



## Pseudo-fault data enhanced relation network for fault detection and localization in train transmission systems

Zhixu Duan <sup>a</sup>, Ruoxin Liu <sup>a</sup>, Zuoyi Chen <sup>a,b,\*</sup>, Hong-Zhong Huang <sup>a,b</sup>

<sup>a</sup> School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, China

<sup>b</sup> Center for System Reliability and Safety, University of Electronic Science and Technology of China, Chengdu, 611731, China



### ARTICLE INFO

#### Keywords:

Train transmission systems  
Fault detection and localization  
Zero faulty data  
Relation network  
Pseudo-fault data

### ABSTRACT

Fault detection and localization is a critical technology for ensuring the safe and efficient operation of train transmission systems (TTS). However, due to the frequent scarcity or absence of fault data, achieving precise detection and localization faces many challenges. To address this challenge, this paper proposes a novel pseudo-fault data enhanced relation network (PE-ERN) for fault detection and localization in TTS. The PE-ERN method generates pseudo-fault data by combining fault information from similar external equipment and out-of-distribution data. This strategy enriches the training dataset, enabling model to extract intrinsic health state information from available health data. Additionally, a feature concatenation mechanism is developed to generate feature pairs by combining health state data with pseudo-fault data. This mechanism uncovers both health-unique and health-inherent attributes, which enhances the PE-ERN's ability to distinguish between fault and health states. Experimental results from subway TTS cases demonstrate that the proposed PE-ERN method outperforms existing fault detection techniques, achieving superior fault localization and detection accuracy across various fault modes, including single-fault, component-level compound-fault, and system-level compound-fault scenarios.

### 1. Introduction

As a critical component of modern transportation, railway transit has experienced remarkable advancements fueled by Industry 4.0 technologies (Zhao et al., 2025; Dou et al., 2025). With increasing operational speed and system complexities, prognostics and health management of railway systems, particularly train transmission systems (TTS), have emerged as critical concerns (Ding et al., 2024a; Gao et al., 2025; Liu et al., 2022). The reliability and safety of railway operations hinge on the condition of critical transmission components, such as traction motors, gearboxes, and axle boxes (Ding et al., 2024b; Chen et al., 2023a). Failures in these components can precipitate catastrophic consequences, underscoring the urgent need for continuous monitoring and real-time fault detection in TTS. By integrating smart sensors, the internet of things, and advanced data analytics, fault detection systems can significantly reduce downtime, mitigate economic losses, and enhance operational safety.

Deep learning techniques have become widely adopted for TTS fault detection, with approaches like convolutional neural networks (CNN) (Wang et al., 2023; Yu and Zhao, 2020), residual networks (ResNet)

(Chen et al., 2023b; Wang et al., 2024; Mao et al., 2024), and graph neural networks (GNN) (Mao et al., 2024; Gao et al., 2024; Liang et al., 2024) showing promising results. However, the efficacy of these approaches typically relies on the availability of substantial labeled fault data, which is often scarce in real-world industrial settings (Wang et al., 2025a; Xu et al., 2025; Li et al., 2024). While health data is relatively abundant, the shortage of fault data poses a significant barrier to the practical deployment of deep learning models (Zhao et al., 2024; Wang et al., 2025b). Moreover, operating TTS in a faulty condition is generally unacceptable, necessitating proactive monitoring to prevent fault escalation. The scarcity of fault data remains a primary challenge in leveraging deep learning for effective TTS fault detection.

To address the challenge of fault detection in the absence of fault data, transfer learning (TL) has emerged as a potential solution, transferring from labeled data in a source domain to an unlabeled target domain (Wang et al., 2025c; Chen et al., 2025a; Zhang et al., 2024). TL enables the transfer of knowledge from a labeled source domain to an unlabeled target domain and has shown success in TTS fault detection and diagnosis, particularly when fault data is simulated in controlled experimental setups. (Cao et al., 2022; Wang et al., 2025d; Guo et al.,

\* Corresponding author. School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, China.  
E-mail address: [zuoyichen@uestc.edu.cn](mailto:zuoyichen@uestc.edu.cn) (Z. Chen).

2019). However, the effectiveness of TL is often constrained by the degree of similarity between the data distributions of the source and target domains. Mechanical differences across machines frequently result in substantial distributional disparities, requiring time-intensive model fine-tuning for each unique target system.

When fault samples are unavailable, fault detection can be reframed as a binary classification task: distinguishing between healthy and faulty states. Recent research has developed fault detection methods that rely solely on health state data, identifying faults by assessing the similarity between health data and data from unknown states (Chen et al., 2022, 2025b; Ma et al., 2025). Fault detection techniques like Autoencoders (Jiang et al., 2017; Liu et al., 2018), long short-term memory networks (Lei et al., 2019; Xiang et al., 2021), and graph neural networks (Ma et al., 2023; Chen et al., 2023c) perform nonlinear mappings to correlate inputs and outputs through health state data. Fault states are identified by evaluating the feature similarities between the fault and health states. However, Fault detection models trained exclusively on health data often struggle to create robust nonlinear mappings, rendering them overly sensitive to normal variations in data, especially in the presence of noise or unexpected fluctuations in health signals. Additionally, most existing fault detection models focus solely on determining whether a system is healthy or faulty, without pinpointing the specific location of the fault. In industrial contexts, both detecting and localizing faults are critical for devising effective maintenance and repair strategies.

This paper proposes a novel pseudo-fault data enhanced relation network (PE-ERN) tailored for fault detection and localization in TTS, particularly when actual fault data is absent. The method addresses the challenge of zero-fault-data scenarios by generating pseudo-fault data, enabling robust fault detection. Given the complex interactions among adjacent TTS components—such as traction motors, gearboxes, and axle boxes—that often hinder accurate fault localization, PE-ERN employs a dedicated relation network (ERN) for each key component.

These networks construct feature pairs by combining pseudo-fault data with health state data, allowing precise similarity assessments to determine each component's fault state. Extensive testing on subway TTS demonstrates PE-ERN's superior performance in detecting and localizing faults across diverse scenarios, including single faults, component-level compound faults, and system-level compound faults, underscoring its robustness and adaptability for real-world industrial applications.

The key contributions of this work are as follows: 1) A novel PE-ERN is proposed to tackle the challenge of detecting and locating faults in TTS under zero-fault-data conditions. By overcoming the reliance on scarce fault data, this method effectively handles complex fault scenarios, offering a significant advancement over traditional methods that struggle with data limitations; 2) To address the scarcity of fault data, a pseudo-fault dataset is constructed by leveraging fault data from external equipment with similar failure modes to the TTS. Additionally, out-of-distribution data (OODD) are generated by applying a soft Brownian offset to health state data. This approach enriches the training set, enhancing the ERN model's ability to capture intrinsic health state patterns while reducing over-sensitivity to normal variations in health data; 3) A feature concatenation mechanism (FCM) is produced to establish feature pairs between the health state and the pseudo-fault data, uncovering both health-unique and health-inherent attributes. This mechanism significantly improves the model's capacity to distinguish subtle differences between fault and health states, thereby enhancing detection accuracy and fault localization across diverse fault types.

The organization of the paper is as follows: Section 2 covers the foundational concepts of the relation network and soft Brownian offset. The methodology is detailed in Section 3. Section 4 discusses the analysis of experimental results, while Section 5 provides the conclusion of the study.

## 2. Preliminaries

### 2.1. Relation network

The relation network is originally developed for few-shot classification (Sung et al., 2018). In contrast to conventional deep neural networks, which depend on extensive datasets, it is capable of achieving accurate classification with minimal training data. As shown in Fig. 1, the relation network consists of two key components: the feature embedding module  $f_\phi(\bullet)$  and the relation module  $g_\phi(\cdot)$ . Initially, two sets are first constructed: the support set  $x_s$  with labeled data and the query set  $x_q$  with unlabeled data. These sets are fed into the embedding module  $f_\phi(\bullet)$  to extract their features, resulting in the corresponding feature vectors:

$$g_s = f_\phi(x_s) \quad (1)$$

$$g_q = f_\phi(x_q) \quad (2)$$

where the feature vectors for the support set  $x_s$  and the query set  $x_q$  are denoted as  $g_s$  and  $g_q$ , respectively.  $\varphi$  represents the learning parameters of the embedding module  $f_\phi(\bullet)$ .

Next, the feature vectors  $g_s$  and  $g_q$  are combined through a concatenation operation  $C(\cdot, \cdot)$ , as follows:

$$g_{s,q} = C(g_s, g_q) = C(f_\phi(x_s), f_\phi(x_q)) \quad (3)$$

where  $g_{s,q}$  is the feature pairs by combining  $g_s$  and  $g_q$ . Finally, this feature pairs are passed into the relation module  $g_\phi(\cdot)$  to calculate the relation score:

$$r_{s,q} = g_\phi(g_{s,q}) = g_\phi(C(f_\phi(x_s), f_\phi(x_q))) \quad (4)$$

where  $r_{s,q}$  denotes the relation score between  $x_s$  and  $x_q$ .  $\phi$  is the parameter of the  $g_\phi(\cdot)$ .

During the training phase, the loss function of the relation network is calculated using the mean square error (MSE) as follows:

$$L = \sum_{s=1}^m \sum_{q=1}^n \left( r_{s,q} - r_{s,q}^{real} \right)^2 \quad (5)$$

where  $r_{s,q}$  represents the predicted relation score between  $x_s$  and  $x_q$ , and  $r_{s,q}^{real}$  is the true relation score. The true score is 1 when  $x_s$  and  $x_q$  belong to the same class, and 0 if they are from different classes.

### 2.2. Soft Brownian offset

Utilizing the Soft Brownian Offset method (Moller et al., 2021), which draws inspiration from the Gaussian hyperspherical offset approach, OODD is produced by projecting a data point  $x \in \mathcal{R}^d$  onto a hypersphere. The radius  $r$  for this projection follows a normal distribution  $r \sim \mathcal{N}(0, I_D)$ , with  $I_D$  being the identity matrix,  $\sigma$  representing the standard deviation. A vector  $v$  is drawn from the distribution  $\mathcal{N}(0, I_D)$ , and its magnitude is normalized to form a unit vector. In this way, out-of-distribution data (OODD) are evenly distributed on the hypersphere, as described by the following equation:

$$\gamma(\mu, \sigma) = rv|v|^{-1} + \sigma n \quad (6)$$

However, the Gaussian hyperspherical offset method assumes that the data points follow a normal distribution, which may limit its applicability. To overcome this, the soft Brownian offset method was introduced, which takes inspiration from Brownian motion. This method iteratively shifts a data point  $x \in X$  away from its starting position in the dataset  $X$ , seeking to distance it as much as possible from the original data distribution. The process starts by randomly picking a sample point  $x$  from the dataset. Over several iterations, this point is fine-tuned until

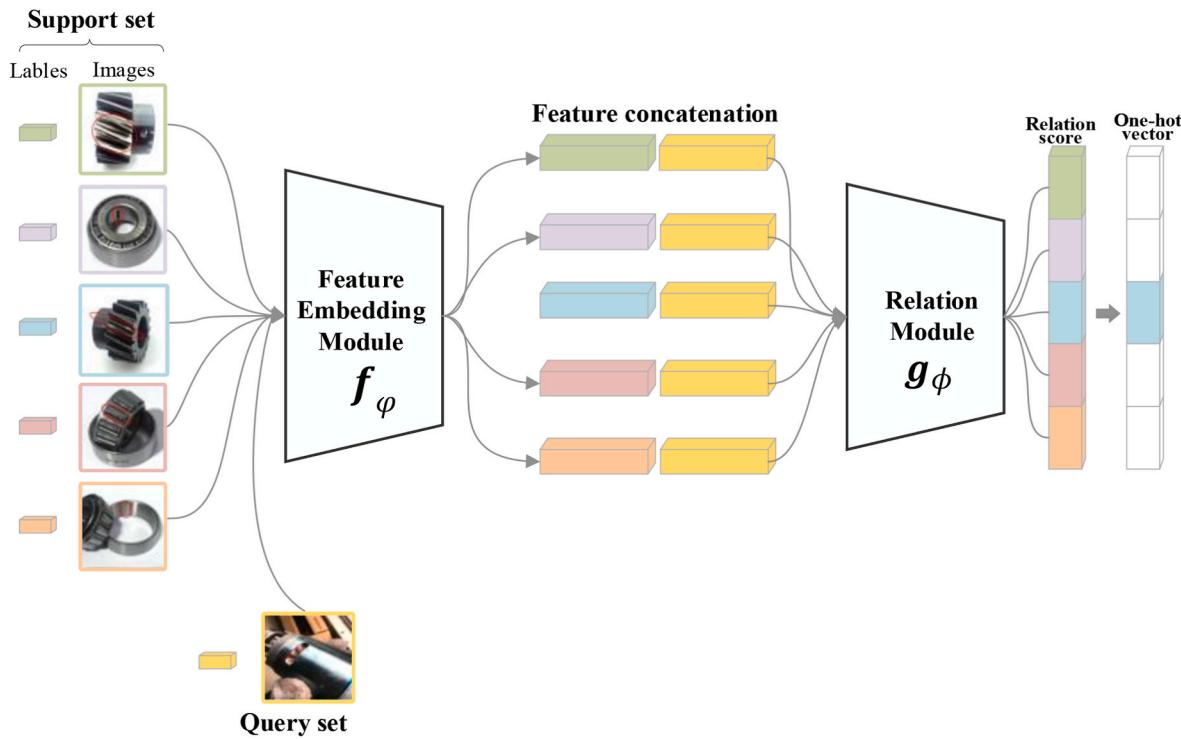


Fig. 1. Relation network structure.

