Advanced Engineering Informatics 70 (2026) 104132

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Full length article

Dynamic graph meta-learning with multi-sensor spatial dependencies for cross-category small-sample fault diagnosis in ZDJ9-RTAs

Yuhan Huang [a], Xiaoxi Hu [b,*], Fei Chen [c], Jingming Cao [d], Yiming He [e], Qi Ming [f], Huan Wang [g], Jianing Wang [h], Tao Tang [b]

[a] School of Mechatronics Engineering, Harbin Institute of Technology, 92 West Dazhi St., Harbin, 150001, Heilongjiang, China
[b] State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing, 100044, China
[c] State Key Laboratory of Water Resources Engineering and Management, Wuhan University, Wuhan, 430072, Hubei, China
[d] School of Automation and Intelligence, Beijing Jiaotong University, Beijing, 100044, China
[e] State Key Laboratory of Intelligent Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, 430074, China
[f] College of Computer Science, Beijing University of Technology, Beijing, 100124, China
[g] Department of Systems Engineering, City University of Hong Kong, Hong Kong, China
[h] School of Power and Mechanical Engineering, Wuhan University, Wuhan, 430072, Hubei, China

ARTICLE INFO

Keywords:
Railway turnout actuator
Fault diagnosis
Small sample
Graph neural network
Cross-category meta-learning

ABSTRACT

Railway Turnout Actuators (RTAs) suffer from small-sample fault diagnosis challenges because their high reliability limits fault data availability. Existing RTA small-sample methods show two major limitations—homogeneous category dependency, which prevents adaptation to newly emerging fault types, and single-sensor bias, which neglects spatial dependencies among multi-sensor signals. Focusing on the widely deployed ZDJ9-RTA, this study proposes a dynamic graph meta-learning framework to overcome these issues. The framework integrates (1) a category-shift meta-learning strategy based on dynamic task construction for rapid cross-category adaptation, and (2) a Graph-based Multi-Sensor Information Fusion Network (GMSIFN) that automatically builds sensor topology graphs and fuses multi-sensor features through GRU-gated attention to capture spatial fault propagation paths. Extensive validation on two real-world multi-sensor ZDJ9-RTA datasets — from a railway test line and in-service field deployments — demonstrates consistent superiority across 12 cross-category tasks (covering 2 to 5 novel class scenarios), achieving an average accuracy of 84.3% (3-shot) and gains of 16.1% over optimal baselines, while the field dataset exceeds 97% average accuracy in representative tasks, with peak performance up to 99.95%. These results confirm the proposed framework's effectiveness, thereby extending small-sample studies to dynamically evolving industrial maintenance environments. Code is available at: https://github.com/huang-yu-han/GMSIFN.

1. Introduction

The railway industry has entered a stage of asset saturation, where extensive infrastructure and equipment demand continuous monitoring and diagnostics to ensure reliable operation [1–3]. Railway Turnout Actuators (RTAs) serve as critical devices in modern industrial railways, where their reliable operation is essential for ensuring safe train routing and network efficiency [4]. As illustrated in Fig. 1, taking the widely deployed ZDJ9 turnout actuator as a representative example, RTAs perform precise switch rail movements through coordinated operation of a motor drive, a mechanical transmission subsystem, and a switch circuit controller incorporating a position feedback mechanism. The complex dynamic interactions between these subsystems create multiple potential failure modes that may lead to severe operational disruptions and substantial economic losses [5]. Given the mission-critical nature of RTAs operation, developing advanced diagnostic methods specifically for RTAs systems has emerged as a crucial research focus in railway intelligent maintenance [6].

Early research on fault diagnosis for RTAs predominantly relied on signal processing techniques, such as wavelet transform and empirical mode decomposition [7,8]. These methods focused on manually extracting time–frequency domain features from sensor data, a process heavily dependent on expert knowledge and often time-consuming.

