss of label classifier, and L_d $(\cdot,\cdot)$ denotes the loss of the domain classifier, with x_l^i and x_d^i being the corresponding losses for the ith sample. The trade-off parameter λ controls the contribution of the domain predictor's loss in the overall loss function.

In DANN, an adversarial training mechanism is applied to optimize equation (1). Specifically, a GRL is embedded into the network to facilitate domain-invariant learning. As a result, the parameters of the DANN are updated using stochastic gradient descent during backpropagation. The parameter updates are defined as follows:

$$\left(\hat{\theta}_f,\hat{\theta}_y\right)=\arg\min_{\theta_f,\theta_l} E\left(\theta_f,\theta_l,\hat{\theta}_d\right), \tag{2}$$

$$\left(\hat{\theta}_d\right)=\arg\max_{\theta_d} E\left(\hat{\theta}_f,\hat{\theta}_l,\theta_d\right). \tag{3}$$

2.3. SPA method

The SPA is a specific application of the mathematical saddle-point technique in statistics [34]. By utilizing the cumulant-generating function (CGF), the SPA method provides a precise approximation of the probability density function (PDF) for input random variables. This approach is effective in scenarios where the distributions of the random variables are unknown.

CGF refers to the natural logarithm of the moment-generating function (MGF) of a random variable [35]. This function is used to describe the cumulant characteristics of the random variable. The CGF can be expressed as:

$$K_{X_i}(t)=\ln\left[M_{X_i}(t)\right]. \tag{4}$$

For a random variable X, the MGF is defined as:

$$M_{X_i}(t)=\int_{-\infty}^{+\infty}\exp(tx_i)f_{Xi}(x_i)\,\mathrm{d}x_i, \tag{5}$$

where X refers to a random variable, and x denotes an instance of X. For a vector $\mathbf{X}$ containing random variables $[X_1, X_2, \ldots, X_n]^{\mathrm{T}}$. its observed counterpart is expressed as $\mathbf{x}=[x_1, x_2, \ldots, x_n]^{\mathrm{T}}$.

When the distribution type is known, the CGF of a random variable can be obtained by the CGF formula of the corresponding distribution. Additionally, the properties of the CGF are described by follows formula:

$$K_{Y_1}(s)=\sum_{j=1}^{n}K_{X_j}(a_j s)+\gamma s \quad \text{if } Y_1=\sum_{j=1}^{n}a\xi_j X_j+\gamma, \tag{6}$$

where $X_1, \ldots X_i, \ldots, X_n$ denote independent random variables, with their CGFs represented as $K_{X1}(t)$, $\ldots K_{Xi}(t)$, $\ldots$, $K_{Xn}(s)$. The function $Y=g(\mathbf{X})$ is approximated using a first-order Taylor series expansion, given as:

$$Y=g(x^*)+\sum_{j=1}^{n}\left.\frac{\partial g}{\partial X_j}\right|_{\mathrm{X}=\mathrm{x}^*}(X_j-x_j^*), \tag{7}$$

where $\mathbf{x}^*=[x_1^*, \ldots, x_i^*, \ldots, x_n^*]$ and the expansion point x_i^* takes a value within the defined domain of $\boldsymbol{X}$. The mean μ_i and the mode x_i^p are used to evaluate the expansion function. Using equation (6) the CGF of Y is derived and expressed as follows:

$$K_Y(t)=K_X\left\{\left[g\left(\mathrm{x}^*\right)-\sum_{i=1}^{n}\left.\frac{\partial g}{\partial X_i}\right|_{X=\mathrm{x}^*}\mathrm{x}_i^*\right]+\sum_{i=1}^{n}\left.\frac{\partial g}{\partial X_i}\right|_{X=\mathrm{x}^*}X_i\right\}. \tag{8}$$

Considering the distributions type of a random variable is unknown, the CGF cannot be counted by equations (4) and (5) directly [36]. Therefore, the power expansion method is employed to estimate the CGFs of several distributions [37]. This method can be expressed as:

$$\tilde{K}_Y(s)=\sum_{\iota=\iota}^{\infty}\kappa_\iota\frac{s^\iota}{\iota!}, \tag{9}$$

in which κ_ι $(\iota=1, 2, \ldots)$ represents the ιth cumulant of Y, and the first four cumulants of the random variables in equation (9) are determined as follows:

$$\begin{cases}\kappa_1=\mu'_1=E(Y)\\ \kappa_2=\mu'_2-{\mu'_1}^2=E(Y-E(Y))^2\\ \kappa_3=\mu'_3-3\mu'\mu'_2+{\mu'_1}^3=E(Y-E(Y))^3\\ \kappa_4=\mu'_4-4\mu'\mu'_3-3{\mu'_2}^2+12{\mu'_1}^2\mu'_2-6{\mu'_1}^4\\ \quad =E(Y-E(Y))^4-3(\kappa_2)^2\end{cases}, \tag{10}$$

where $\mu'_\iota(\iota=1,2,3,4)$ represents the first four central moments, and E denotes the mathematical expectation [38]. Subsequently, the SPA is subsequently employed to describe the distribution by the following function:

$$f_Y(y)=\left\{\frac{1}{2\pi K_Y''(s)}\right\}^{1/2}\exp\left[K_Y(s)-sy\right]. \tag{11}$$

Here, $K_Y'(\cdot)$ and $K_Y''(\cdot)$ represent the first- and second-order derivatives of the CGF, respectively. Specifically, the saddle point s is obtained as the real root of the following equation:

$$K_Y'(s) = y. \tag{12}$$

Using equation (12), the saddle point is determined as follows:

$$\tilde{\kappa}_1 s^0 + \frac{1}{1!}\kappa_2 s^1 + \frac{1}{2!}\kappa_3 s^2 + \frac{1}{3!}\kappa_4 s^3 = 0, \tag{13}$$

in which $\tilde{\kappa}_1 = \kappa_1 - y$. Then, introducing Cardano's method to obtain a reasonable saddle point. This method is applied to depressed cubic $t^3 + pt + q = 0$. To eliminate the s^2 term in equation (13), the substitution formula is applied as follows:

$$s = x - \frac{\kappa_2}{3\kappa_3}. \tag{14}$$

Substituting equation (14) into equation (13) yields the standard form of p and q.

$$p = \frac{3\kappa_3\kappa_1 - \kappa_2^2}{3\kappa_3^2},\ q = \frac{2\kappa_2^3 - 9\kappa_3\kappa_2\kappa_1 + 27\kappa_3^2\kappa_0}{27\kappa_3^3}. \tag{15}$$

The Final, using Cardano's method to obtain the saddle-point s. The formula as follow:

$$s = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}. \tag{16}$$

3. Fault diagnosis model for distribution boxes

This model is designed to address the cross-domain and open-set challenges associated with the distribution boxes of the rolling group. The DANN model introduced as the foundational framework to address the fault diagnosis challenges across different operational domains. Additionally, the SPA method is integrated to tackle potential open-set faults in unknown scenarios.

3.1. Feature extractor for enhanced channel information exchange

Depthwise separable convolution produces more compact and discriminative feature representations compared to traditional convolutional methods [39]. Consequently, incorporating depthwise separable convolution as the feature extractor enhances both domain alignment and classification performance within DANN. Additionally, the compact and discriminative representations can improve the accuracy of modeling distributions. However, depthwise convolution processes each channel separately, which limits the exchange of spatial information between channels. To address the limitation, a GCN is integrated to enhance inter-channel dependencies [40]. GCN enables weighted integration of information across the channels in depthwise separable convolution and replaces the traditional approach of direct concatenation. The proposed feature extractor is shown in figure 6.