its soft distance  $d^*$  exceeds a set threshold. The soft distance  $d^*$  is defined as  $d^* = d_{min}(\hat{x}, X)$ , where  $\hat{x}$  s the updated data point. To set the boundary for this adjustment, a rejection probability  $\rho$  is used, which ranges between 0 and 1, and is expressed by the following equation:

$$\rho(d^*, d^-, \sigma) = \left(1 + \exp\left(\frac{d^* + d^-}{\sigma d^- k}\right)\right)^{-1} \quad (7)$$

As the values of  $d^*$  and  $d^-$  increase, the probability  $\rho(d^*, d^-, \sigma)$  gradually decreases. To ensure a high probability when  $d^* = 0$ , the parameter  $k = 7$  is recommended, and  $\sigma$  should be adjusted between 0 and 1 to control the softness of the boundary. The probability  $\rho(d^*, d^-, \sigma)$  decreases as both  $d^*$  and  $d^-$  increase. To ensure that  $\rho$  remains high when  $d^* = 0$ , it is advised to set the parameter  $k = 7$  and modify the boundary softness using  $\sigma$ , is constrained within the interval  $[0, 1]$ . Fig. 2 illustrates how the soft Brownian offset method generates

data points that lie outside the conventional distribution. These OODD points are located close to the feature space of the health state data, making them suitable for use as fault data. This proximity enables them to enhance fault detection accuracy, especially when the training is based solely on health state data.

### 3. Methodology

#### 3.1. The proposed framework

A typical TTS comprises a traction motor, a driving gearbox, and two axle boxes. Faults in these components can pose serious safety risks, underscoring the need for advanced intelligent solutions capable of detecting and localizing such faults. However, the scarcity of fault samples and the complex interactions between adjacent components in the TTS make it challenging to accurately detect and localize faults using standalone fault detection models. To address these difficulties, this study proposes constructing independent fault detection networks for the traction motor, driving gearbox, and axle boxes, supported by pseudo-fault data to improve model performance. As illustrated in Fig. 3, the PF-ERN-based fault detection and localization framework is composed of two main stages: the training phase and the detection phase.

During the training phase, sensors are installed at each key component of the TTS to monitor its operating conditions. Historical health state data for the  $K$  key component is collected and represented as  $\mathcal{D}_H$  ( $\mathcal{D}_H^1, \mathcal{D}_H^2, \dots, \mathcal{D}_H^K$ ), with  $\mathcal{D}_H^K$  indicating the data pertaining to the  $K$ -th component. Additionally, a pseudo-fault database  $\mathcal{D}_F$  is constructed to enable ERN to fully extract hidden information from the health state. The pseudo-fault database  $\mathcal{D}_F$  comprises external equipment fault datasets  $\mathcal{D}_E$  and corresponding OODD  $\mathcal{D}_O$  ( $\mathcal{D}_O^1, \mathcal{D}_O^2, \dots, \mathcal{D}_O^K$ ) generated from the health data of each key component. Subsequently, corresponding ERN models are constructed for each key detection component based on the training dataset  $\mathcal{D}[\mathcal{D}_H^K; \mathcal{D}_F^K(\mathcal{D}_E, \mathcal{D}_O^K)]$ . The ERN primarily consists of a residual shrinkage feature extractor (RSFE), a feature concatenation mechanism (FCM), and a Kolmogorov-Arnold network

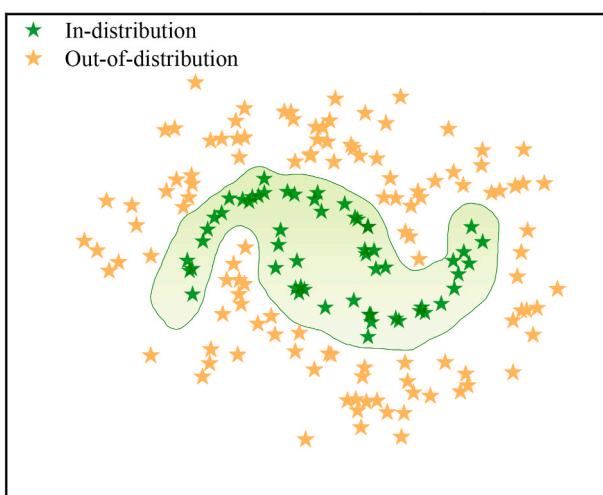


Fig. 2. Producing OODD within the latent space.

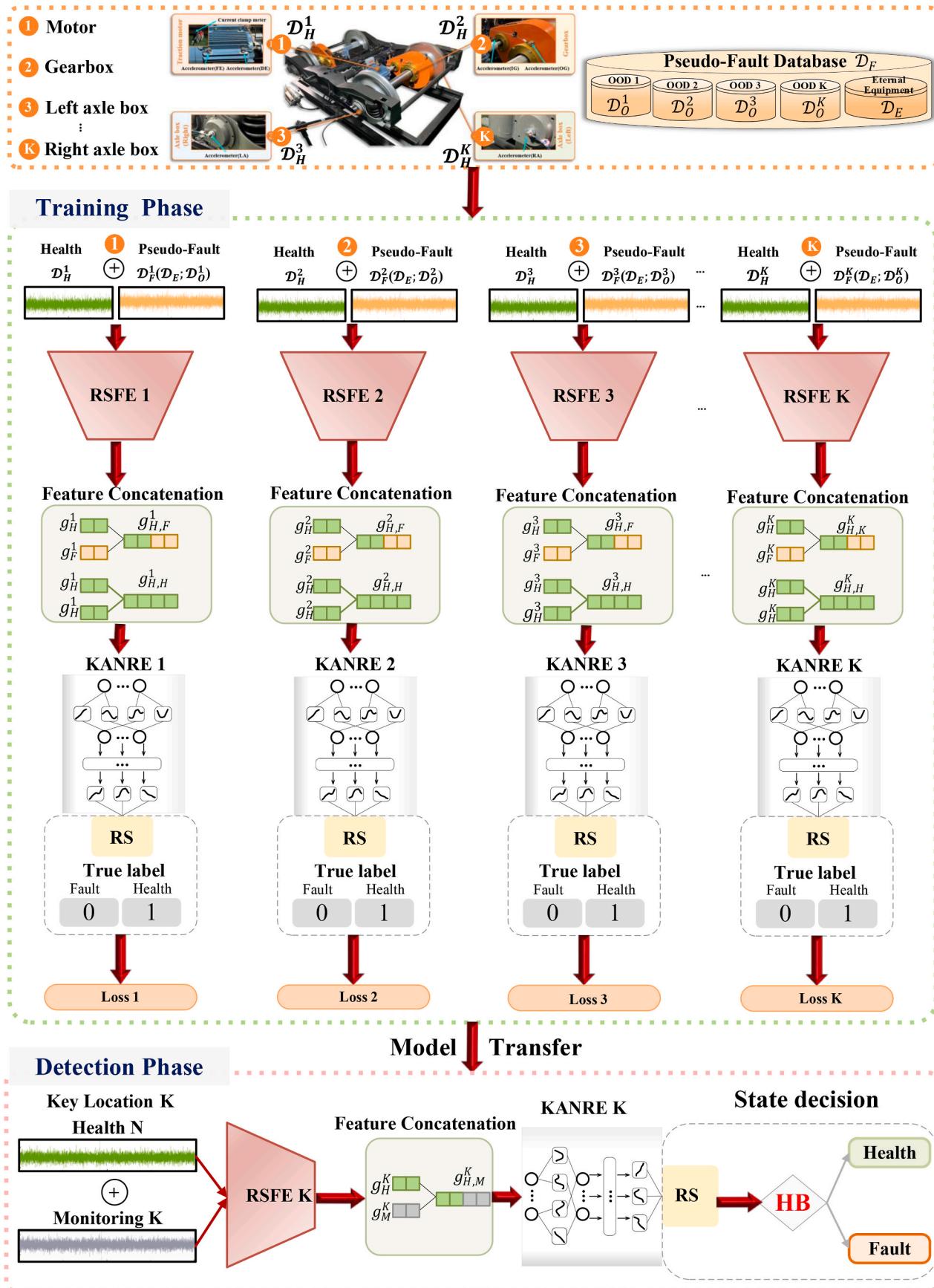


Fig. 3. Fault detection and localization framework base on PF-ERN.

relation evaluator (KANRE). Specifically, the RSFE extracts features from the training dataset  $\mathcal{D}[\mathcal{D}_H^K; \mathcal{D}_F^K(\mathcal{D}_E, \mathcal{D}_0^K)]$ . The resulting feature sets,  $g_H^K$  and  $g_F^K$ , represent the health state features and the pseudo-fault features at the  $K - th$  key component, respectively. These feature sets are then processed by FCM, which concatenates  $g_H^K$  and  $g_F^K$  to generate feature pairs  $(g_{H,F}^K, g_{H,F}^K)$ . The KANRE analyzes these feature pairs, assessing their similarity and computing corresponding relation scores. This process effectively evaluates the similarity between health state features and pseudo-fault features to achieve fault detection for  $K - th$  component. Each ERN is optimized by minimizing the loss between predicted and actual relation scores, enabling it to effectively evaluate feature relationships. In summary, the fault detection and localization process based on PF-ERN  $\mathbb{M}(\mathcal{D}_H, \mathcal{D}_F; \zeta)$  for TTS, using the  $K$  ERN model  $M_K(\mathcal{D}_H^K, \mathcal{D}_F^K; \theta_K)$ , as outlined in the following workflow.

$$\mathbb{M}(\mathcal{D}_H, \mathcal{D}_F; \zeta) = \begin{cases} M_1(\mathcal{D}_H^1, \mathcal{D}_F^1; \theta_1) \\ \vdots \\ M_K(\mathcal{D}_H^K, \mathcal{D}_F^K; \theta_K) \end{cases} \rightarrow \begin{cases} S_1(\mathcal{D}_H^1, \mathcal{D}_F^1); S_1(\mathcal{D}_H^1, \mathcal{D}_F^1) \\ \vdots \\ S_K(\mathcal{D}_H^K, \mathcal{D}_F^K); S_K(\mathcal{D}_H^K, \mathcal{D}_F^K) \end{cases} \rightarrow \begin{cases} r_{\mathcal{D}, \mathcal{D}}^1 \\ \vdots \\ r_{\mathcal{D}, \mathcal{D}}^K \end{cases} \rightarrow \begin{cases} (C_H \text{ or } C_F)^1 \\ \vdots \\ (C_H \text{ or } C_F)^K \end{cases} \quad (8)$$

Here,  $\theta_K$  represents the learnable parameters of the  $K$  ERN model.  $S_K(\mathcal{D}_H^K, \mathcal{D}_F^K)$  denotes the similarity between  $\mathcal{D}_H^K$  and  $\mathcal{D}_F^K$  at the  $K - th$  key detection component. Similarly,  $r_{\mathcal{D}, \mathcal{D}}^K$  and  $(C_H \text{ or } C_F)^K$  correspond to the relation scores and the predicted condition for the  $K - th$  key detection component, where  $C_H$  denotes the health state and  $C_F$  indicates the fault state.