* Corresponding author.
E-mail addresses: 23S108172@hit.edu.cn (Y. Huang), xiaoxhu@bjtu.edu.cn (X. Hu), chenfei@whu.edu.cn (F. Chen), jingmingcao@bjtu.edu.cn (J. Cao), heyiming@hust.edu.cn (Y. He), mingqi@bjut.edu.cn (Q. Ming), hwan36@cityu.edu.hk (H. Wang), 2023302081196@whu.edu.cn (J. Wang), ttang@bjtu.edu.cn (T. Tang).

https://doi.org/10.1016/j.aei.2025.104132
Received 26 June 2025; Received in revised form 20 November 2025; Accepted 21 November 2025

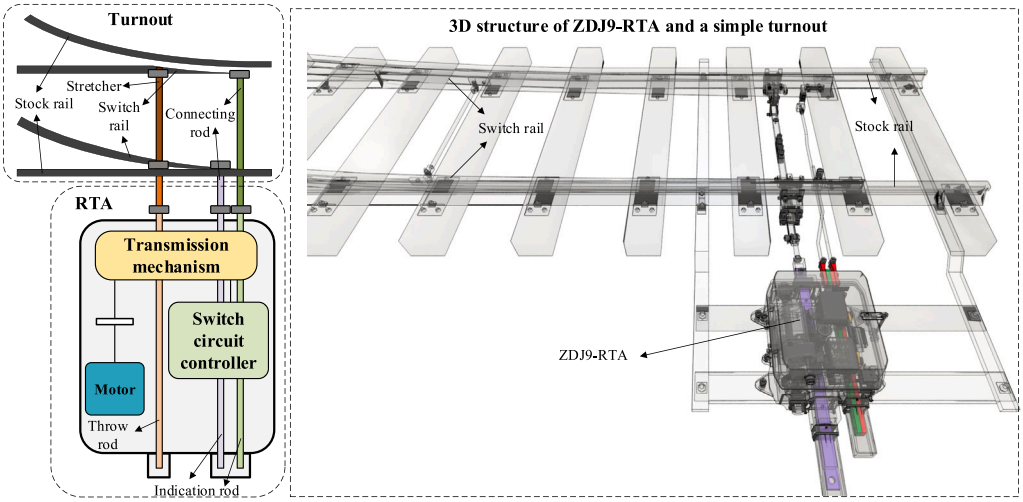

Fig. 1. Schematic diagram a ZDJ9-RTA and a turnout.

Subsequently, the field began to shift towards conventional machine learning approaches [9], which offered a degree of automation in fault diagnosis. For example, support vector machine [10] and random forests [11] could learn decision boundaries from engineered features to classify various fault types. However, these methods suffered from two fundamental limitations: diagnostic accuracy was constrained by the quality of human-crafted features, and they struggled to establish accurate nonlinear correlations between fault manifestations and their corresponding fault categories. The advent of deep learning in recent years has provided a transformative solution to these limitations. Distinct from conventional approaches, deep neural networks demonstrate the capability to autonomously learn discriminative features directly from raw sensor inputs, eliminating dependency on expert-designed features while maintaining superior nonlinear approximation capacity [5,12,13]. This technological breakthrough has catalyzed innovative applications of advanced deep architectures: Chen et al. [14] pioneered the integration of multi-head self-attention mechanisms with residual networks to analyze temporal characteristics in RTA power curves for fault classification. Chen et al. [15] developed a novel dual-scale convolutional neural network to capture multi-resolution fault effectively features from RTA vibration signals. Hu et al. [16] proposed a hybrid diagnostic framework incorporating both spatial and channel attention mechanisms to enhance RTA fault diagnosis accuracy. These approaches have substantially improved diagnostic performance by autonomously learning discriminative fault representations. However, a fundamental limitation persists—deep learning models inherently require large-scale labeled training data to achieve high performance [17]. In contrast, real-world RTA operations yield extremely scarce fault samples due to the component's stringent reliability standards (where failures may lead to significant operational interruptions and financial losses) [18,19]. This intrinsic contradiction between data-hungry deep learning paradigms and limited fault data availability presents a critical research challenge, underscoring the urgent need to develop small-sample learning methodologies for practical RTA fault diagnosis.