Unlike traditional convolution, depthwise separable convolution has two steps. The first is a depthwise convolution that processes each input channel independently. The second is a point convolution using 1×1 windows for feature fusion [41]. The corresponding representation is as follows:

$$Y_{h,\omega,c} = \sum_{kh=0}^{K_h-1}\sum_{k\omega=0}^{K_\omega} X_{h+kh,\omega+k\omega,c} \cdot W_{kh,k\omega,c}^{(d)}, \tag{17}$$

where $W^{(d)} \in \mathbf{R}^{K_h \times K_\omega \times C}$ is the convolutional kernel of the depthwise convolution, with K_h and K_ω denoting the height and width of the convolutional kernel, respectively. A point convolution is then used to fuse the output of the depthwise convolution. The representation of the point convolution is as follows:

$$Z_{h,\omega,k} = \sum_{c=0}^{C-1} Y_{h,\omega,c} \cdot W_{c,k}^{(p)}, \tag{18}$$

where $W^{(p)} \in \mathbf{R}^{C \times K}$ represents the convolutional kernel of the point convolution, and K is number of output channels from the depthwise convolution.

To enhance inter-channel dependencies, the feature matrix is combined with the weight matrix of the GCN. The formula can be expressed as:

$$H_{GCN} = \hat{A} \cdot H_{conv} \cdot W_{GCN}, \tag{19}$$

where $H_{conv} \in \mathbf{R}^{N \times C}$ is the output feature matrix of convolution process, $\hat{A} \in \mathbf{R}^{N \times N}$ is the adjacency matrix, and W_{GCN} is the weight of GCN.

3.2. SPA in open-set fault identification

Deep feature representations often exhibit complex and unstable distributions, especially under cross-domain conditions, which makes direct distribution modeling challenging. To address this issue, the proposed method constructs class-wise distributions based on the distance between each sample and its corresponding class center, rather than directly modeling the high-dimensional features. During training, data from the source domain are used to establish the distributions for each known class, which are subsequently employed to diagnose unknown faults in the target domain. The process of establishing a distribution for a class is shown in figure 7, with a detailed description of the distribution establishment provided below.

The f_i^C denote the feature representation of a source sample, where i represents the ith sample, and c represents the cth class. Let $s_i = f_i^c$ represent the ith sample from class c. Each class is represented by the mean feature m_c, calculated as the average of the feature representations of training samples from class c:

$$m_c = \text{mean}\,(s_i)\,. \tag{20}$$

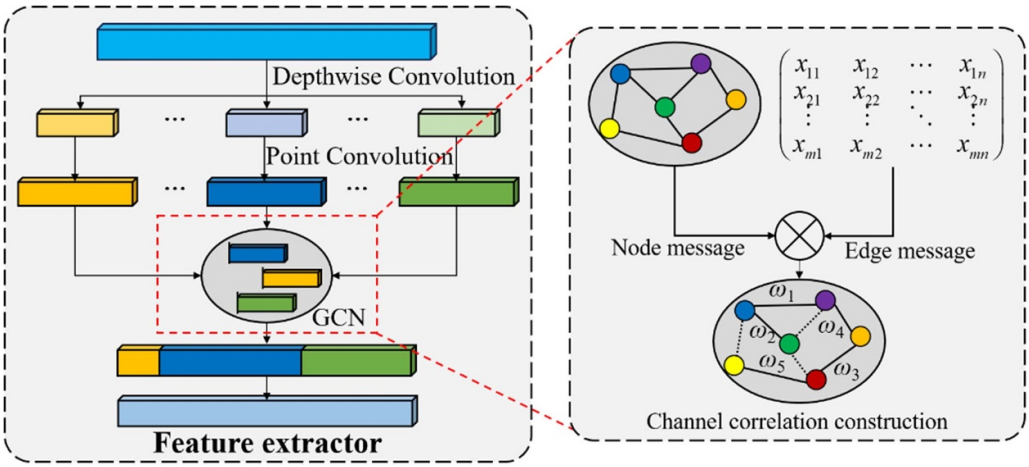


Figure 6. Structure of feature extractor.

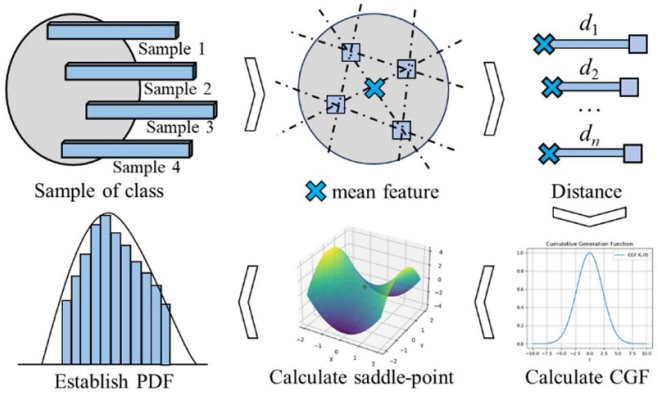


Figure 7. Saddle-point approximation for distribution.

The distance from each sample to its corresponding class mean m_c is calculated as follows:

$$d_i^c = \|s_i - m_c\|. \tag{21}$$

Due to the uncertainty in the source domain data distribution, estimating it directly using the PDF formula may be inaccurate. Thus, an approximate CGF is constructed to aid in deriving the PDF. For each class c, the cumulative amount is calculated as follows:

$$\begin{cases} \kappa_c^{(1)} = \frac{1}{n_c}\sum_{i=1}^{n_c} d_i^c \\ \kappa_c^{(2)} = \frac{1}{n_c}\sum_{i=1}^{n_c}\left(d_i^c - \kappa_c^{(1)}\right)^2 \\ \kappa_c^{(3)} = \frac{1}{n_c}\sum_{i=1}^{n_c}\left(d_i^c - \kappa_c^{(1)}\right)^3 \\ \kappa_c^{(4)} = \frac{1}{n_c}\sum_{i=1}^{n_c}\left(d_i^c - \kappa_c^{(1)}\right)^4 \end{cases}, \tag{22}$$

where n_c is the number of samples in class c. The CGF $K_c(t)$ for class c can is expressed as follow:

$$K_c(t) \approx \sum_{i=1}^{4} \kappa_c^{(i)} \frac{t^i}{i!}, \tag{23}$$

where $\kappa_c^{(i)}$ donates the cumulative amount obtained through sample estimation.

Then, the PDF of each class can be estimated based on the approximate CGF and SPA methods without needing a known PDF. The saddle-point is calculated as follows:

$$\frac{\partial K_c(t)}{\partial t} = 0, \tag{24}$$

and PDF of class c is approximated by saddle-point method, as shown in the following formula:

$$p_c(f) \approx \exp\left(K_c(s_c)\right) \Big/ \sqrt{\frac{\partial^2 K_c(s_c)}{\partial t^2}}. \tag{25}$$

With SPA, the distributions for each class in training phase is estimated. In the testing phase, the distance from each sample to m_c is then used to identify unknown samples.

3.3. Adaptive threshold set for distribution

After the distribution is established during the training process, thresholds are required to make decisions on the test samples. Considering that different classes exhibit varied data distributions, employing a fixed threshold for all lacks adaptability. To overcome this limitation, an adaptive threshold mechanism utilizing a set of candidate thresholds is introduced. This mechanism generates candidate thresholds in small step sizes, facilitating a precise search for the optimal value within the potential range for each class. The adaptive threshold approach assigns a unique threshold to each class, enabling more accurate identification of unknown samples.

First, define a set of adaptive thresholds $\{\Omega_c\}$ for each category c, and the set is defined as follow:

$$\{\Omega_c\} = [\ldots, \varepsilon_c - 2\gamma, \varepsilon_c - \gamma, \varepsilon_c, \varepsilon_c + \gamma, \varepsilon_c + 2\gamma, \ \ldots], \tag{26}$$

where ε_c denotes the initial distance reference for category c, and γ controls the interval of the candidate thresholds, resulting in a finite discrete candidate set.