During the detection phase, real-time monitoring data  $\mathcal{D}_M(\mathcal{D}_M^1, \mathcal{D}_M^2, \dots, \mathcal{D}_M^K)$  are collected from the  $K$  key components within the TTS, where  $\mathcal{D}_M^K$  represents the signals collected from the  $K - th$  key detection component. These monitoring data, which may reflect either a healthy or faulty state, are processed by the proficiently trained PF-ERN model  $\mathbb{M}(\mathcal{D}_H, \mathcal{D}_F; \zeta)$ . The model analyzes the monitoring data  $\mathcal{D}_M(\mathcal{D}_M^1, \mathcal{D}_M^2, \dots, \mathcal{D}_M^K)$  by comparing them against historical health state data  $\mathcal{D}_H$ , generating relation scores that indicate the similarity between  $\mathcal{D}_M(\mathcal{D}_M^1, \mathcal{D}_M^2, \dots, \mathcal{D}_M^K)$  and  $\mathcal{D}_H$  for each component. This process is formalized as follows:

$$\mathbb{M}(\mathcal{D}_H, \mathcal{D}_M; \zeta) = \begin{cases} M_1(\mathcal{D}_H^1, \mathcal{D}_M^1; \theta_1) \\ \vdots \\ M_K(\mathcal{D}_H^K, \mathcal{D}_M^K; \theta_K) \end{cases} \rightarrow \begin{cases} S_1(\mathcal{D}_H^1, \mathcal{D}_M^1) \\ \vdots \\ S_K(\mathcal{D}_H^K, \mathcal{D}_M^K) \end{cases} \rightarrow \begin{cases} r_{\mathcal{D}_H, \mathcal{D}_M}^1 \\ \vdots \\ r_{\mathcal{D}_H, \mathcal{D}_M}^K \end{cases} \quad (9)$$

The state of each key component in the TTS is determined using the following strategy:

$$\begin{cases} \text{Health}, r_{\mathcal{D}_H, \mathcal{D}_M}^1 \geq HB \\ \vdots \\ \text{Fault}, r_{\mathcal{D}_H, \mathcal{D}_M}^K < HB \end{cases} \quad (10)$$

where  $HB$  represents the health baseline. If  $r_{\mathcal{D}_H, \mathcal{D}_M}^K \geq HB$ , the  $K - th$  key detection component of TTS is classified as being in a health state. Conversely, if  $r_{\mathcal{D}_H, \mathcal{D}_M}^K < HB$ , the  $K - th$  key detection component is identified as exhibiting a fault. Using this strategy, the PF-ERN model  $\mathbb{M}(\mathcal{D}_H, \mathcal{D}_F; \zeta)$  effectively performs machinery fault detection and localization.

### 3.2. Pseudo-fault data enhanced relation network structure

The PF-ERN consists of four key modules: the pseudo-fault database construction, the Continuous Wavelet Transform (CWT), the RSFE, the FCM, and KANRE. Each module contributes uniquely to the overall functionality of the PF-ERN.

#### (1) Pseudo-Fault Database Construction

The pseudo-fault database  $\mathcal{D}_F$  is systematically constructed to improve the performance of the ERN by enabling it to effectively extract the hidden and latent information from the health state data. This database serves as a crucial module in the fault detection and localization framework. It provides a synthetic yet realistic representation of fault scenarios, enabling the model to train effectively even without

actual fault data. By generating pseudo-fault data, the training dataset is enriched, which improves the model's ability to generalize across different fault types.

The pseudo-fault database  $\mathcal{D}_F$ , as shown in Fig. 4, consists of external equipment fault datasets  $\mathcal{D}_E$  and corresponding out-of-distribution datasets  $\mathcal{D}_O(\mathcal{D}_0^1, \mathcal{D}_0^2, \dots, \mathcal{D}_0^K)$ . The external equipment fault datasets  $\mathcal{D}_E$  are constructed by leveraging data from similar fault types observed in other equipment, based on the common fault types occurring in the TTS. On the other hand, the OODD  $\mathcal{D}_O(\mathcal{D}_0^1, \mathcal{D}_0^2, \dots, \mathcal{D}_0^K)$  are generated from historical health state data  $\mathcal{D}_H(\mathcal{D}_H^1, \mathcal{D}_H^2, \mathcal{D}_H^3, \mathcal{D}_H^K)$  collected from key component in the TTS. These health data are processed by the OODD generator  $\mathbb{G}(\cdot)$  to produce OODD  $\mathcal{D}_O(\mathcal{D}_0^1, \mathcal{D}_0^2, \dots, \mathcal{D}_0^K)$ . According to equations (6) and (7), the generation mechanism of  $\mathbb{G}(\cdot)$  is as follows:

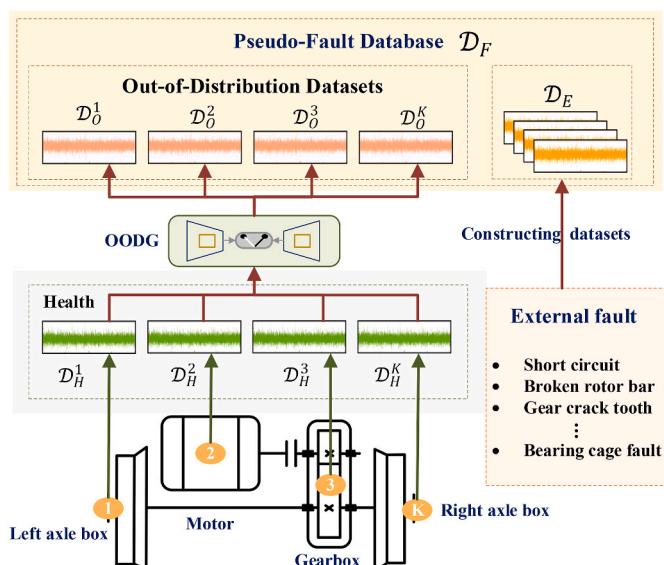


Fig. 4. Pseudo-fault database construction.

$$\mathcal{D}_O = \sum_{i=1}^K \mathbb{G}(\mathcal{D}_H^i, N) = \sum_{i=1}^K \sum_{j=1}^N \left( \mathbf{x}_H^i + a_j \cdot \frac{\mathcal{I}_j}{|\mathcal{I}_j|} + \sigma n_j \right) \quad (11)$$

where  $N$  represents the number of OODD samples to be generated for each health state data  $\mathcal{D}_H^i$ . The radius  $a_j$  for the  $j - th$  sample is drawn from normal distribution  $\mathcal{N}(0, I_D)$ , while  $\mathcal{I}_j$  is a random vector sampled from  $\mathcal{N}(0, I_D)$ , normalized by its magnitude  $|\mathcal{I}_j|$ . Consequently,  $K$  corresponding training datasets  $\mathcal{D}_F^K(\mathcal{D}_E, \mathcal{D}_0^K)$  are constructed for the  $K - th$  ERN models.

The combination of  $\mathcal{D}_E$  and  $\mathcal{D}_O$  into the pseudo-fault database  $\mathcal{D}_F$  creates a robust training set that merges external fault knowledge with system-specific fault simulations. This approach not only enriches the dataset but also enables the model to effectively extract hidden and latent information from the health state data, improving its generalization and fault detection capabilities across a wide range of scenarios.

## (2) CWT

The CWT (Yan et al., 2014) is utilized on raw vibration signals to identify health-related traits at multiple resolutions, creating time-frequency images that reflect key characteristics. The wavelet transform is defined by the following equation:

$$\mathcal{W}\mathcal{T}_f(\omega, \tau) = |\omega|^{-\frac{1}{2}} \int_{-\infty}^{+\infty} f(t)\varphi\left(\frac{t-\tau}{\omega}\right) dt \quad (12)$$

where  $\omega$  represents the scaling factor,  $\tau$  is the translation parameter,  $f(\cdot)$  is the input signal,  $\varphi(\cdot)$  is the mother wavelet function. All training datasets are transformed into time-frequency images using CWT and fed into the RSFE.

## (3) RSFE

The impact of noise on feature extraction is particularly pronounced in real-world industrial scenarios, significantly undermining the accuracy and reliability of fault detection (Wang et al., 2025e). To this end, Residual shrinkage network (RSN) (Zhao et al., 2020) serves as the backbone for feature extraction. The RSN is effective in suppressing noise and enhancing the representativeness of extracted features. Unlike traditional ResNets, the RSN incorporates a soft thresholding

mechanism that filters out irrelevant features while preserving key data. This approach also helps maintain stable gradients during back-propagation, mitigating common issues in deep networks such as vanishing and exploding gradients.

The RSN architecture incorporates four residual shrinkage units, along with an average pooling layer. Each residual shrinkage unit, includes a specialized module for calculating soft thresholds. This module uses Global Average Pooling (GAP) to compute the absolute values of feature  $x$ , generating one-dimensional vectors. These values are then passed through a fully connected layer to determine scaling factors, which are constrained between  $[0, 1]$ , using a sigmoid function:

$$\alpha = \frac{1}{1 + e^{-x}} \quad (13)$$

where  $x$  is the output from the fully connected layer. The threshold  $\tau$  is for the soft thresholding is then computed as:

$$\tau = a \cdot \text{average } |x_{i,j,c}| \quad (14)$$

Here,  $x_{i,j,c}$  refers to the feature at position  $(i,j)$  and channels  $c$ . The threshold  $\tau$  is carefully calibrated to ensure efficient noise suppression while retaining key features. This design enhances the module's ability to handle noisy data while maintaining robust feature extraction.

The RSFE can extract features  $(g_H^1, g_H^2, \dots, g_H^K)$  and  $(g_F^1, g_F^2, \dots, g_F^K)$  for the  $\mathcal{D}_H(\mathcal{D}_H^1, \mathcal{D}_H^2, \dots, \mathcal{D}_H^K)$  and  $\mathcal{D}_F(\mathcal{D}_F^1, \mathcal{D}_F^2, \dots, \mathcal{D}_F^K)$ , respectively.

## (4) FCM

To enhance the ability of the PF-ERN to evaluate similarity between the feature sets  $(g_H^1, g_H^2, \dots, g_H^K)$  and  $(g_F^1, g_F^2, \dots, g_F^K)$ , a novel FCM process is introduced, which creates feature pairs. The FCM employs two distinct concatenation mechanisms to uncover both health-unique and health-inherent attributes. These splicing mechanisms enable the model to capture both the distinctive patterns that separate healthy and faulty states, and the general health-related characteristics across multiple detection components in the TTS, as shown in Fig. 5.

The first FCM focuses on revealing health-unique attributes by concatenating the health features  $(g_H^1, g_H^2, \dots, g_H^K)$  with pseudo-fault features  $(g_F^1, g_F^2, \dots, g_F^K)$ . This concatenation operation serves to expose the subtle differences between the normal and pseudo-fault states for each

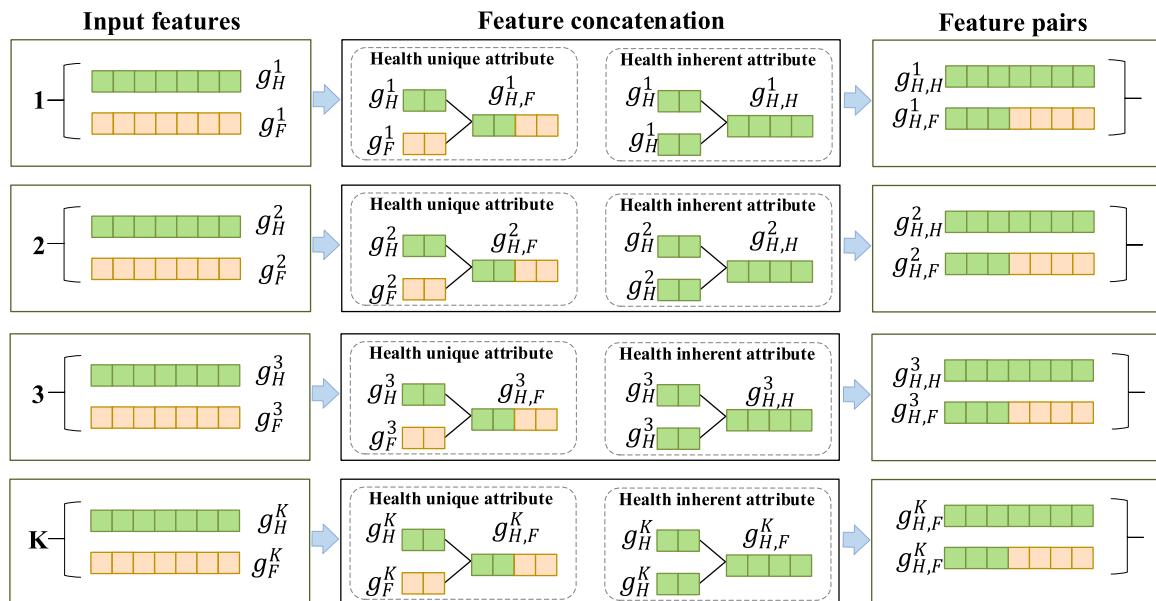


Fig. 5. Feature concatenation mechanism.

key detection component within the TTS. Specifically, for each component, the concatenation is expressed as follows:

$$\begin{cases} g_{H,F}^1 = [g_H^1, g_F^1] \\ \vdots \\ g_{H,F}^K = [g_H^K, g_F^K] \end{cases} \quad (15)$$

where  $g_H^k$  represents the health features of the  $K - th$  key detection component, and  $g_F^k$  denotes the pseudo-fault features at same component. The notation  $[., .]$  indicates the concatenation operation applied between health features and pseudo-fault features for each component. This concatenation operation is repeated for all key components, resulting in a set of feature pairs  $(g_{H,F}^1, g_{H,F}^2, \dots, g_{H,F}^K)$ . These pairs are designed to capture and emphasize the distinguishing features of the health and fault states at each critical component.