To address this fundamental challenge, several pioneering studies have explored small-sample learning methodologies for RTA fault diagnosis. He et al. [6] proposed a flexible semi-supervised meta-learning network that utilizes a dual-channel heterogeneous convolution kernel to extract multi-level fault features from limited samples. Lao et al. [18] developed a semi-supervised weighted prototypical network incorporating a dual-scale neural network to enhance feature representation capabilities under scarce fault data conditions. In another significant contribution, Lao et al. [20] introduced a balanced regularized prototypical network with adaptive current fusion strategies to mitigate the impact of data imbalance in few-shot scenarios. Zhong et al. [21] further advanced this field by integrating a 1D-signal channel attention mechanism with meta-learning to improve cross-domain diagnostic performance with small samples. Despite these notable advances in small-sample diagnosis for RTAs, two critical limitations persist in current approaches:

1. Conventional small-sample learning paradigms typically employ identical fault categories during both training and testing phases, which contradicts the practical railway maintenance requirements where novel fault types continuously emerge [22,23]. This homogeneous category learning framework lacks the capability to rapidly adapt existing fault knowledge to newly encountered fault categories.
2. Existing methods predominantly rely on single-sensor data analysis, whereas real-world RTA faults typically manifest as comprehensive changes across multiple monitoring parameters. This single-sensor dependency fails to capture the complete operational information of the equipment [24], resulting in suboptimal diagnostic performance under complex RTA failure conditions.

To bridge the gap between existing small-sample diagnosis research for RTAs and practical railway maintenance requirements, we establish a novel cross-category small-sample diagnosis scenario for RTAs that fundamentally redefines conventional settings through two critical innovations: (1) cross-category rapid adaptation requiring models to diagnose emerging fault categories absent from training data by leveraging limited novel fault samples, simulating real-world scenario where maintenance systems encounter previously unseen failure modes and must achieve rapid diagnosis to prevent operational disruptions; and (2) multi-sensor data integration mandating the use of correlated multi-sensor monitoring data rather than single-sensor inputs to capture comprehensive equipment states reflective of industrial monitoring environments. Unlike traditional settings where identical fault categories appear in both training and testing phases, our scenario requires models to develop cross-category adaptation capabilities during training that transfer knowledge to novel fault types, achieve rapid adaptation to novel faults using limited samples during deployment, and process synchronized multi-sensor data streams to exploit complementary diagnostic information—directly addressing the dynamic nature of railway maintenance where diagnostic systems must continuously handle evolving fault patterns while leveraging comprehensive monitoring infrastructure.

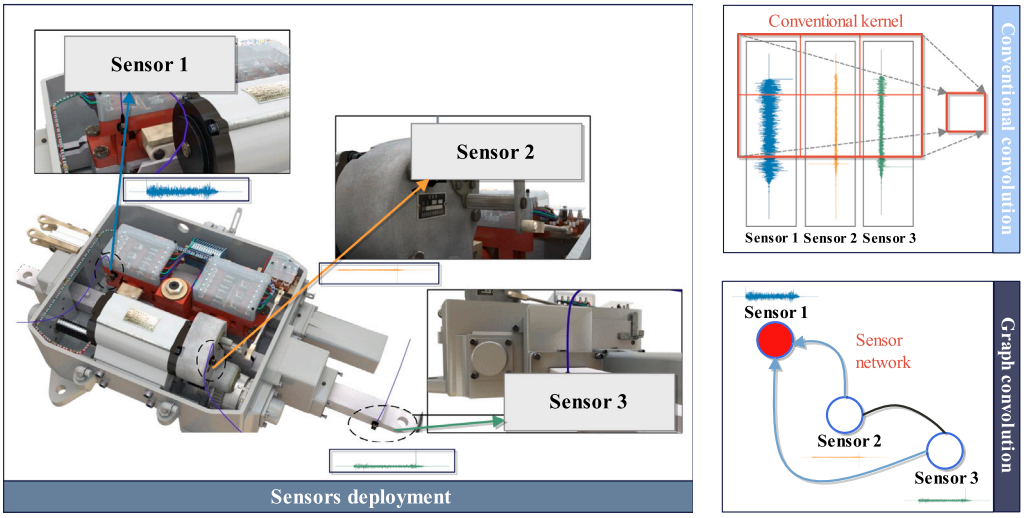

Fig. 2. Illustration of convolutional and graph convolutional operations. (a) The conventional convolution operation computes the sum of pointwise multiplications between multi-sensor signals and a learned convolution kernel. (b) In graph convolution, each sensor signal is treated as a node, and the operation aggregates the node's own features with those of its neighbors through learnable parameters. Unlike traditional convolution, graph convolution inherently captures multi-sensor topological structures by adaptively modeling node-wise dependencies.