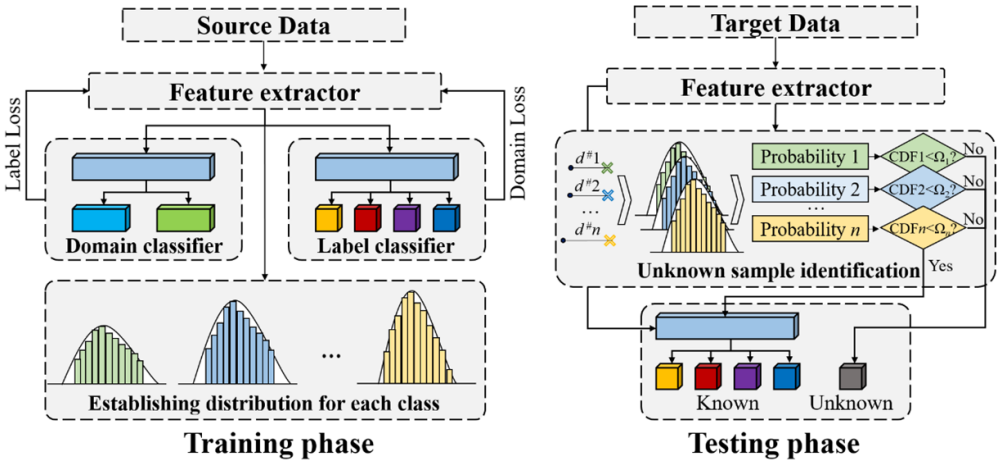


Figure 8. The structure of the fault diagnosis model.

During each training epoch, the classification effectiveness is evaluated using the H-value for each threshold set $\{\Omega_c\}$. Specifically, after the forward pass of the current epoch, sample-to-center distances are computed and corresponding classification decisions are obtained under each candidate threshold. The H-value is then calculated for each candidate threshold based on these decisions. The threshold that produces the highest H-value is then chosen as the optimal threshold, the $\{\Omega_c\}$ selection is present as follow:

$$\{\Omega_c^*\} = \arg \max_{\Omega \in \{\Omega_c\}} H\text{-value}\left(\{\Omega_c\}\right). \tag{27}$$

The selected optimal threshold $\{\Omega_c^*\}$ is updated at each training epoch and fixed after training for the inference stage.

The adaptive threshold search is limited to the training phase and conducted over a small discrete candidate set in a low-dimensional statistical space, resulting in minimal additional computational cost while providing enhanced robustness in practical engineering applications.

3.4. The fault diagnosis process of the distribution boxes in rolling mill group

The model employs adversarial learning to address the issue of data exhibiting indistinct fault characteristics data. Within the open-set method, the workflow of proposed model is divided into training and testing phases to illustrate how the SPA method identifies unknown samples, as shown in figure 8.

The main workflow of proposed model is presented in algorithm 1. The model primarily consists of a feature extractor for representation learning and a label classifier to perform classification. Additionally, the identification of open-set is divided into training phase and testing phase. Detailed steps are provided in algorithms 2 and 3.

During the training process, the SPA method is employed to establish the distributions for each known class, which facilitates the identification of unknown faults. The algorithm is detailed as follows.

During testing, the distribution established in the training process is utilized to identify unknown faults. Additionally, an adaptive threshold method is incorporated to enhance accuracy when identifying these faults.

Algorithm 1. Overview of proposed model.

Initialize network parameters
Construct a grid of candidate hyperparameter values
Forward-propagation:
Input: Source data $\boldsymbol{S}$ and target data $\boldsymbol{T}$.

1) Feed the source data and target data into feature extractor G_f to obtain source samples features $G_f(x_i^S)$ and target samples features $G_f(x_i^T)$;
2) Feed the $G_f(x_i^S)$ and $G_f(x_i^T)$ into label classifier G_l and domain classifier G_d simultaneously;
3) Feed $G_f(x_i^S)$ into label classifier G_l to obtain label classification $\hat{y}_i^S = G_f(x_i^S)$ and compute the cross-entropy loss L_{class} between $\hat{y}_i^S$ and y_i^S;
4) Apply gradient reversal on $G_f(x_i^S)$ and $G_f(x_i^T)$ using gradient reversal layer (GRL) with weight λ
5) Feed $G_f(x_i^S)$ and $G_f(x_i^T)$ into the domain classifier G_d to predict domain labels d_i (0 for source, 1 for target);
6) Utilize the SPA method to estimate distribution for open-set recognition. Training step and testing step are provided in algorithms 2 and 3.
7) Compute the domain classification loss L_{domain} for d_i;
8) Compute the total loss $L = L_{\text{class}} + \lambda \cdot L_{\text{domain}}$;
9) Update G_l, G_l and G_d by backpropagating the total loss L using Adam optimizer;

Output: The proposed model.
Select the model with the highest accuracy under hyperparameter values.

4. Experimental analysis of fault diagnosis model

4.1. Dataset acquisition

The finishing rolling process consists of seven finishing mills (F1–F7), each equipped with a distribution box, so there are seven distribution boxes (D1–D7) in total in the finishing mill group. Faults in distribution boxes lead to abnormal vibration, so vibration acceleration sensors are selected for fault signals

Algorithm 2. SPA method for open-set recognition (training).

Input: source samples feature $G_f(x_i^S)$ from feature extractor G_f;
1) Let $s_i = f_i^c$ represent the ith sample from class c;
2) Compute class mean features for each class,
$m_c = \text{mean}(s_i)$;
3) Compute each sample distance to m_c, $d_i^c = \|s_i - m_c\|$;
4) Construct approximate CGF for each class,
$K_c(t) \approx \sum_{i=1}^{4} \kappa_c^{(i)} t^i / i!$;
5) Calculate saddle-point by $\partial K_c(t)/\partial t = 0$;
6) Model the PDF of class c by SPA,
$p_c(f) \approx \exp(K_c(s_c)) \Big/ \sqrt{\partial^2 K_c(s_c)/\partial t^2}$;
Output: PDF of each class.

Algorithm 3. SPA method for open-set recognition (testing).

Input: target samples feature $G_f(x_i^T)$ from feature extractor G_f; per-class mean features m_c, distribution parameters (τ_c, κ_c, λ_c), candidate threshold set […, ε-2γ, ε-γ, ε, $\varepsilon + \gamma$, $\varepsilon + 2\gamma$, …].
1) For each sample, compute the distance $d_i^{T,c} = \left\| s_i^{T,c} - m_c \right\|$;
2) Compute the CDF of $d_i^{T,c}$,
$\gamma_c(d_i^{T,c}) = 1 - \exp\left(-\left\| d_i^{T,c} - \tau_c \right\| \Big/ \lambda_c\right)^{\kappa_c}$
3) **for** $\{\Omega_c\} = [\ldots, \varepsilon_c\text{-}2\gamma, \varepsilon_c\text{-}\gamma, \varepsilon_c, \varepsilon_c + \gamma, \varepsilon_c + 2\gamma, \ldots]$ **do**
1) Compute the H-score under $\{\Omega_c\}$
2) Update the $\{\Omega_c^*\}$ under the highest H-score,
$\{\Omega_c^*\} = \arg\max_{\Omega \in \{\Omega_c\}} H\text{-score}(\{\Omega_c\})$
end for
Output: unknown sample; if $\gamma_c(d_i^{T,c}) > \{\Omega_c^*\}$ for any class c.

collection. When abnormal vibrations occur, this phase data will be analyzed for fault diagnosis. However, data exhibiting indistinct fault characteristics reduce the diagnostic accuracy for distribution box D6. To diagnose faults in distribution box D6, high-quality data from distribution box D1 are used as the source domain, while data from distribution box D6 are used as the target domain. Additionally, to evaluate the robustness of the proposed model, data from distribution box D6 are collected under three working conditions, each corresponding to the production of a different type of steel. These working conditions result in variations in operating parameters such as load and speed, while the fault categories remain unchanged. Reviewing faults historical, four common faults shared by Distribution Box D1 and Distribution Box D6 were selected, and one rare fault from Distribution Box D6 was designated as the unknown fault. The common faults are lubricant leakage, gear surface cracks, pitting on the tooth flank, and spalling on the gear tooth surface, while the rare fault is flat-headed sleeve tooth cracks. The appearances of these five faults are shown in figure 9.