The second FCM aims to uncover the health-inherent attributes by concatenating the health state features of each component with themselves. This captures the fundamental, consistent patterns within the health state, regardless of the specific fault scenario. This operation is defined as:

$$\begin{cases} g_{H,H}^1 = [g_H^1, g_H^1] \\ \vdots \\ g_{H,H}^K = [g_H^K, g_H^K] \end{cases} \quad (16)$$

The resulting feature pairs  $(g_{H,H}^1, g_{H,H}^2, \dots, g_{H,H}^K)$  represent the inherent characteristics shared by the health states across different components in the TTS. By focusing on the health-inherent attributes, the PF-ERN method effectively captures the commonalities within the health state, ensuring that the model is not overly sensitive to noise or fluctuations in the data.

## (5) KANRE

The KANRE is engineered to capture the deep and intricate relationships within feature pairs, which include health state features  $(g_{H,H}^1, g_{H,H}^2, \dots, g_{H,H}^K)$  and pseudo-fault features  $(g_{H,F}^1, g_{H,F}^2, \dots, g_{H,F}^K)$ .

Leveraging its multi-layer architecture, as depicted in Fig. 6, the KANRE iteratively extracts complex relationships between these feature pairs. It achieves this by modeling the connection between inputs and outputs through the composition of univariate continuous functions at each layer. The input activation values  $g_i$  are transmitted to the subsequent layer, where they undergo a series of one-dimensional nonlinear transformations, yielding the output  $g_{l+1}$ . This layer-to-layer propagation is mathematically expressed as:

$$g_{l+1,j} = \sum_{i=1}^{n_l} \phi_{l,j,i}(g_{l,i}) \quad (17)$$

Here,  $g_{l,i}$  represents the activation value of the  $i - th$  neuron in the  $l - th$  layer, and  $\phi_{l,j,i}$  is the activation function for that neuron. With each layer's transformation, the KANRE progressively uncovers and models the intricate relationships within the feature pairs  $(g_{H,H}^1, g_{H,H}^2, \dots, g_{H,H}^K)$  and  $(g_{H,F}^1, g_{H,F}^2, \dots, g_{H,F}^K)$ . The activation function  $\phi(g)$  in the KANRE is composed of a basis function and a spline function. Specifically, the activation function within the KANRE is composed of two components: a basis function  $b(g)$  and a B-spline function  $spline(g)$ . Specifically, the activation function is constructed as a weighted combination of these components:

$$\phi(g) = \omega_b(g) + \omega_s \text{spline}(g) \quad (18)$$

Where, SiLU activation function is used for the basis function  $b(g)$

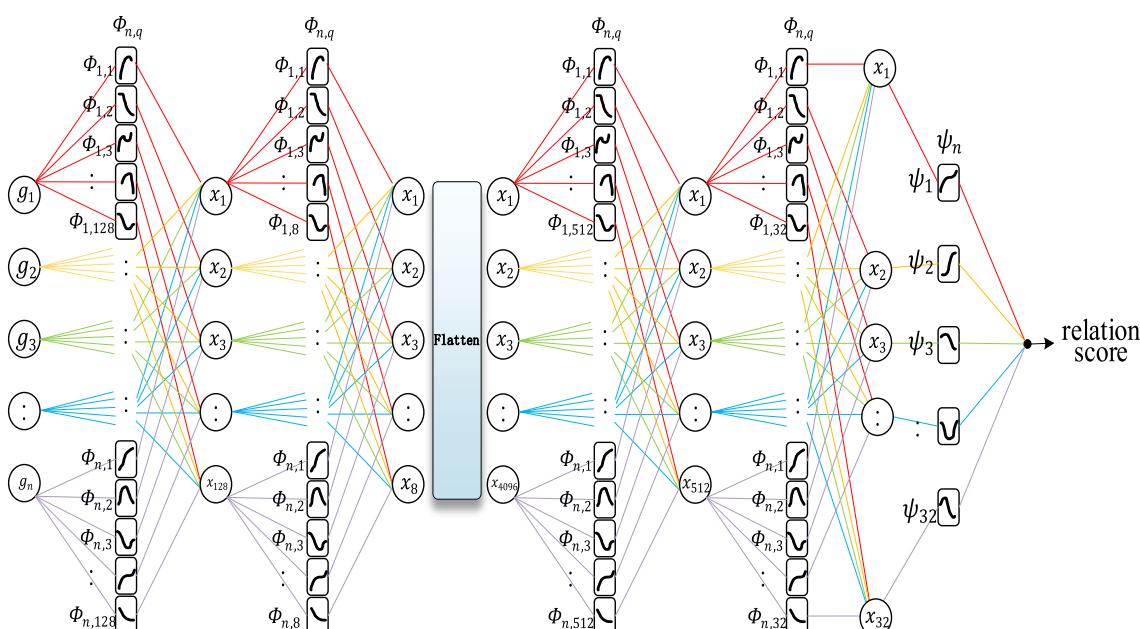
$$\text{SiLU}(g) = \frac{g}{1 + e^{-g}} \quad (19)$$

The spline function  $spline(g)$  is represented as a linear combination of B-spline basis functions:

$$\text{spline}(g) = \sum_i c_i B_i(g) \quad (20)$$

where  $c_i$  is trainable coefficients, and  $B_i(g)$  is the  $i - th$  B-spline basis function. This combination of basis and spline functions allows the KANRE to adapt its nonlinear representations based on input features, thereby improving its flexibility and effectiveness.

Through successive layer-wise propagation, the KANRE constructs



**Fig. 6.** Kolmogorov-Arnold Network relation evaluator structure.



Fig. 7. Experimental platform of subway train transmission systems.

and refines the complex relationships between the input feature pairs. In the final layer, information from all preceding layers is aggregated via a weighted summation to generate the ultimate output. For a KANRE with a depth of 4, this output is represented as:

$$\text{KAN}(g) = (\phi_4 \cdot \phi_3 \cdot \phi_2 \cdot \phi_{L-2})g_o \quad (21)$$

Each layer  $\phi_l$  consists of a matrix of one-dimensional functions that incrementally extract and model the similarities and differences between the feature pairs. The final output layer consolidates all features into a relation score matrix, which quantifies the similarity or correlation between the feature pairs:

$$r = \text{GELU}\left(BN\left(\text{Concat}\left(\sum_{i=1}^N F_i \otimes g\right)\right)\right) \quad (22)$$

where, GELU serves as the activation function, BN is batch normalization, Concat represents concatenation, and  $\otimes$  denotes the interaction between features. The use of GELU helps to effectively mitigate overfitting in the KANRE. The final relation score  $r$  lies within the range  $[0, 1]$ , reflecting the probability that the feature pair belongs to the same class. For each key detection component, a relation score matrix is computed. The combined relation score matrix is represented as

$$R_p = \begin{cases} r_p^1 = \begin{bmatrix} r_{H,F}^1 & \dots & r_{H,F}^1 \\ r_{H,H}^1 & \dots & r_{H,H}^1 \\ \vdots & \dots & \vdots \\ r_{F,F}^1 & \dots & r_{F,F}^1 \end{bmatrix} \\ \vdots \\ r_p^K = \begin{bmatrix} r_{H,F}^K & \dots & r_{H,F}^K \\ r_{H,H}^K & \dots & r_{H,H}^K \\ \vdots & \dots & \vdots \\ r_{F,F}^K & \dots & r_{F,F}^K \end{bmatrix} \end{cases} \quad (23)$$

where  $r_p^K$  refers to predicted relation score corresponding to  $K - th$  key detected component.

### 3.3. Training strategy

The PF-ERN model's is calculated by evaluating the difference between the predicted relation score  $R_p$  and actual relation score  $R_T$ . Below, the matrix layout depicting  $R_T$  is presented by

$$R_T = \begin{cases} r_T^1 = \begin{bmatrix} 0 & \dots & 0 \\ 1 & \dots & 1 \\ \vdots & \dots & \vdots \\ 1 & \dots & 1 \end{bmatrix} \\ \vdots \\ r_T^K = \begin{bmatrix} 0 & \dots & 0 \\ 1 & \dots & 1 \\ \vdots & \dots & \vdots \\ 1 & \dots & 1 \end{bmatrix} \end{cases} \quad (24)$$

Following this, the loss is computed using the MSE function, which is expressed as:

$$L_{MSE} = \sum (R_p - R_T)^2 = \sum_{i=1}^K Loss_i = (r_p^i - r_T^i)^2 \quad (25)$$

Here,  $r_p^i$  and  $r_T^i$  represent the predicted and actual relation scores for the  $i - th$  feature pair. The total loss across all pairs is the sum of the squared differences. Let  $\theta_{f_{RSFE}}$  and  $\theta_{f_{KANRE}}$  represent the parameters of the RSFE and KANRE, respectively. The loss function is then optimized with respect to these parameters by minimizing the following objective:

$$\delta_{f_{RSFE}}^*, \delta_{f_{KANRE}}^* \leftarrow \underset{\delta_{f_{RSFE}}, \delta_{f_{KANRE}}}{\operatorname{argmin}} L_{MSE}(\theta_{f_{RSFE}}, \theta_{f_{KANRE}}) \quad (26)$$

In this equation,  $\delta_{f_{RSFE}}^*$  and  $\delta_{f_{KANRE}}^*$  represent the optimal parameters that minimize the loss function, ensuring the alignment of the predicted relation scores with the actual scores. This optimization allows the PF-ERN to improve its ability to model the complex relationships between feature pairs.

## 4. Experiment study

### 4.1. Experimental setup and dataset description

A fault simulation experiment using subway TTS (Ding et al., 2024a) is carried out to showcase the effectiveness of the PF-ERN in detecting and locating faults. As illustrated in Fig. 7, the experimental platform is

designed and scaled at a 1:2 ratio relative to the actual subway train bogies. The transmission system consists of a motor, a gearbox, and two axle boxes. The entire system is powered by a three-phase asynchronous AC motor, with the load applied through electro-hydraulic load equipment. The motor utilizes SKF6205-2RSH bearings, and the reduction gearbox is equipped with helical gears. The driving gear has 16 teeth, while the driven gear features 107 teeth. The axial bearing for the driving gear is an HRB 32305 model, and the axle box bearing type is HRB 352213.

The subway TTS is equipped with four key detection components: traction motors, driving gearboxes, and left and right axle boxes (see Fig. 7). Vibration sensors are assigned to each component for monitoring purposes. To investigate various fault scenarios, experiments were conducted across three modes: single fault mode, component-level compound fault mode, and system-level compound fault mode. As detailed in Table 1, a total of 21 fault experiments were carried out, consisting of 16 single-fault experiments, 2 component-level Furukawa fault experiments, and 3 system-level compound fault experiments. Fig. 8 presents photographs illustrating the fault.

Simulations conducted during the experiments. Vibration sensors were configured with a sampling rate of 64 kHz, and was run for a duration of 120 s. For each experiment, 1000 samples were recorded for each fault condition and 2000 samples for the health condition, with each sample containing 2048 data points. The training set included 800 samples from the healthy state, and 200 health state samples were reserved for validation. The test dataset contained 1000 samples representing both the health and fault states.

The external equipment fault datasets as one of the pseudo-fault data are obtained from the high-speed train axle box bearing test bench and the gearbox test bench, as shown in Fig. 9. The specific types and quantities of the pseudo-fault datasets are summarized in Table 2. The high-speed train axle box bearing test bench is capable of simulating three single fault scenarios and two composite fault scenarios. On the other hand, the gearbox test bench can simulate a total of fourteen fault scenarios, including six single fault scenarios and eight composite fault scenarios. Additionally, fifteen OODD were generated using the OODD Generator. All pseudo-fault datasets are used exclusively as training datasets.

#### 4.2. Experimental results

The PF-ERN training setup parameters include a learning rate of 0.001, a batch size of 8, and the Adam optimizer. Prior to performing fault detection and localization using the PF-ERN method on the TTS, the *HB* value must first be determined. The *HB* value range is set between 0.9 and 0.99, and the optimal *HB* value for each key component is identified using the validation dataset. Fig. 10 displays the average accuracy of the validation dataset for various *HB* values, with each value being tested in 10 repeated trials. The PF-ERN method consistently achieves detection accuracy than 90 % across all *HB* values. When the *HB* value is configured to 0.95, the detection accuracy at each key component is relatively balanced leading to the selection of 0.95 as the optimal *HB* values.