Building upon this scenario, we develop an integrated framework addressing both requirements through category-shift meta-learning within the Model-Agnostic Meta-Learning (MAML) framework [25] combined with a Graph-based Multi-Sensor Information Fusion Network (GMSIFN). For cross-category adaptation, our approach constructs meta-training tasks by randomly sampling different category subsets from available training classes during each episode (e.g., Task 1 samples categories A,B; Task 2 samples C,D), deliberately creating inter-task category variations. This design forces the model to develop rapid adaptation capabilities through gradient-based meta-optimization, enabling efficient small-sample fine-tuning to novel fault categories during deployment. For multi-sensor processing, as illustrated in Fig. 2, the GMSIFN architecture fundamentally differs from translation-invariant convolutional methods [26] by explicitly modeling topological dependencies among sensors—treating each sensor as a node and performing graph-level classification to holistically diagnose faults using multi-sensor data. This architecture sequentially executes three core modules: first, the unsupervised graph construction module transforms raw multi-channel sensor data into graph-structured representations by dynamically inferring node relationships through signal similarity measurements, eliminating predefined spatial constraints [27]; second, the node feature aggregation Module updates node embeddings through GRU-gated attention mechanisms that selectively fuse neighborhood features with intrinsic node states via multi-layer message passing, preserving critical node-specific information; finally, the graph classification module utilizes the updated graph representations to combine global graph embeddings with localized node features, forming composite representations for final diagnosis. The co-optimization of these strategies enables simultaneous cross-category adaptability and multi-sensor fusion—delivering a comprehensive solution for railway maintenance scenarios with evolving fault patterns and data scarcity. The contributions of this study are summarized as follows:

1. We propose the first cross-category small-sample fault diagnosis scenario for RTAs. Unlike prior studies that assume identical fault categories during training and testing, our scenario considers non-overlapping categories between these stages, which better reflects the evolving and diverse fault types commonly encountered in real-world railway maintenance operations.
2. To address the rapid adaptation requirements of this new scenario, we first develop a dynamic meta-learning training mechanism that randomly samples subsets of fault categories during training to emulate varying fault distributions, thus enhancing the model's cross-category transferability. Subsequently, during testing, we adopt a "small-sample fine-tuning + full-set evaluation" strategy to ensure the engineering reliability of diagnostic results.
3. We further propose a graph-based multi-sensor fusion architecture comprising an unsupervised topology construction layer and a gated graph attention mechanism, which can effectively capture spatial dependencies among multiple sensors. This architecture enables adaptive aggregation of sensor features and effectively models fault propagation patterns across mechanical components, outperforming traditional single-sensor or simple feature concatenation methods.
4. Extensive experiments conducted on two real-world multi-sensor ZDJ9-RTA datasets collected from railway test lines and field operations demonstrate that our method achieves significant and consistent improvements over state-of-the-art baselines across multiple cross-category diagnosis tasks, validating its practical applicability under dynamic railway maintenance conditions.

The remainder of this paper is structured as follows: Section 2 presents our methodology. Section 3 details the experimental studies and analyzes the results, and Section 4 concludes our work.

2. Methodology

2.1. Problem formulations

For cross-category small-sample fault diagnosis in multi-sensor RTAs, we formally define: Let $\mathcal{Y} = \mathcal{Y}_{\text{train}} \cup \mathcal{Y}_{\text{test}}$ denote the fault category space, where $\mathcal{Y}_{\text{train}} = \{1, \ldots, D_s\}$ represents D_s known fault categories accessible during training, and $\mathcal{Y}_{\text{test}} = \{D_s + 1, \ldots, M\}$ denotes $M - D_s$ novel fault categories emerging during testing. Each sample is a multi-channel time series $\boldsymbol{X} \in \mathbb{R}^{T \times C}$ (T: time length, C: sensor channels). The testing dataset is partitioned into two disjoint subsets: a small adaptation set $\mathcal{I} = \{(\boldsymbol{X}_i, y_i)\}_{i=1}^{|\mathcal{I}|}$ for rapid model adaptation and a large evaluation set $\mathcal{T} = \{(\boldsymbol{X}_j, y_j)\}_{j=1}^{|\mathcal{T}|}$ for performance verification, satisfying $|\mathcal{T}| \gg |\mathcal{I}|$. The core objective is to obtain parameters θ^* such that the diagnostic model f_{θ^*} achieves minimal empirical risk when tested on $\mathcal{T}$ after adaptation using $\mathcal{I}$:

$$\mathcal{R}(\theta^*) = \frac{1}{|\mathcal{T}|} \sum_{(\boldsymbol{X}_j, y_j) \in \mathcal{T}} \mathcal{L}(f_{\theta^*}(\boldsymbol{X}_j), y_j) \rightarrow \min, \tag{1}$$