The rotating components within the distribution box generate periodic mechanical vibration pulses during operation. These vibration signals can be effectively captured by accelerometer-based vibration sensors. Typically, triaxial accelerometers are mounted on the casing to collect operational data. The installation location of the sensor is illustrated in figure 10.

Before training the model, the data is preprocessed and organized to ensure learning efficiency and accuracy. The segmentation results shown in table 1, where the sample length is set to 512 points based on the 1250 Hz sensor frequency and approximate 600 rpm roller speed. The data from distribution box D1 serves as the source domain data, while data from distribution box D6 is used as the target domain data. Additionally, to enhance the reliability of performance evaluations beyond a static split, random cross-validation is employed. In the experimental section, the proposed model and each comparative model undergo five iterations, with the results reported as the best outcome across these runs.

To visualize the fault signals, the first 16 000 data points are plotted in a time domain waveform at a response frequency of 1250 Hz. The waveforms are shown in figure 11.

4.2. Evaluation index

To evaluate the diagnostic performance of the proposed model, three evaluation protocols are introduced. Diagnosis capability of known fault and unknown fault described as:

$$\begin{cases} \text{Acc}_{\text{known}} = \frac{M_s}{N_s} \\ \text{Acc}_{\text{unknown}} = \frac{M_u}{N_u} \end{cases}, \tag{28}$$

where M_s is the number of known-class samples, and N_s is the total number of samples in all known classes. Similarly, M_u denotes the number of unknown class samples, and N_u represents the total number of samples in all unknown classes.

To evaluate the capability of model in distinguishing unknown faults, the H-score index is introduced [37], defined as:

$$H\text{-score} = 2 \cdot \frac{\text{Acc}_{\text{known}} \cdot \text{Acc}_{\text{unknown}}}{\text{Acc}_{\text{known}} + \text{Acc}_{\text{unknown}}}. \tag{29}$$

Finally, overall accuracy is employed to measure the comprehensive fault diagnosis capability of model, calculated as follows:

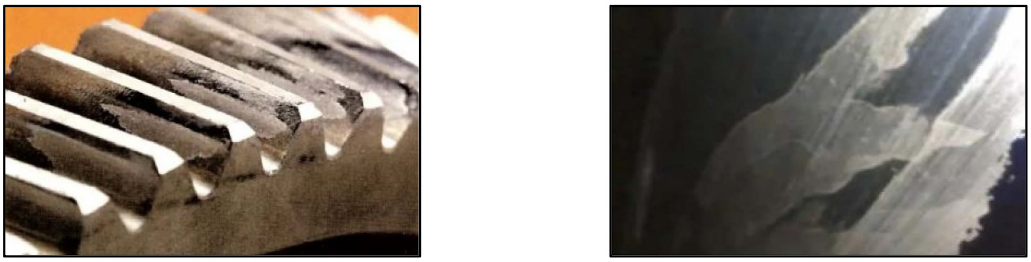
$$\text{accuracy} = \frac{M_s + M_u}{N_s + N_u}. \tag{30}$$

4.3. Results comparison

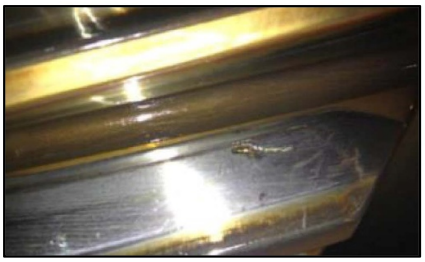
To confirm the practical performance of the proposed model in engineering, three different working condition signals are used in the experiment. The experiments are conducted on a computer equipped with an Intel Core i7-9750H processor, an Nvidia GeForce GTX 1660 Ti graphics card, and 16 GB of memory. The model is developed using Python 3.7 and PyTorch. The relevant hyperparameters are selected through a grid search, and the details are provided in table 2.

The testing history for the target domain (distribution box D6) is shown in figure 12. Due to the nature of adversarial

(a). Lubricant leakage (b). Gear surface crack



(c). Pitting of tooth flank (d). Gear tooth surface spalling



(e). Flat-headed sleeve tooth crack

Figure 9. Appearance of the faults.

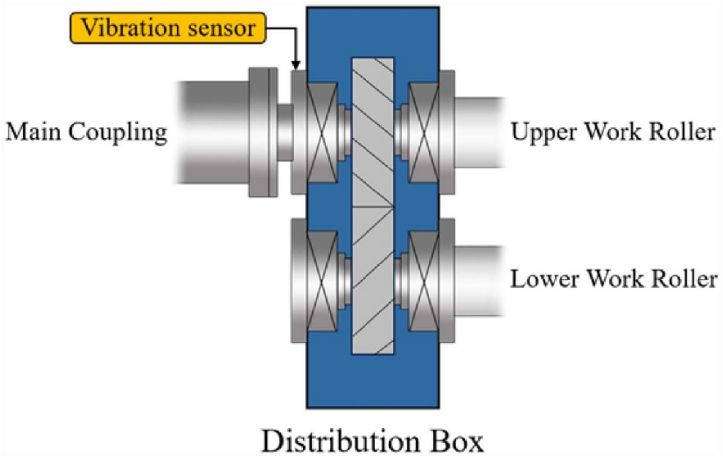


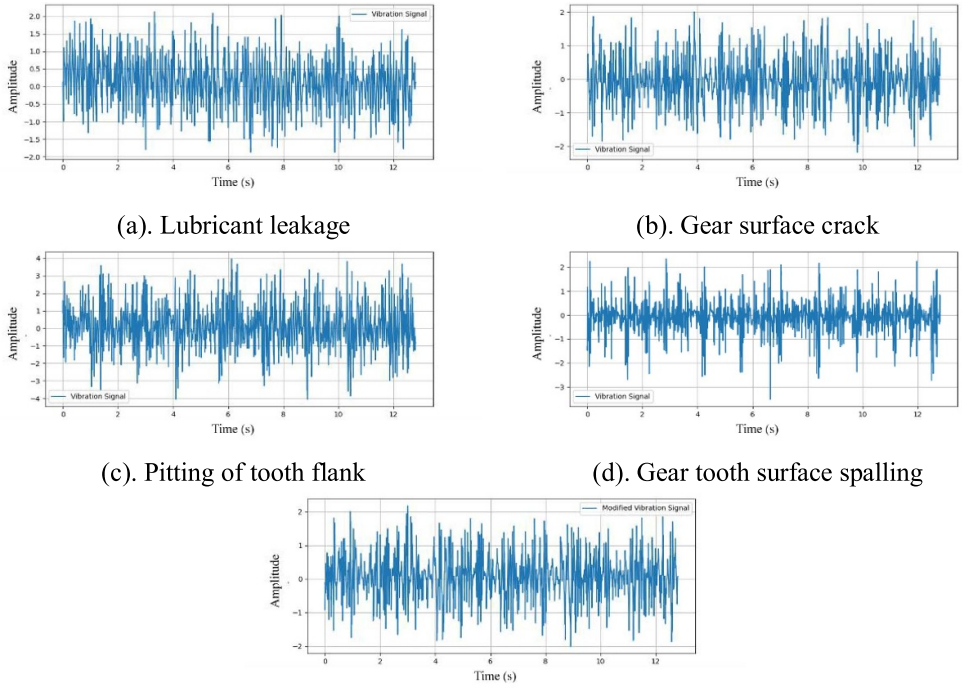
Figure 10. Vibration acceleration sensor installation location.

learning and the influence of open-set conditions, the accuracy during the learning process exhibits fluctuations. But after 125 epochs, the testing accuracy for all three working conditions relatively stabilizes and consistently surpasses 90%. The final accuracies of three conditions are 91.30%, 95.03% and 96.07% respectively.