The PF-ERN method is initially evaluated in the context of single fault detection and localization modes to assess its effectiveness. As the dataset lacks single fault data for the right axle box, the focus of the tests is on other components for fault detection. The test results, summarized in Fig. 11, are derived by averaging data from ten identical trials to minimize error. The PF-ERN method demonstrates excellent performance on the motor and left axle box, with accuracy and recall rates exceeding 97.5 %, and many reaching up to 99.5 %. For fault state M2, M3, and LA3, both accuracy and recall rates are perfect at 100 %, indicating flawless detection. However, the performance on the gearbox is less robust compared to the other components. For faults state such as G6 and G7, detection results approach 100 %, while for fault state like G2 and G4, accuracy and recall rates remain just above 60 %. The

**Table 1**  
Health state in the datasets.

Mode	Fault location	Fault type	Fault code	Sample size
Single Fault	Motor	Health	M0	2000
		Short circuit	M1	1000
		Broken rotor	M2	1000
		bar		
		Bearing fault	M3	1000
		Bowed axis	M4	1000
		Health	G0	2000
		Gear cracked	G1	1000
		tooth		
		Gear worn	G2	1000
		tooth		
		Gear missing	G3	1000
		tooth		
		Gear chipped	G4	1000
		tooth		
		Bearing inner race fault	G5	1000
		Bearing outer race fault	G6	1000
		Bearing rolling element fault	G7	1000
		Bearing cage fault	G8	1000
	Left axle boxes	Health	LA0	2000
		Bearing inner race fault	LA1	1000
		Bearing outer race fault	LA2	1000
		Bearing rolling element fault	LA3	1000
		Bearing cage fault	LA4	1000
		Health	RA0	2000
Component-level compound-fault	Gearboxes	Gear cracked tooth, Bearing inner race fault	G1+ G5	1000
		Gear worn tooth, Bearing inner race fault	G2+ G5	1000
		Gear missing tooth, Bearing inner race fault	G3+ G5	1000
		Gear chipped tooth, Bearing inner race fault	G4+ G5	1000
		Bearing inner race fault, Bearing outer race fault	LA1+ LA2	1000
		Bearing outer race fault, Bearing rolling element fault	LA2+ LA3	1000

(continued on next page)

**Table 1 (continued)**

Mode	Fault location	Fault type	Fault code	Sample size
	Left axle box	Bearing outer race fault, Bearing cage fault	LA2+ LA4	1000
	Left axle box	Bearing inner race fault, Bearing outer race fault, Bearing rolling element fault	LA1+ LA2+ LA3	1000
	Left axle box	Bearing inner race fault, Bearing outer race fault, Bearing cage fault	LA1+ LA2+ LA4	1000
	Left axle box	Bearing inner race fault, Bearing outer race fault, Bearing rolling element fault, Bearing cage fault	LA1+LA2+LA3+LA4	1000
System-level compound-fault	Motor, Gearbox	Short circuit, Gear missing tooth	M1+G3	1000
	Motor, Gearbox	Short circuit, Bearing inner race fault	M1+G5	1000
	Motor, Gearbox	Broken rotor bar, Gear missing tooth	M2+G3	1000
	Motor, Gearbox	Broken rotor bar, Bearing inner race fault	M2+G5	1000
	Motor, Gearbox	Bearing fault, Gear missing tooth	M3+G3	1000
	Motor, Gearbox	Bearing fault, Bearing inner race fault	M3+G5	1000
	Motor, Gearbox	Bowed axis, Gear missing tooth	M4+G3	1000
	Motor, Gearbox	Bowed axis, Bearing inner race fault	M4+G3	1000
	Motor, Left axle box	Short circuit, Bearing inner race fault	M1+LA1	1000
	Motor, Left axle box	Broken rotor bar, Bearing inner race fault	M1+LA1	1000
	Motor, Left axle box	Bearing fault, Bearing	M1+LA1	1000

**Table 1 (continued)**

Mode	Fault location	Fault type	Fault code	Sample size
		inner race fault		
	Motor, Left axle box	Bowed axis, Bearing inner race fault	M1+LA1	1000
	Gearbox, Left axle box	Gear missing tooth, Bearing inner race fault	G3+LA1	1000
	Left axle box, Right axle box	Bearing inner race fault	LA1, RA1	1000
		Bearing inner race fault		
	Motor, Gearbox, Left axle box	Short circuit, Gear missing tooth, Bearing inner race fault	M1+G3+LA1	1000
	Motor, Gearbox, Left axle box	Short circuit, Bearing inner race fault	M1+G5+LA1	1000
	Motor, Gearbox, Left axle box	Broken rotor bar, Gear missing tooth, Bearing inner race fault	M2+G3+LA1	1000
	Motor, Gearbox, Left axle box	Broken rotor bar, Bearing inner race fault	M2+G5+LA1	1000
	Motor, Gearbox, Left axle box	Bowed axis, Gear missing tooth, Bearing inner race fault	M4+G3+LA1	1000
	Motor, Gearbox, Left axle box	Bowed axis, Bearing inner race fault	M4+G5+LA1	1000
	Motor, Left axle box, Right axle box	Short circuit, Bearing inner race fault	M1+LA1+ RA1	1000
	Motor, Left axle box, Right axle box	Broken rotor bar, Bearing inner race fault	M2+LA1+ RA1	1000
	Motor, Gearbox, Left axle	Short circuit, Bearing inner race fault	M1+G5+LA1+ RA1	1000

(continued on next page)

**Table 1 (continued)**

Mode	Fault location	Fault type	Fault code	Sample size
	box, Right axle box	Bearing inner race fault, Bearing inner race fault		

findings demonstrate that the PF-ERN approach effectively identifies and pinpoints faults within the TTS under diverse fault conditions.

Fig. 12 displays the distribution of relation scores generated by the PF-ERN method across various states of different components. These relation scores, ranging from 0 to 1, are visualized using color bars that transition from red (indicating fault states) to green (representing health states). The scores are computed based on the model's similarity assessment between feature pairs of the input data and reference health states, with higher scores reflecting greater similarity to healthy conditions. In most instances, the color bars are predominantly either red or green, demonstrating the PF-ERN method's ability to accurately differentiate between fault and health states. This indicates that the model effectively evaluates feature pair similarity, yielding reliable relation scores for fault detection. However, for fault types G3 and G4 in the gearbox, we observe a mixture of green and red bars, signaling a higher incidence of misclassifications. This can be attributed to the complexity of these fault types, which may produce feature patterns that partially resemble healthy states, posing challenges for accurate detection. These misclassifications point to potential areas for enhancing the model's performance, such as refining feature extraction techniques or improving similarity metrics. Nevertheless, the overall findings in Fig. 13 affirm that the PF-ERN method excels at detecting and localizing faults in the TTS by assessing feature pair similarity, with only specific

cases showing reduced performance.

Furthermore, Fig. 13 provides a supplementary perspective on understanding model performance by visualizing the feature distributions across different components. The results reveal that for the motor component, the features of the healthy state exhibit almost no overlap with those of the faulty states, enabling clear separation. This feature distribution aligns closely with the relation scores for the motor shown in Fig. 12, highlighting the model's reliability in distinguishing motor states. In contrast, the feature distribution for the gearbox component presents a more complex scenario. While most faulty states can be differentiated from the healthy state, the features of faults G2, G3, and G4 show significant overlap with the healthy state. This overlap directly contributes to the alternating red and green color bars observed in the gearbox relation scores in Fig. 12 for faults G2, G3, and G4, indicating a higher misclassification rate for these specific faults. Upon further analysis, this misclassification likely arises from the similarity in feature patterns between faults G2, G3, and G4 and the healthy state in certain key dimensions, making it difficult for the model to fully distinguish them. Overall, the PF-ERN method demonstrates effective differentiation between healthy and faulty states in TTS, underscoring its practical value in fault detection, while also pointing to potential areas for improvement when addressing complex fault features.

#### 4.3. Result discussion

To systematically analyze the impact of pseudo-fault data on the fault detection performance of the PF-ERN method, a motor component fault detection experiment is specifically designed. In this experiment, the pseudo-fault dataset was generated using the Soft Brownian Offset technique, with the key parameter  $k$  set to 7. This parameter value was selected based on experimental results from reference (Moller et al., 2021), striking a balance between computational efficiency and the quality of the generated pseudo-fault data. As shown in Fig. 14, the

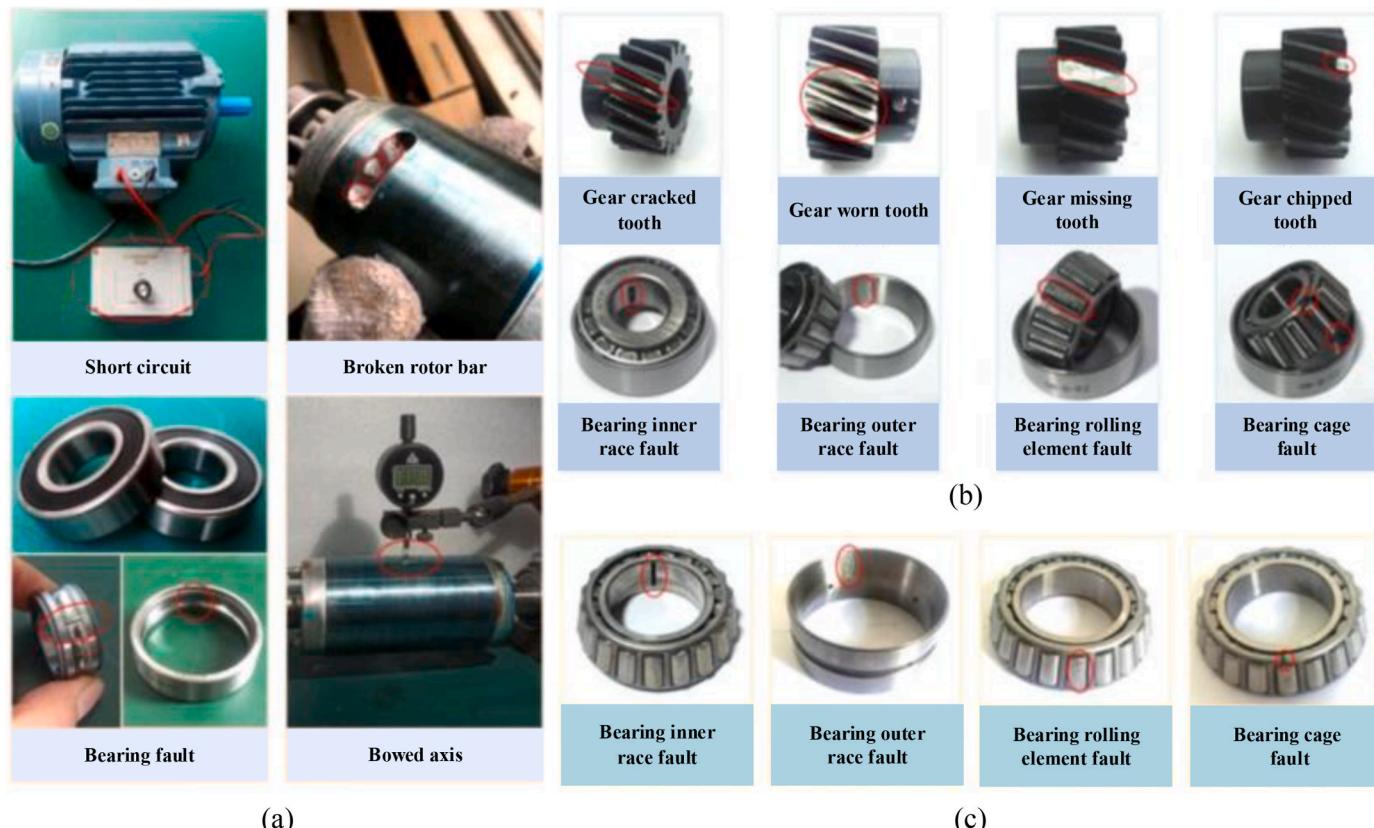


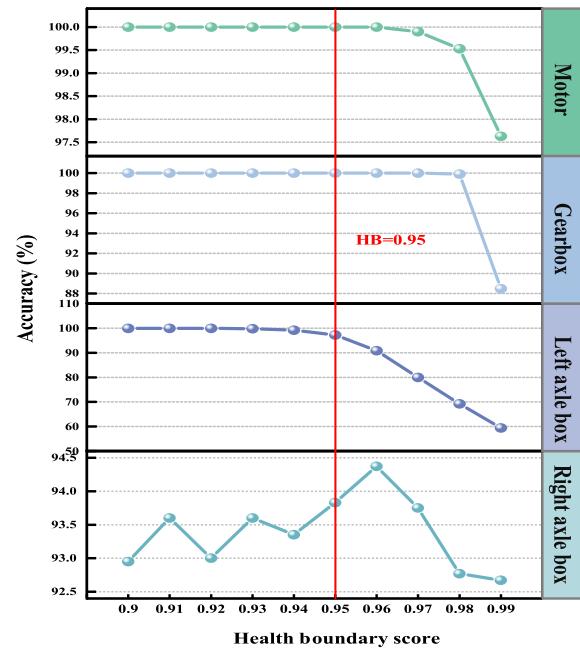
Fig. 8. Faults of different transmission components. (a) Motor. (b) Gearbox. (c) Axle box.