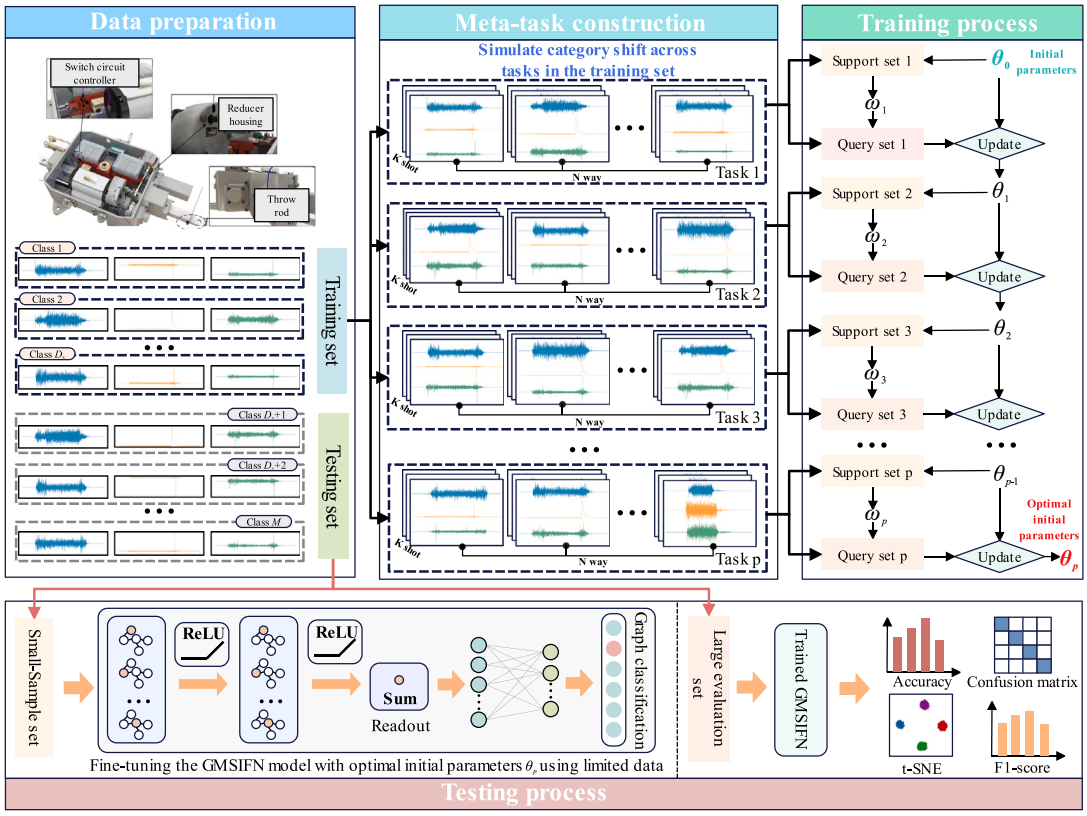

Fig. 3. Framework of cross-category small-sample fault diagnosis for multi-sensor RTAs.

where $\mathcal{L}$ is the cross-entropy loss function quantifying diagnostic error.

The parameter θ^* is obtained via: (1) Meta-training over $\mathcal{Y}_{\text{train}}$ with category-shifted tasks yields initial parameters θ_p; (2) Rapid adaptation using $\mathcal{I}$:

$$\theta^* = \theta_p + \arg\min_{\Delta\theta} \frac{1}{|\mathcal{I}|} \sum_{(\boldsymbol{X}_i, y_i)\in\mathcal{I}} \mathcal{L}(f_{\theta_p+\Delta\theta}(\boldsymbol{X}_i), y_i), \tag{2}$$

where $\theta_p \in \mathbb{R}^d$ is the meta-learned initialization, $\Delta\theta \in \mathbb{R}^d$ denotes parameter adjustment, and $|\mathcal{I}|$ is constrained to be small ($|\mathcal{I}| \ll |\mathcal{T}|$).