To further verify the advantages of the proposed model in feature extractor and unknown fault identification, an ablation experiment is conducted to assess the contribution of each component. Several related open-set methods are introduced for comparison to assess the effectiveness of the proposed model. The comparison includes the OpenMax approach

Table 1. Segmentation of fault samples.

Fault type	Sample length	Source domain(D1)	Target domain(D6)	
		Training/testing	Training	Testing
Lubricant leakage	512 * 1	225/97	225	97
Gear surface crack	512 * 1	225/97	225	97
Pitting of tooth flank	512 * 1	225/97	225	97
Gear tooth surface spalling	512 * 1	225/97	225	97
Flat−headed sleeve tooth crack	512 * 1	0/0	0	97



(a). Lubricant leakage (b). Gear surface crack

(c). Pitting of tooth flank (d). Gear tooth surface spalling

(e). Flat−headed sleeve tooth crack

Figure 11. Waveform of the fault signals in the time domain.

Table 2. Values of hyperparameters.

Hyperparameters	Values	Description
Epoch	150	Number of training epochs
Batch size	64	Number of samples per batch
Learning rate	0.0001	Initial learning rate for optimizer
Optimizer	Adam	Optimization algorithm
Gamma	0.1	Learning rate decay factor
Decay step size	25	Step size for learning rate decay
Weight decay	0.0001	L2 regularization factor to prevent overfitting
Candidate thresholds	0.01	Interval of the candidate thresholds

[42], the EVT-based Weibull distribution method [23], and improved OpenMax approach [23]. Both method [42] and method [30] combine Softmax values with a Weibull distribution when handling open-set samples. In contrast, method [23] and the proposed approach directly use the source data distribution to distinguish open-set samples. In addition, the feature extractor in DANN plays a crucial role in addressing cross-domain and open-set challenges. We also introduce the depthwise separable CNN-BiLSTM method [5], Depthwise separable convolution model [43] and Standard DANN network model [44] for comparison to assess the performance of the proposed feature extractor. The details of the models for comparison are presented in table 3.

The performance comparison of fault diagnosis models is shown in table 4. By comparing accuracy and H-score, the proposed model has better diagnostic performance. For a detailed analysis, the results from condition 1, which show the poorest performance, are selected for visualization.

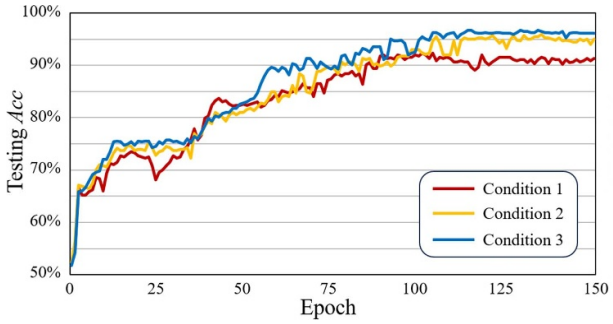


Figure 12. Training history of target domain (Distribution Box D6).

Table 3. Comparison methods information.

Model	Feature extractor	Unknown fault identification
M1	Proposed feature extractor	OpenMax [42]
M2	Proposed feature extractor	EVT [23]
M3	Proposed feature extractor	Improved OpenMax [30]
M4	Depthwise separable convolution with Bi-LSTM [5]	SPA method with adaptive threshold
M5	Depthwise separable convolution [43]	SPA method with adaptive threshold
M6	Domain-Adversarial Neural Networks [44]	SPA method with adaptive threshold
Proposed	Depthwise separable convolution with GCN	SPA method with adaptive threshold

Table 4. Comparison results of different methods.

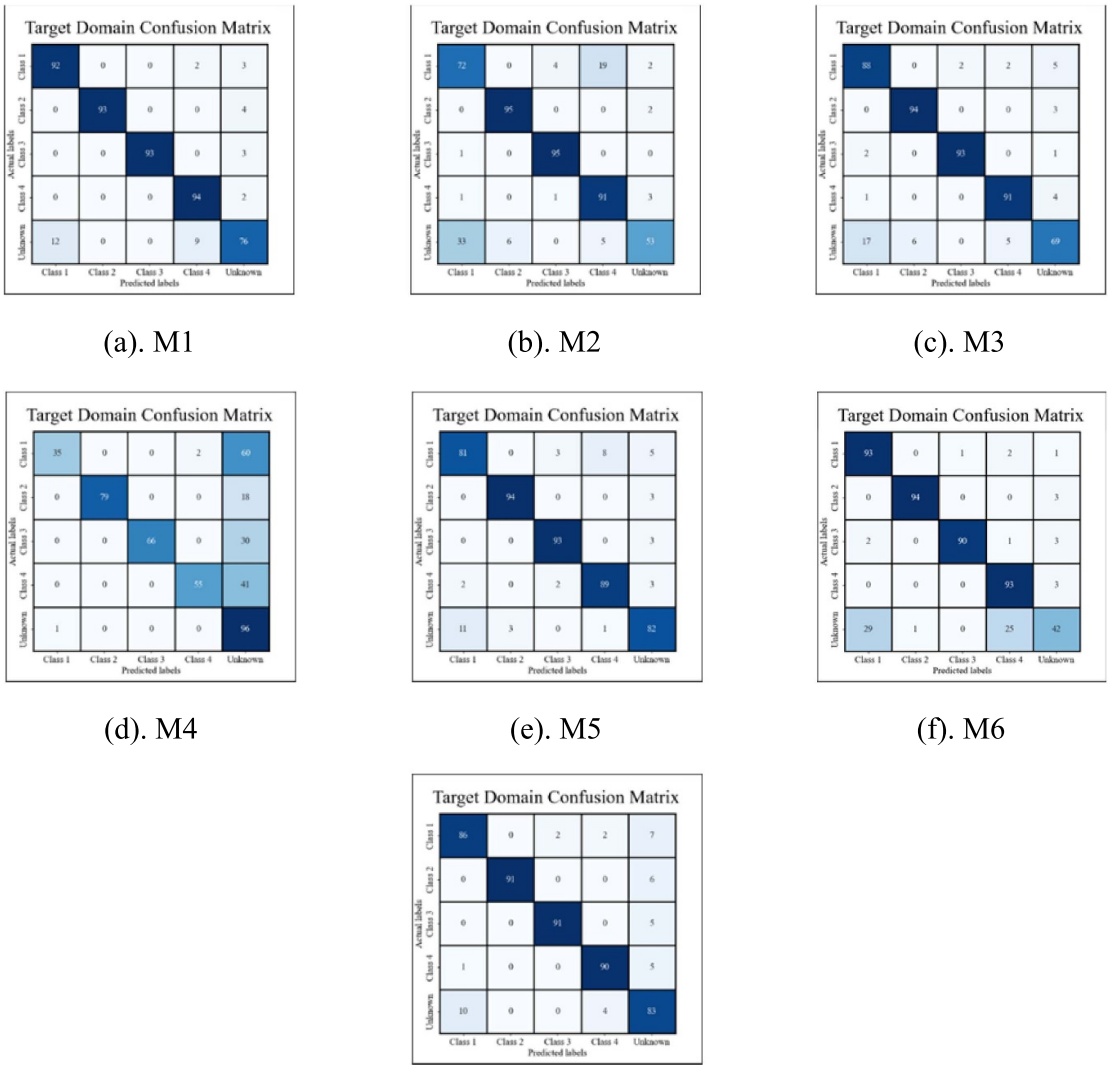
	Evaluation index for different working condition of distribution box D6							
	Condition 1		Condition 2		Condition 3		Average	
Model	Accuracy	*H*-score	Accuracy	*H*-score	Accuracy	*H*-score	Accuracy	*H*-score
M1	88.20%	86.43%	94.00%	92.31%	92.75%	85.90%	91.65%	88.21%
M2	84.06%	68.41%	91.72%	85.90%	94.20%	92.84%	89.99%	82.38%
M3	90.06%	81.28%	94.82%	93.63%	95.24%	95.09%	93.37%	90.00%
M4	68.53%	75.39%	85.09%	80.07%	80.33%	79.97%	77.98%	78.48%
M5	90.89%	88.33%	94.20%	92.42%	95.45%	94.42%	93.51%	91.72%
M6	85.30%	84.22%	89.13%	90.12%	89.88%	87.66%	88.10%	87.33%
Proposed	91.30%	89.01%	95.03%	91.10%	96.07%	93.94%	94.13%	91.35%