**Fig. 9.** Experimental bench for external equipment failures as part of the generation of pseudo-fault data. (a) high-speed train axle box bearing test bench; (b) gearbox test bench.

**Table 2**  
Detailed description of data construction for the Pseudo-fault data.

Pseudo-fault data	Fault type	Sample size
High-speed train axle box test bench	Bearing inner ring fault	1000
	Bearing outer ring fault	1000
	Bearing rolling element fault	1000
	Bearing inner race fault, Bearing outer race fault	1000
	Bearing inner race fault, Bearing rolling element fault	1000
Gearbox test bench	Broken tooth	1000
	Lacking tooth	1000
	Cracked tooth	1000
	Coupling loose	1000
	Bearing roller wear	1000
	Bearing outer ring wear	1000
	Broken tooth, Lacking tooth	1000
	Broken tooth, Cracked tooth	1000
	Broken tooth, Coupling loose	1000
	Broken tooth, Bearing roller wear	1000
	Broken tooth, Bearing outer ring wear	1000
	Lacking tooth, Coupling loose	1000
	Lacking tooth, Bearing roller wear	1000
Out-of-distribution data	Lacking tooth, Bearing outer ring wear	1000
	fifteen out-of-distribution data	15 × 1000



**Fig. 10.** Determination of the PF-ERN HB value for TTS.

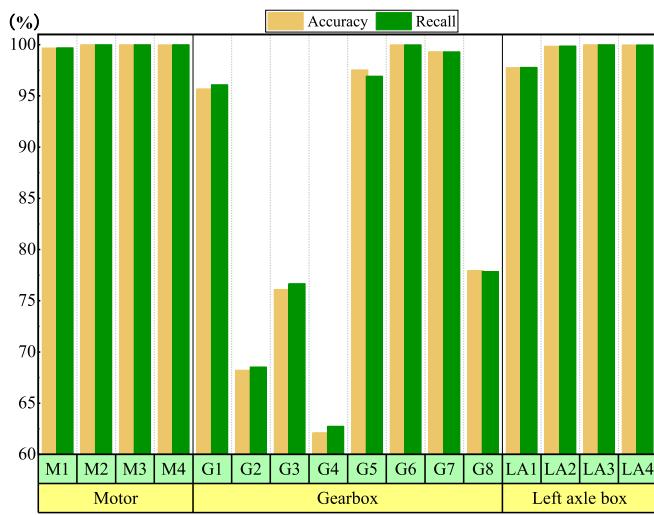


Fig. 11. PF-ERN detection outcomes for single fault modes.

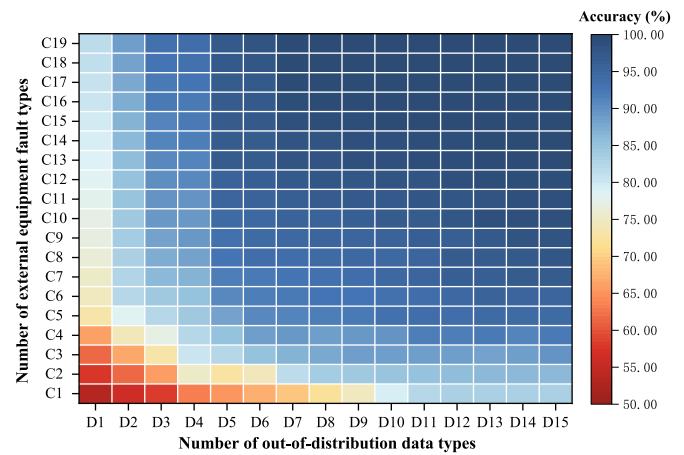


Fig. 14. Influence of the number of pseudo-fault types on the detection performance of PF-ERN method.

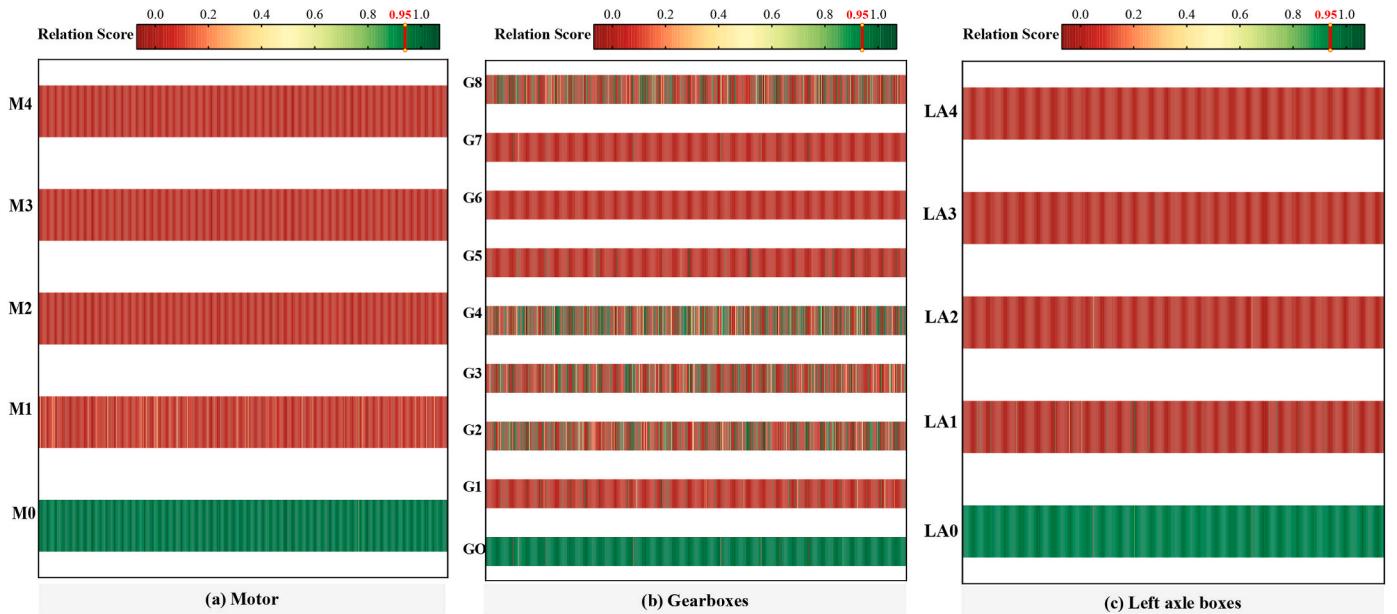


Fig. 12. Relation scores for different states under different tasks.

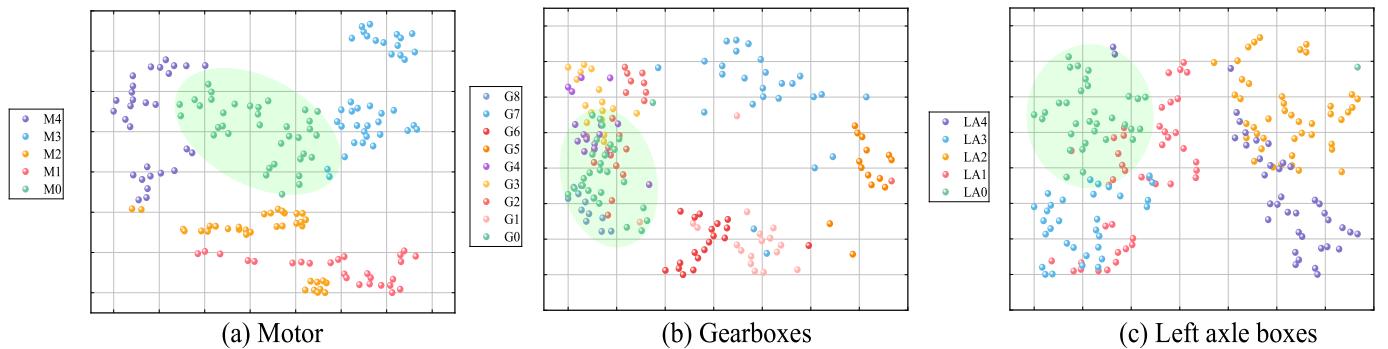


Fig. 13. Visualization of features between different states of different key components.

detection accuracy of the PF-ERN method increases progressively with the inclusion of additional external equipment fault types and OODD fault types. However, this improvement plateaus beyond a certain threshold, where further diversification of fault types yields only

marginal gains in accuracy. This trend suggests that while the diversity of pseudo-fault types significantly enhances the method's detection capability, there exists an intrinsic limit beyond which additional variety contributes little to performance, likely due to saturation in the

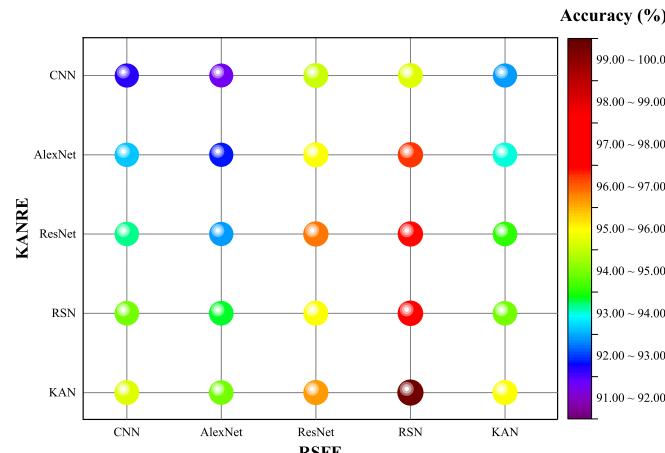
model's ability to differentiate fault signatures. The challenge of pinpointing the exact number of pseudo-fault types required for optimal performance highlights a practical consideration: when computational resources permit, adopting a broad spectrum of pseudo-fault types is advantageous. This approach not only reduces the risk of underfitting caused by limited fault diversity but also strengthens the model's robustness, enabling it to generalize effectively across varied fault conditions encountered in real-world scenarios.

To investigate the contributions of RSFE and KANRE to the detection performance of the PF-ERN method, we conducted experiments testing various model combinations, including CNN, AlexNet, ResNet, RSN, and KAN. As illustrated in Fig. 15, all tested configurations achieved detection accuracies exceeding 91 %, underscoring the robustness and versatility of the PF-ERN framework across diverse deep learning architectures. This consistent high performance suggests that PF-ERN is highly adaptable, maintaining effectiveness regardless of the backbone network employed. Among the evaluated combinations, the pairing of RSN as the RSFE and KAN as the KANRE distinguished itself by achieving an exceptional detection accuracy of 99 %, markedly surpassing other configurations. This standout result highlights a powerful synergy between RSN and KAN, positioning this combination as the optimal setup within the PF-ERN framework. The superior performance can be attributed to the complementary strengths of RSFE's feature extraction capabilities and KANRE's relational evaluation prowess, which together maximize the method's fault detection precision. Based on these compelling findings, we have selected the RSN and KAN combination as the focal point for subsequent experiments. This decision not only capitalizes on the observed synergy but also establishes a robust foundation for further enhancing detection accuracy.

#### 4.4. Complex fault diagnosis scenarios

To further validate the superior performance of the proposed PF-ERN method, fault detection is tested under various compound fault scenarios. As shown in Table 1, compound faults are categorized into component-level and system-level compound faults. The component-level faults refer to cases where multiple fault types occur simultaneously within a single component, while system-level faults involve the simultaneous failure of multiple components.

Fig. 16 showcases the effectiveness of the PF-ERN method in detecting component-level compound faults, achieving remarkable accuracy in both dual and quadruple fault scenarios. The detection results for faulty components are presented here. In quadruple fault cases, precision and recall rates exceed 95 %, markedly outperforming those in dual fault scenarios. This enhancement suggests that the presence of



**Fig. 15.** Influence of different selection of RSFE and KANRE on the detection performance of PF-ERN method.

multiple simultaneous fault types strengthens fault signals, making anomalies more distinct. As a result, PF-ERN's pseudo-fault data enhancement strategy and feature concatenation mechanism are effectively harnessed, enabling highly accurate fault diagnosis.