This empirical risk $\mathcal{R}(\theta^*)$ directly quantifies the model's cross-category small-sample diagnostic capability for RTAs in real-world maintenance scenarios.

2.2. Overall diagnostic process

The framework of cross-category small-sample fault diagnosis for multi-sensor RTAs, as illustrated in Fig. 3, consists of four key phases:

Data Preparation: Multi-channel RTA signals are acquired from critical sensor measurement points, including switch circuit controllers, reducer housings, and throw rods. These multi-channel signals are organized into a comprehensive dataset whose complete fault category space $\mathcal{Y} = \{1, \ldots, M\}$ is partitioned into disjoint meta-training classes $\mathcal{Y}_{\text{train}} = \{1, \ldots, D_s\}$ and meta-testing classes $\mathcal{Y}_{\text{test}} = \{D_s + 1, \ldots, M\}$, establishing a data foundation for small-sample fault diagnosis in multi-sensor scenarios.

Meta-Task Construction: Within the training class space $\mathcal{Y}_{\text{train}}$, we construct each meta-training episode through random sampling with replacement of N-way categories from $\mathcal{Y}_{\text{train}}$, a process that intentionally introduces variation in the category composition across consecutive tasks. This strategic sampling methodology simulates inter-task category shifts, with each dynamically generated task containing a K-shot support set for inner-loop adaptation and a corresponding query set for outer-loop optimization.

Training Phase: The model undergoes iterative training across multiple meta-tasks. For each task, the procedure comprises: (1) adapting model parameters on the few-shot support samples, (2) evaluating the adapted model on the query set, and (3) updating the initial parameters θ_0 using gradients backpropagated from the query set evaluation. Through multi-task iterative optimization, the model acquires optimal initialization parameters θ_p enabling rapid adaptation to novel fault patterns.

Testing Phase: For novel fault categories $\mathcal{Y}_{\text{test}}$, the model first rapidly fine-tunes θ_p on a limited adaptation set containing small samples and subsequently evaluates diagnostic performance on a large-scale evaluation set containing all remaining test samples. This two-stage process quantifies the model's cross-category small-sample diagnostic capability.

Next, we elaborate on the category-shift meta-learning strategy and GMSIFN model implementation.

2.2.1. Category-shift meta-learning strategy

The proposed category-shift meta-learning strategy simulates real-world category variations through dynamically constructed training tasks. During meta-training, we randomly generate p tasks $\{\mathcal{T}_l\}_{l=1}^{p}$, where each task $\mathcal{T}_l$ samples N categories from known fault classes set $\mathcal{Y}_{\text{train}}$, and collects $K + Q$ instances per category to form the task dataset $D_l = \{(x_i, y_i)\}_{i=1}^{N\times(K+Q)}$. These instances are partitioned into support set $\mathcal{S}_l = \{(x_i^s, y_i^s)\}_{i=1}^{N\times K}$ for inner-loop adaptation and query set $\mathcal{Q}_l = \{(x_i^q, y_i^q)\}_{i=1}^{N\times Q}$ for outer-loop optimization.

The training employs a bi-level optimization framework. For inner-loop adaptation, task-specific parameters are computed as:

$$\omega_l = \theta - \alpha\nabla_\theta \mathcal{L}_{\mathcal{S}_l}(f_\theta), \tag{3}$$

where α denotes the inner-loop learning rate.