The confusion matrices of all models under condition 1 are presented in figure 13. From figures 13(a)–(g), it can be observed that the classification of faults within the known classes is generally accurate. However, there is a significant difference in classification performance between known and unknown classes. The results in figures 13(a)–(c), (f) and figure 13(g) are relatively similar, as the same unknown class samples are frequently misclassified into known classes. In contrast, figures 13(d) and (e) show notable differences, where many known class samples are misclassified into the unknown class. According to the results, the proposed feature extraction module demonstrates superior performance in extracting features.

Combined with table 4 and figures 13(a)–(c), (f) and 13(g) for a further analysis, the proposed model demonstrates a balanced capability in distinguishing known classes from unknown classes. And it achieves good classification performance within known classes. The key difference among the four models are their modules for handling the open-set scenario. Therefore, the proposed module for handling open-set scenarios demonstrates superior performance in addressing this challenge.

Apart from evaluating diagnostic accuracy, the feature representations of different models are analyzed by visualization techniques to provide an intuitive understanding. The t-distributed stochastic neighbor embedding (t-SNE) method is employed to explore high-dimensional data representations by mapping deep features onto a 2D feature space. In the low-dimensional space, the feature extracted can be observe intuitively. Condition 1 data is used for analysis, and the feature representations extracted by the feature extractor are shown in figure 14. For M4 and M6, overlap between known samples and unknown samples is observed, leading to higher misclassification rates. In comparison, for M1-3, M5 and Proposed model, the clustering effect of known classes is weaker than in previous models. However, the larger inter-class margin allows the classifier to establish clearer classification boundaries, enhancing the distinction between known classes. These results demonstrate that the proposed model effectively separates the features of each class.

To evaluate the effectiveness of the saddlepoint approximation, we visualize the feature distribution of the source domain data. As illustrated in figure 15, the red curve



(a). M1 (b). M2 (c). M3

(d). M4 (e). M5 (f). M6

(g). Proposed

Figure 13. Confusion matrixes of different methods.

denotes the distribution modeled by the saddle point approximation.

Furthermore, to assess the computational efficiency of various methods in practical applications, the training time is measured as the time required to train a deep model for 150 epochs. The testing time is defined as the time required to predict all test samples during the final epoch. The visualization results are shown in figure 16. Considering that faults in the distribution box leads to downtime and economic losses, a slight increase in training of model is justified.

5. Conclusions

To address the issues of data exhibiting indistinct fault characteristics and unknown faults in distribution boxes of finishing mill, this paper proposes a cross-domain open-set fault diagnosis model. The model employs DANN as the basic framework and incorporates open-set method. To verify the effectiveness of the proposed model, experiments are conducted on historical fault data from distribution boxes. The results show that the proposed method can be applied in engineering.

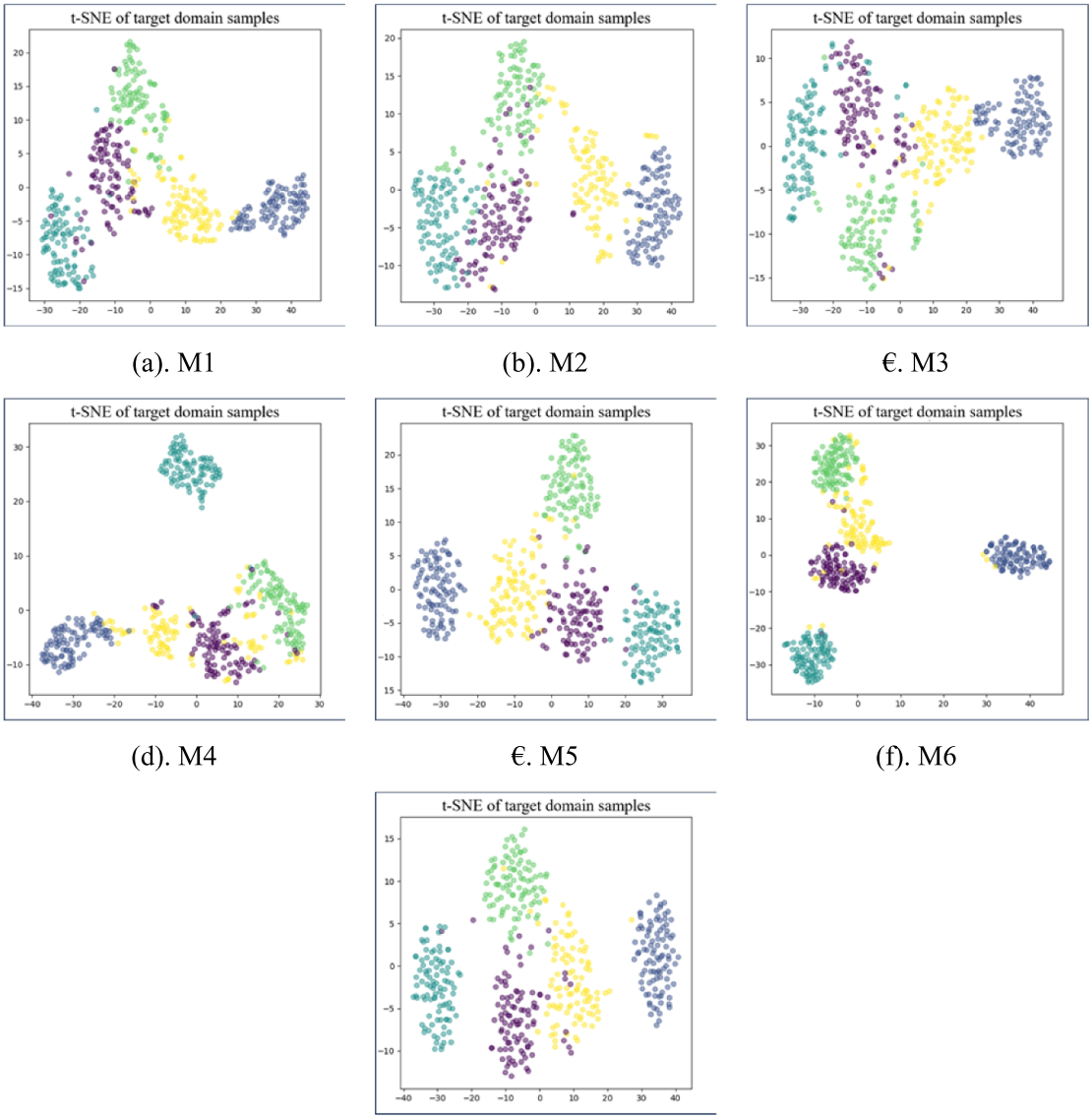
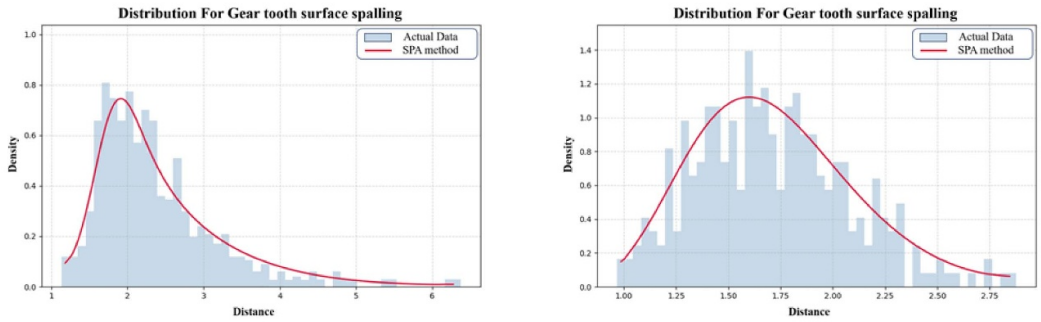


Figure 14. t-SNE visualization results.