Similarly, Fig. 17 illustrates the detection results for system-level compound faults. The PF-ERN method delivers average precision and recall rates above 96 % when identifying faults in the motor, left axle box, and right axle box, underscoring its robustness and versatility across subsystems. While detection performance for gearbox faults is slightly lower, the method maintains strong overall consistency and reliability. The cornerstone of PF-ERN's success lies in its feature concatenation mechanism, which pairs health state data with pseudo-fault data. This approach reveals attributes unique to the health state and intrinsic to its condition, greatly improving the model's ability to differentiate between faulty and healthy states. Consequently, PF-ERN sustains high detection accuracy even in intricate compound fault scenarios, establishing a solid technical basis for precise fault detection and localization in TTS.

#### 4.5. Comparison analysis

This study conducts a comprehensive comparison of deep learning, TL, and few-shot learning (FSL) methods for fault detection in TTS, with a focus on scenarios where fault samples are scarce or unavailable. The deep learning approaches evaluated include CNN, RSN, and autoencoders. For TL methods, Domain-Adversarial Neural Network (DANN) (Ganin et al., 2016) and Deep Subdomain Adaptation Network (DSAN) (Zhu et al., 2021) are assessed, Matching Network (MN) Matching Network (MN) (Vinyals et al., 2016) and Conditional Self-Supervised Learning (CSS) (Wang et al., 2020) represent FSL techniques. These methods are benchmarked against the proposed PF-ERN to evaluate their effectiveness across three fault conditions: single fault mode, component-level compound fault mode, and system-level compound fault mode. To ensure a fair comparison, all models—CNN, RSN, autoencoder, DANN, DSAN, MN, and CSS—were trained, validated, and tested using the same datasets and training parameters as PF-ERN, including learning rate, batch size, and optimizer. The network architectures were also standardized to align with PF-ERN's structure, minimizing the influence of hyperparameter discrepancies and isolating performance differences to the models' inherent capabilities.

Fig. 18 presents the fault detection performance of these techniques under the three specified fault conditions. A notable observation is that detection accuracy remains relatively stable across all methods as fault complexity increases. This stability can be attributed to the amplified distinction between fault and healthy signals in more complex fault modes, which facilitates detection despite the growing intricacy. Notably, Deep learning methods, specifically CNN and ResNet, exhibit the weakest performance, with average accuracies consistently below 60 %. Trained exclusively on health states and pseudo-fault data, these models struggle to generalize to unseen fault types encountered during testing. Their reliance on patterns observed in the training data limits their adaptability, rendering them less effective for detecting novel TTS faults. The autoencoder, ranking third in detection accuracy, adopts a different approach by constructing a health index from healthy data and comparing new inputs against it. While this method excels at identifying deviations from normalcy, its exclusive reliance on healthy state data makes it hypersensitive to minor variations within that state. Consequently, it often produces false positives when encountering faults not explicitly represented during training, undermining its reliability.

Similarly, the performance of these TL methods is moderate, achieving an average detection accuracy of approximately 70 %. In this context, the source domain comprises health states and pseudo-faults, while the target domain includes both health and actual fault states. Despite their ability to adapt knowledge across domains, these methods face challenges due to significant distribution shifts between pseudo-fault and real fault data. This discrepancy hampers their capacity to

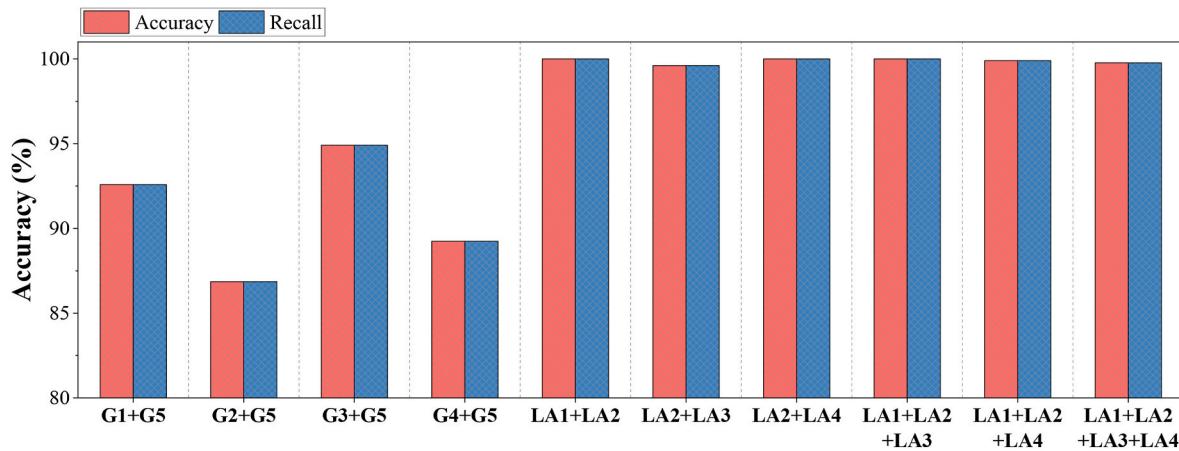


Fig. 16. Detection results of PF-ERN in component-level compound-fault scenario.

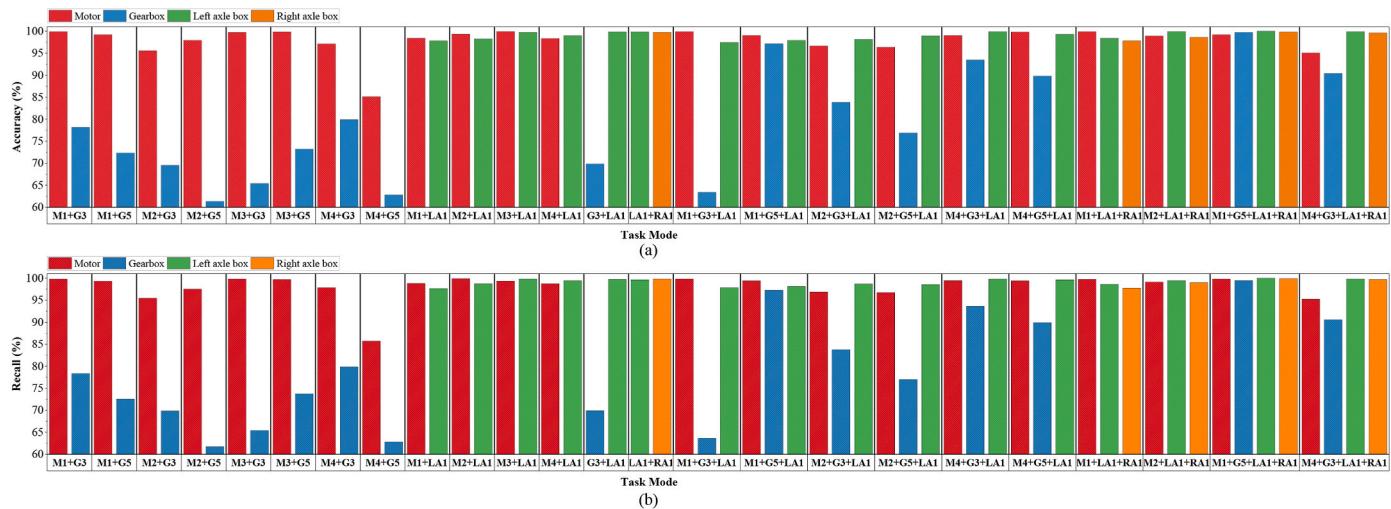


Fig. 17. Detection results of PF-ERN in System-level compound-fault scenario. (a) Accuracy; (b) Recall.

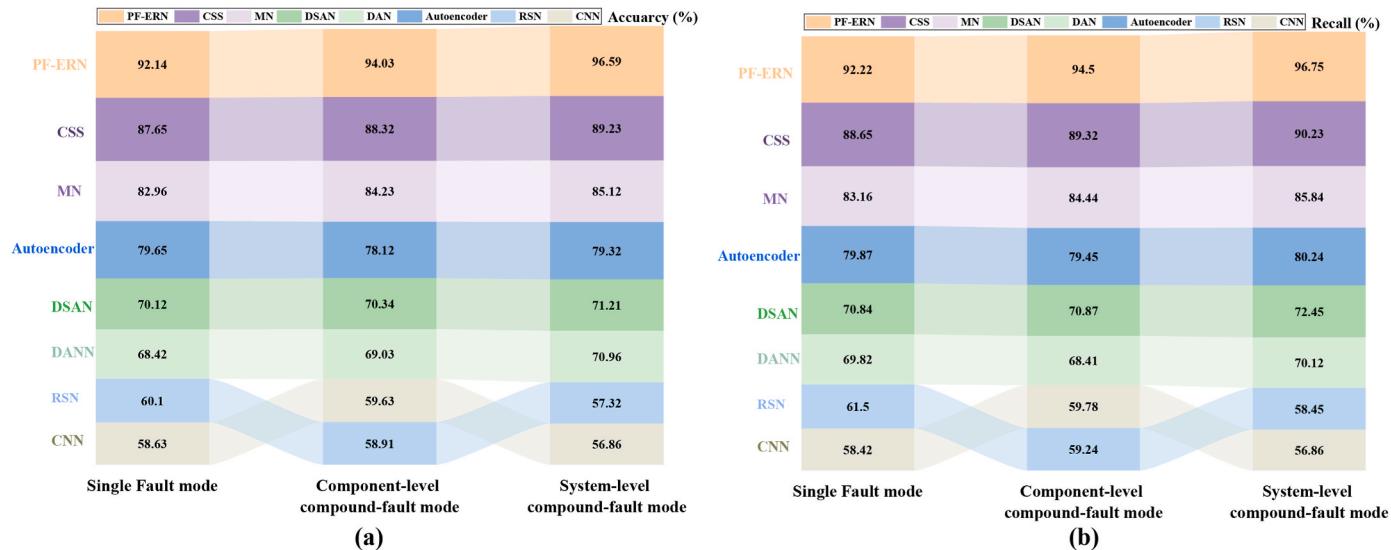


Fig. 18. Differences in detection performance among various methods. (a) Accuracy; (b) Recall.

**Table 3**

Comparison of different methods in terms of training time and detection time.

Method	Training time (min)	Detection time (sec)
CNN	18.4	32.4
RSN	20.8	35.1
DANN	30.7	48.6
DSAN	36.4	50.9
Autoencoder	18.9	33.5
MN	24.5	44.5
CSS	35.6	132.4
PF-ERN	30.9	96.2

extract domain-invariant features, resulting in suboptimal knowledge transfer and detection performance across varying fault domains. Few-shot learning methods exhibit competitive performance, securing the second-best detection results. These methods assess fault states by measuring similarity between observed samples and healthy states, leveraging health-unique features across different states for classification. However, their approach overlooks health-inherent features within a single state, which can compromise classification precision and limit their effectiveness compared to more comprehensive methods.

In contrast, The proposed PF-ERN method stands out as the top performer, surpassing few-shot learning methods by at least 5 % in accuracy. Its superiority stems from a sophisticated analysis of relational patterns between feature pairs, integrating both health-inherent features within the same state and health-unique features across different states. This dual-focus feature extraction enables PF-ERN to capture a more complete representation of fault characteristics, significantly enhancing its detection capabilities. This comparative evaluation underscores the strengths and limitations of various fault detection techniques in TTS systems. While deep learning methods like CNN and ResNet struggle with generalization, and transfer learning methods like DANN and DSAN are constrained by domain shifts, few-shot learning approaches like MN and CSS offer competitive but incomplete solutions. The autoencoder provides a moderate alternative but is prone to false positives. In contrast, PF-ERN's robust feature analysis positions it as the most effective method, offering superior accuracy and adaptability for fault detection in TTS, particularly under data-scarce conditions.

Furthermore, to clearly analyze the performance differences in training and testing times across methods, multiple approaches were compared under a unified experimental environment (equipped with an NVIDIA GeForce RTX 3060 GPU, 12 GB memory). Table 3 presents the time comparison results for training and testing phases. The proposed PF-ERN method requires slightly more training time than some methods but is not the most time-consuming (compared to the CSS method). In the testing phase, its time is also marginally higher. However, Table 1 reflects the total time for processing the entire sample set, while the testing time for a single sample is only milliseconds, fully meeting real-world industrial real-time requirements. Analysis shows that, despite slightly higher time complexity, the proposed PF-ERN method's significant advantage in fault detection accuracy (see Fig. 18) strongly demonstrates its practicality and efficiency for industrial applications.

## 5. Conclusion

This study presents an advanced PE-ERN model tailored for identifying and locating faults within TTS, tackling the issue of sparse or absent fault data. By leveraging pseudo-fault data generated through a combination of external equipment fault datasets and OODD created via soft Brownian offset, The ability of the PE-ERN to detect and localize faults is effectively enhanced. The PE-ERN also incorporates the FCM to create feature pairs, capturing both health-unique and health-inherent attributes, which significantly improves fault detection performance. Experimental results on a subway TTS demonstrate that the proposed PE-ERN method outperforms existing fault detection models. The PE-ERN method achieved fault detection accuracy exceeding 97 % for

most fault states, with traction motors and axle boxes performing particularly well (approaching 99 %). The inclusion of pseudo-fault data significantly enhanced detection and localization, with performance improving as the number of pseudo-fault types increased. This highlights the importance of using diverse pseudo-fault data, especially when actual fault data is scarce. The FCM, which feature pairs health and pseudo-fault data, improved the model's ability to differentiate between healthy and faulty states, enhancing accuracy and sensitivity, particularly in noisy environments. The method demonstrated strong performance in both component-level and system-level compound fault scenarios, achieving average accuracy and recall rates above 95 %.

Despite its success, the method has certain limitations. The detection performance still relies on the quality and diversity of the pseudo-fault data, and the model's accuracy could further improve with more comprehensive fault data. Additionally, the proposed approach may face challenges in real-time applications, particularly with more dynamic operational conditions.

In future work, the integration of a broader range of fault scenarios and real-time fault detection capabilities will be explored. The pseudo-fault data generation process will be further optimized, with a focus on generalizing it to produce data representative of multiple domains. Additionally, the method will be extended to cross-domain fault diagnosis using techniques such as domain adaptation and transfer learning to address differences in data distribution across various domains. This extension is expected to enhance the model's robustness and applicability in diverse industrial environments, establishing it as a more universal tool for fault detection and localization. Furthermore, the relational network architecture will be deepened to improve model performance. Meta-learning frameworks will also be investigated to enhance generalization across imbalanced datasets and open-set conditions.

## CRediT authorship contribution statement

**Zhixu Duan:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Data curation, Conceptualization. **Ruoxin Liu:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Zuoyi Chen:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation. **Hong-Zhong Huang:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition.

## Declaration of competing interest

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

## Acknowledgments

This research was partially funded by the National Natural Science Foundation of China (Grant No. 52405093), the Natural Science Foundation of Sichuan Province (Grant No. 2025ZNSFSC1269), the National Science and Technology Major Project (Grant No. J2019-IV-0001-0068), the College Students' Innovative Entrepreneurial Training Plan Program under UESTC (Grant No. X202510614040).

## Data availability

Data will be made available on request.

## References

- Cao, H.R., Shao, H.D., Zhong, X., Deng, Q.W., Yang, X.K., Xuan, J.P., 2022. Unsupervised domain-share CNN for machine fault transfer diagnosis from steady speeds to time-varying speeds. *J. Manuf. Syst.* 62, 186–198.
- Chen, Z.Y., Li, Z.R., Wu, J., Deng, C., Dai, W., 2022. Deep residual shrinkage relation network for anomaly detection of rotating machines. *J. Manuf. Syst.* 65, 579–590.
- Chen, Z.Y., Wang, X.Q., Wu, J., Deng, C., Zhang, D.D., 2023a. Relational conduction graph network for intelligent fault diagnosis of rotating machines under small fault samples. *IEEE T Instrum Meas* 72, 1–11.
- Chen, Z.Y., Wu, K., Wu, J., Deng, C., Wang, Y.H., 2023b. Residual shrinkage transformer relation network for intelligent fault detection of industrial robot with zero-fault samples. *Knowl-Based Syst* 268, 110452.
- Chen, D.Y., Liu, R.A., Hu, Q.H., Ding, S.X., 2023c. Interaction-aware graph neural networks for fault diagnosis of complex industrial processes. *IEEE T Neur Net Lear* 34 (9), 6015–6028.
- Chen, Z.Y., Huang, H.Z., Deng, Z.W., Wu, J., 2025a. Shrinkage mamba relation network with out-of-distribution data augmentation for rotating machinery fault detection and localization under zero-faulty data. *Mech. Syst. Signal Process.* 224.
- Chen, Y., Zhang, X.Y., Li, D.Z., Zhou, J.L., 2025b. A new class of fault detection and diagnosis methods by fusion of spatially distributed and time-dependent features. *J Process Contr* 146.
- Ding, A., Qin, Y., Wang, B., Guo, L., Jia, L.M., Cheng, X.Q., 2024a. Evolvable graph neural network for system-level incremental fault diagnosis of train transmission systems. *Mech. Syst. Signal Process.* 210, 111175.
- Ding, A., Qin, Y., Wang, B., Liu, H., Chen, D.W., Tao, L., Guo, L., 2024b. Brownian distance covariance-based few-shot learning framework considering noisy labels for fault diagnosis of train transmission systems. *IEEE T Ind Inform* 21, 136–145.
- Dou, Q., Lu, D.G., Zhang, B.Y., 2025. Physical resilience assessment of road transportation systems during post-earthquake emergency phase: with a focus on restoration modeling based on dynamic Bayesian networks. *Reliab Eng Syst Safe* 257.
- Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V., 2016. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* 17.
- Gao, Z., Jiang, W.X., Wu, J., Dai, T.J., Zhu, H.P., 2024. Nonlinear slow-varying dynamics-assisted temporal graph transformer network for remaining useful life prediction. *Reliab Eng Syst Safe* 248.
- Gao, Z., Jiang, W.X., Wu, J., Wang, Y.H., Zhu, H.P., 2025. A customized dual-transformer framework for remaining useful life prediction of mechanical systems with degraded state. *Mech. Syst. Signal Process.* 230.
- Guo, L., Lei, Y.G., Xing, S.B., Yan, T., Li, N.P., 2019. Deep convolutional transfer learning network: a new method for intelligent fault diagnosis of machines with unlabeled data. *IEEE T Ind Electron* 66 (9), 7316–7325.
- Jiang, G.Q., He, H.B., Xie, P., Tang, Y.F., 2017. Stacked multilevel-denoising autoencoders: a new representation learning approach for wind turbine gearbox fault diagnosis. *IEEE T Instrum Meas* 66 (9), 2391–2402.
- Lei, J.H., Liu, C., Jiang, D.X., 2019. Fault diagnosis of wind turbine based on Long Short-term memory networks. *Renew Energ* 133, 422–432.
- Li, G.Q., Wei, M.R., Shao, H.D., Liang, P.F., Q Duan, C., 2024. Wavelet knowledge-driven transformer for intelligent machinery fault detection with zero-fault samples. *IEEE Sens. J.* 24 (21).
- Liang, P.F., Xu, L.T., Shuai, H.Q., Yuan, X.M., Wang, B., Zhang, L.J., 2024. Semisupervised subdomain adaptation graph convolutional network for fault transfer diagnosis of rotating machinery under time-varying speeds. *IEEE-Asme T Mech* 29 (1).
- Liu, H., Zhou, J.Z., Zheng, Y., Jiang, W., Zhang, Y.C., 2018. Fault diagnosis of rolling bearings with recurrent neural network based autoencoders. *Isa T* 77, 167–178.
- Liu, Y., Zhang, D.B., Yang, Y.T., Liang, R.Q., 2022. Fault characteristics analysis of high-speed train transmission systems. *Shock Vib.* 2022.
- Ma, X.X., Wu, J., Xue, S., Yang, J., Zhou, C., Sheng, Q.Z., Xiong, H., Akoglu, L., 2023. A comprehensive survey on graph anomaly detection with deep learning. *IEEE T Knowl Data En* 35 (12), 12012–12038.
- Ma, Y.H., Ye, F., Li, D.Z., Zhou, J.L., 2025. A novel dynamic nonlinear non-Gaussian approach for fault detection and diagnosis. *Neurocomputing* 624.
- Mao, Z.H., Wang, H., Jiang, B., Xu, J., Guo, H.F., 2024. Graph convolutional neural network for intelligent fault diagnosis of machines via knowledge graph. *IEEE T Ind Inform* 20 (5), 7862–7870.
- Moller, F., Botache, D., Huseljic, D., Heidecker, F., Bieshaar, M., Sick, B., 2021. Out-of-distribution detection and generation using soft brownian offset sampling and autoencoders. *IEEE Comput Soc Conf* 46–55.
- Sung, F., Yang, Y.X., Zhang, L., Xiang, T., Torr, P.H.S., Hospedales, T.M., 2018. Learning to compare: relation network for few-shot learning. *Proc Cvpr IEEE* 1199–1208.
- Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K., Wierstra, D., 2016. Matching networks for one shot learning. *Adv Neur In* 29.
- Wang, K.X., Chen, Y., Guo, H.K., Wen, L.F., Shen, S.J., 2020. Geometric pretraining for monocular depth estimation. *IEEE Int Conf Robot.* <https://doi.org/10.1109/ICRA40945.2020.9196847>.
- Wang, H., Liu, Z.L., Peng, D.D., Zuo, M.J., 2023. Interpretable convolutional neural network with multilayer wavelet for Noise-Robust Machinery fault diagnosis. *Mech. Syst. Signal Process.* 195, 110314.
- Wang, S.Y., Tian, J.Y., Liang, P.F., Xu, X.F., Yu, Z.Z., Liu, S.Y., Zhang, D.L., 2024. Single and simultaneous fault diagnosis of gearbox via wavelet transform and improved deep residual network under imbalanced data. *Eng Appl Artif Intel* 133, 108146.
- Wang, C.D., Yang, J.L., Jie, H.M., Zhao, Z.Y., Wang, W.S., 2025a. An energy-efficient mechanical fault diagnosis method based on neural-dynamics-inspired metric SpikingFormer for insufficient samples in industrial internet of things. *IEEE Internet Things* 12 (1), 1081–1097.
- Wang, C.D., Liu, X.F., Yang, J.L., Jie, H.M., Gao, T.Y., Zhao, Z.Y., 2025b. Addressing unknown faults diagnosis of transport ship propellers system based on adaptive evolutionary reconstruction metric network. *Adv. Eng. Inform.* 65.
- Wang, C.D., Yang, J.L., Jie, H.M., Tao, Z., Zhao, Z.Y., 2025c. A lightweight progressive joint transfer ensemble network inspired by the Markov process for imbalanced mechanical fault diagnosis. *Mech. Syst. Signal Process.* 224.
- Wang, C.D., Shu, Z., Yang, J.L., Zhao, Z.Y., Jie, H.M., Chang, Y.Q., Jiang, S.Q., See, K.Y., 2025d. Learning to imbalanced open set generalize: a meta-learning framework for enhanced mechanical diagnosis. *IEEE T Cybernetics* 55 (3), 1464–1475.
- Wang, C.D., Jie, H.M., Yang, J.L., Gao, T.Y., Zhao, Z.Y., Chang, Y.Q., See, K.Y., 2025e. A multi-source domain feature-decision dual fusion adversarial transfer network for cross-domain anti-noise mechanical fault diagnosis in sustainable city. *Inform Fusion* 115.
- Xiang, L., Wang, P.H., Yang, X., Hu, A.J., Su, H., 2021. Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism. *Measurement* 175, 109094.
- Xu, J., Zhao, Y.J., Bao, W.M., Hao, C., 2025. Fault diagnosis of motor bearing in complex scenarios based on Mamba and Indicative Contrastive Learning. *Eng Appl Artif Intel* 146.
- Yan, R.Q., Gao, R.X., Chen, X.F., 2014. Wavelets for fault diagnosis of rotary machines: a review with applications. *Signal Process.* 96, 1–15.
- Yu, W.K., Zhao, C.H., 2020. Broad convolutional neural network based industrial process fault diagnosis with incremental learning capability. *IEEE T Ind Electron* 67 (6), 5081–5091.
- Zhang, Y.C., Ding, J.L., Li, Y.B., Ren, Z.H., Feng, K., 2024. Multi-modal data cross-domain fusion network for gearbox fault diagnosis under variable operating conditions. *Eng Appl Artif Intel* 133.
- Zhao, M.H., Zhong, S.S., Fu, X.Y., Tang, B.P., Pecht, M., 2020. Deep residual shrinkage networks for fault diagnosis. *IEEE T Ind Inform* 16 (7), 4681–4690.
- Zhao, Y.J., Guo, H., Guo, X.W., 2024. Lithium-ion battery state of health estimation method based on frequency-domain-assisted data decoupling and time-frequency features compensation learning network. *IEEE T Instrum Meas* 73.
- Zhao, X., Liu, Z.R., Wu, C.S., Jin, T.T., 2025. Joint optimization of maintenance and speed selection for transportation systems. *Reliab Eng Syst Safe* 257.
- Zhu, Y.C., Zhuang, F.Z., Wang, J.D., Ke, G.L., Chen, J.W., Bian, J., 2021. Deep subdomain adaptation network for image classification. *IEEE T Neur Net Lear* 32 (4), 1713–1722.