The main contributions of the proposed model lie in the design of the feature extractor and method to identify unknown faults. In DANN, a feature extractor integrates depthwise separable convolution with GCN is proposed. This structure overcomes the channel independence inherent in traditional depthwise separable convolution. To identify unknown faults, the SPA method is employed to establish the distributions. Unlike traditional methods, the SPA method establishes distributions directly from the data, without relying on distribution assumptions and distribution formulas.

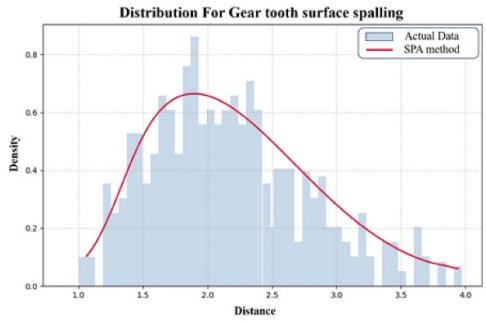
To evaluate the performance of the proposed model, we compared the proposed model with several existing diagnostic models. The results show that the proposed feature extractor performs better at features extracting. Feature quality affects both adversarial learning and the establishment of distributions. Additionally, the SPA method achieves higher accuracy in establishing distributions for identifying unknown faults. The comparison results indicate that the proposed model achieves the highest average accuracy (94.13%) and the highest average H-score (91.35%). Even under the worst data-quality conditions, the accuracy and H-score remain close to 90%, reaching 91.30% and 89.01%, respectively. This demonstrates that the proposed model can maintain reliable diagnostic performance even when the fault data has weak fault characteristics.

However, compound faults may occur in the distribution box, and it remains unknown whether the signals show identifiable sub-fault features. Future research will focus on classifying these compound faults as known sub-faults rather than mistakenly treating them as new faults.

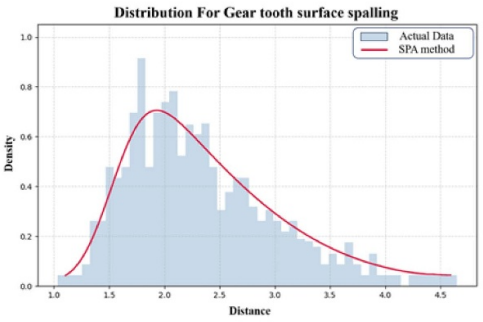


(a). Lubricant leakage

(b). Gear surface crack



(c). Pitting of tooth flank



(d). Gear tooth surface spalling

Figure 15. Distribution visualization results.

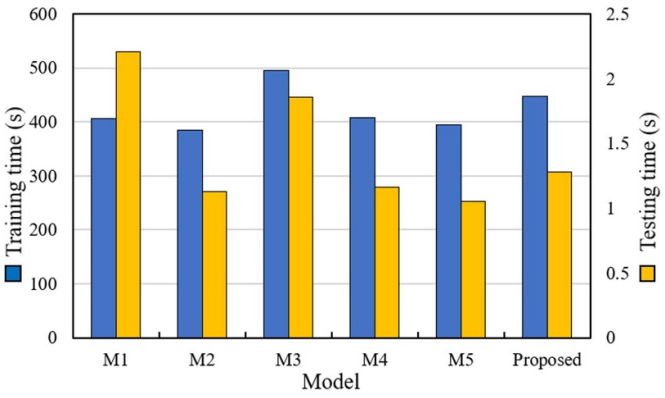


Figure 16. Comparison of the training time and testing time.

Data availability

The datasets during the current study are available from the corresponding author upon reasonable request.

Conflict of interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding statement

This project is supported by the Fundamental Research Fund for the Central Universities (2232023D-17) and the Shanghai Scientific Research project (No. 22511103604).

ORCID iD

Tangbin Xia 0000-0001-9121-1716

References

[1] Zhao S, Bao L, Hou C, Bai Y and Yu Y 2024 Multi-source domain adversarial graph convolutional networks for rolling mill health states diagnosis under variable working conditions *Struct. Health Monit.* **23** 3505–24

[2] Yildiz S K, Forbes J F, Huang B, Zhang Y, Wang F, Vaculik V and Dudzic M 2009 Dynamic modelling and simulation of a hot strip finishing mill *Appl. Math. Modelling.* **33** 3208–25

[3] Zhou X and Ben X 2023 Maintenance modelling for work rolls in hot finishing mill group with constraint of thermal character *Int. J. Prod. Res.* **62** 1846–61

[4] He H, Shao J, Wang X, Yang Q, Liu Y, Xu D and Sun Y-Z 2021 Research and application of approximate rectangular section control technology in hot strip mills *J. Iron Steel Res. Int.* **28** 279–90

[5] Zhang W, He J, Ma C, Gao W and Li G 2025 Compound fault diagnosis method of rotating machinery using multi-view multi-label feature selection based on label compression and local label correlation *Adv. Eng. Inf.* **65** 103310

[6] Li W, Liu D, Li Y, Hou M, Liu J, Zhao Z, Guo A, Zhao H and Deng W 2025 Fault diagnosis using variational autoencoder GAN and focal loss CNN under unbalanced data *Struct. Health Monit.* **24** 1859–72

[7] Zhao C, Sun J, Lin S and Peng Y 2021 Fault diagnosis method for rolling mill multi row bearings based on AMVMD-MC1DCNN under unbalanced dataset *Sensor* **21** 5494

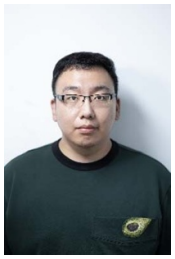
[8] Yu Y, Shi P, Tian J, Xu X and Hua C 2023 Rolling mill health states diagnosing method based on multi−sensor

information fusion and improved DBNs under limited datasets *ISA Trans.* **134** 529–47
[9] Shi P, Gao H, Yu Y, Xu X and Han D 2022 Intelligent fault diagnosis of rolling mills based on dual attention−guided deep learning method under imbalanced data conditions *Measurement* **204** 111993
[10] Guo Y, Zhou D, Chen H, Yue X and Cheng Y 2024 Fault intelligent diagnosis for distribution box in hot rolling based on depthwise separable convolution and Bi−LSTM *Processes* **12** 1999
[11] Peng R, Zhang X and Shi P 2022 Bearing fault diagnosis of hot-rolling mill utilizing intelligent optimized self-adaptive deep belief network with limited samples *Sensors* **22** 7815
[12] Li X, Jin W, Xu X and Yang H 2022 A domain-adversarial multi-graph convolutional network for unsupervised domain adaptation rolling bearing fault diagnosis *Symmetry* **14** 2654
[13] Hou D, Shi P, Chen J and Shi P 2024 Improved GNN based on graph−transformer: a new framework for rolling mill bearing fault diagnosis *Trans. Inst. Meas. Control.* **46** 2804–15
[14] Wang C, Xin C and Xu Z 2021 A novel deep metric learning model for imbalanced fault diagnosis and toward open-set classification *Knowl.-Based Syst.* **220** 106925
[15] Wang L, Gao Y, Li X and Gao L 2024 Self-supervised-enabled open-set cross-domain fault diagnosis method for rotating machinery *IEEE Trans. Ind. Inform.* **20** 10314–24
[16] Lundgren A and Jung D 2022 Data-driven fault diagnosis analysis and open-set classification of time-series data *Control Eng. Pract.* **121** 105006
[17] Razavi-Far R, Hallaji E, Saif M and Ditzler G 2019 A novelty detector and extreme verification latency model for nonstationary environments *IEEE Trans. Ind. Electron.* **66** 561–70
[18] Tian Y, Wang Z, Zhang L, Lu C and Ma J 2018 A subspace learning-based feature fusion and open-set fault diagnosis approach for machinery components *Adv. Eng. Inf.* **36** 194–206
[19] Peng P, Lu J, Xie T, Tao S, Wang H and Zhang H 2023 Open-set fault diagnosis via supervised contrastive learning with negative out-of-distribution data augmentation *IEEE Trans. Ind. Inform.* **19** 2463–73
[20] Jung D 2020 Data-driven open-set fault classification of residual data using Bayesian filtering *IEEE Trans. Control. Syst. Technol.* **28** 2045–52
[21] Zhang X, Wang J, Han B, Zhang Z, Yan Z, Jia M and Guo L 2022 Feature distance-based deep prototype network for few-shot fault diagnosis under open-set domain adaptation scenario *Measurement* **201** 111522
[22] Cui Z, Lu Y, Yan X and Cui S 2024 Compound fault diagnosis of diesel engines by combining generative adversarial networks and transfer learning *Expert Syst. Appl.* **251** 123969
[23] Wang H, Xu Z, Tong X and Song L 2023 Cross-domain open set fault diagnosis based on weighted domain adaptation with double classifiers *Sensors* **23** 2137
[24] Zhu J, Huang C, Shen C and Shen Y 2022 Cross-domain open-set machinery fault diagnosis based on adversarial network with multiple auxiliary classifiers *IEEE Trans. Ind. Inform.* **18** 8077–86
[25] Zhao C and Shen W 2022 Adaptive open set domain generalization network: learning to diagnose unknown faults under unknown working conditions *Reliab. Eng. Syst. Saf.* **226** 108672
[26] Chen Z, Liao Y, Li J, Huang R, Xu L, Jin G and Li W 2023 A multi-source weighted deep transfer network for open-set fault diagnosis of rotary machinery *IEEE Trans. Cybern.* **53** 1982–93
[27] Zhang Z, Nie G, Shao M, Li L, Zhou J and Shao S 2023 Multi-sample-distances-fusion- and generalized-Pareto-distribution-based open-set fault diagnosis of rolling bearing *Nonlinear Dyn.* **111** 11407–28
[28] Wang X, Zhang H, Qiao X, Ma K, Tao S *et al* 2024 Generalized out-of-distribution fault diagnosis (GOOFD) via internal contrastive learning *IEEE Trans. Ind. Inform.* **20** 9987-9996
[29] Yu X *et al* 2022 Deep-learning-based open set fault diagnosis by extreme value theory *IEEE Trans. Ind. Inform.* **18** 185–96
[30] Zhang B, Zhou C, Li W, Ji S, Li H, Tong Z and Ng S-K 2022 Intelligent bearing fault diagnosis based on open set convolutional neural network *Mathematics* **10** 3953
[31] Gao Y, Ma A J, Gao Y, Wang J and Pan Y 2020 Adversarial open set domain adaptation via progressive selection of transferable target samples *Neurocomputing* **410** 174–84
[32] Fan Y, Guo L, Sun Y, Luo T, Hou K *et al* 2025 Wasserstein bi-classifier adversarial learning network for machinery fault diagnostics *Struct. Health Monit.* **24** 3363–81
[33] Verstraete D B, Droguett E L, Meruane V, Modarres M and Ferrada A 2019 Deep semi-supervised generative adversarial fault diagnostics of rolling element bearings *Struct. Health Monit.* **19** 390–411
[34] Mao W, Liu Y, Ding L, Safian A and Liang X 2021 A new structured domain adversarial neural network for transfer fault diagnosis of rolling bearings under different working conditions *IEEE Trans. Instrum. Meas.* **70** 1-13
[35] Zhou D, Pan E, Zhang X and Zhang Y 2020 Dynamic model-based saddle-point approximation for reliability and reliability-based sensitivity analysis *Reliab. Eng. Syst. Saf.* **201** 106972
[36] Huang B and Du X 2006 Uncertainty analysis by dimension reduction integration and saddlepoint approximations *J. Mech. Des.* **128** 26
[37] He J, Huang W, Liu Y, Qian C, Ma C, Gao W and Jin X 2025 Data imbalance fault diagnosis method based on an ensemble multi-scale convolutional attention network *Mech. Syst. Signal Process.* **236** 112934
[38] Zhou D, Zhang X and Zhang Y 2016 Dynamic reliability analysis for planetary gear system in shearer mechanisms *Mech. Mach. Theory* **105** 244–59
[39] Wang C, Li X, Yuan P, Su K, Xie Z and Wang J 2025 An integrated approach for mechanical fault diagnosis using maximum mean square discrepancy representation and CNN-based mixed information fusion *Struct. Health Monit.* **24** 3674–87
[40] Liu T, Zhang Y, Huang G, Liu C and Jiang D 2025 Fault identification of aero-engine central rod-fastened rotor system based on multi-source information fusion and graph convolutional network *Struct. Health Monit.* **0**
[41] Xiang Q, Xiang Q, Lai J, Lei L, Song Y, He J and Li R 2024 Quadruplet depth-wise separable fusion convolution neural network for ballistic target recognition with limited samples *Expert Syst. Appl.* **235** 121182
[42] Bendale A and Boult T E 2016 Towards open set deep networks *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 1563–72
[43] Qin Y, Zhang T, Qian Q and Mao Y 2024 Large model for rotating machine fault diagnosis based on a dense connection network with depthwise separable convolution *IEEE Trans. Instrum. Meas.* **73** 3520512
[44] Ganin Y, Ustinova U, Ajakan H, Germain P, Larochelle H, Laviolette F, Marchand M and Lempitsky V 2016 Domain-adversarial training of neural networks (arXiv:1505.07818)

Yonglin Guo received the BE degree from the Shandong Jianzhu University, Jinan, China, in 2023. He is currently pursuing his Master's degree at the School of Mechanical Engineering, Donghua University, Shanghai, China. His research interest is fault diagnosis.

Shihao Li is a Master's candidate in the School of Mechanical Engineering at Dong Hua University, Shanghai, China. His research interests focus on reliability systems and failure probability calculations.

Tangbin Xia (Member, IEEE) received the PhD degree in mechanical engineering (industrial engineering) from Shanghai Jiao Tong University, Shanghai, China, in 2014. He is currently a Professor and the Vice Dean of the School of Mechanical Engineering, Shanghai Jiao Tong University. His research interests include intelligent maintenance systems, prognostics and health management, and advanced manufacturing. Dr Xia is a member of IISE, ASME, and INFORMS.



Di Zhou received his PhD degree in Mechanical Engineering and Automation from Northeastern University, China. He is currently an Associate Professor in the School of Mechanical Engineering at Donghua University. His research interests include mechanical reliability analysis and mechanical dynamics.

Ershun Pan received the BS and MS degrees in mechanical design and manufacturing from Northeastern University, Shenyang, China, in 1997, and the PhD degree in mechanical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2000.