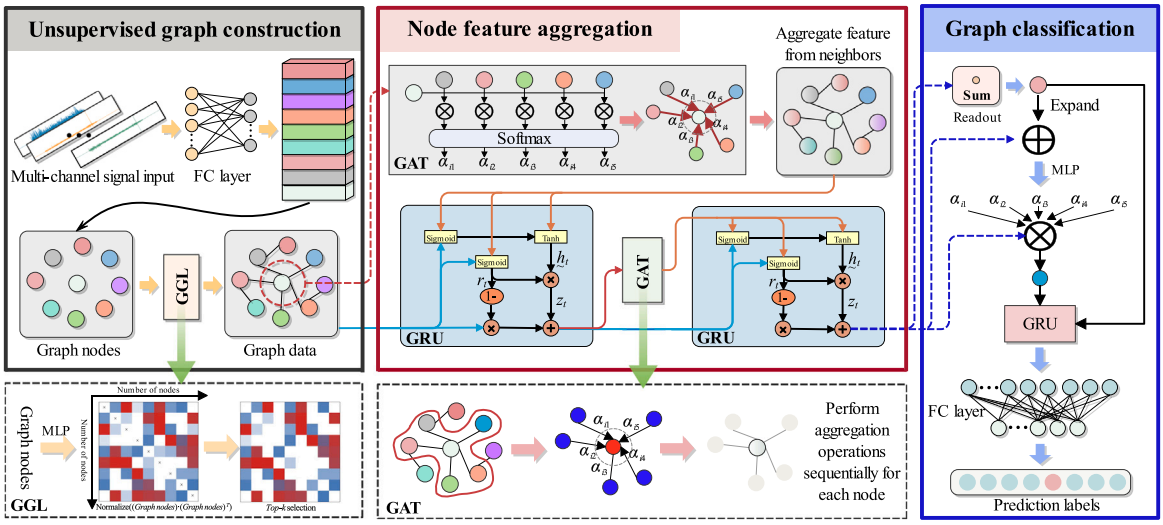

Fig. 4. Framework of GMSIFN model.

The outer-loop optimization updates the initial parameters using gradients from query sets:

$$\theta_l \leftarrow \theta - \beta \nabla_\theta \sum_{l=1}^{p} \mathcal{L}_{\mathcal{Q}_l}(f_{\omega_l}), \tag{4}$$

where β denotes the outer-loop learning rate. This enables the model to learn both rapid adaptation within tasks and meta-knowledge for handling category shifts.

The category shifts between tasks follow a dynamic sampling strategy where the overlap $|\mathcal{Y}_l \cap \mathcal{Y}_{l+1}|$ between consecutive tasks follows a binomial distribution with parameters $(N, N/D_s)$.

During testing on novel fault categories $\mathcal{Y}_{\text{test}}$, the model first performs rapid adaptation using the small-sample adaptation set $\mathcal{I} = \{(\boldsymbol{X}_i, y_i)\}_{i=1}^{|\mathcal{I}|}$ through gradient-based optimization:

$$\theta^* = \theta_p - \alpha \nabla_\theta \mathcal{L}_{\mathcal{I}}(f_{\theta_p}), \tag{5}$$

where θ_p denotes the meta-learned initialization parameters, α is the adaptation learning rate, and $\mathcal{L}_{\mathcal{I}}$ represents the cross-entropy loss computed over the adaptation set.

The adapted model then generates predictions for samples in the large-scale evaluation set $\mathcal{T} = \{(\boldsymbol{X}_j, y_j)\}_{j=1}^{|\mathcal{T}|}$:

$$\hat{y} = f_{\theta^*}(\boldsymbol{X}). \tag{6}$$

This strategy endows the model with two crucial capabilities: (1) gradient-based rapid adaptation to novel fault categories using limited samples, and (2) inherent robustness to category shifts acquired through meta-training with varying task compositions.

2.2.2. GMSIFN model

This paper proposes the GMSIFN model to enhance fault diagnosis performance for RTA in multi-sensor environments. As illustrated in Fig. 4, the network adopts a three-stage processing architecture that effectively integrates multi-sensor fault information through the synergistic operation of: (1) unsupervised graph construction, (2) node feature aggregation, and (3) graph classification. In the unsupervised graph construction stage, the GMSIFN network constructs sensor topology graphs in an unsupervised manner, relying solely on similarity computation between nodes without requiring any manual labeling. First, feature compression is applied to the raw multi-sensor signals through a learnable linear transformation that projects high-dimensional signals into a low-dimensional feature space:

$$H = \text{LeakyReLU}(X W_q + b_q), \tag{7}$$

where $H \in \mathbb{R}^{C\times d}$ represents the compressed multi-sensor feature matrix, $X \in \mathbb{R}^{C\times T}$ denotes the input signal tensor with C sensor channels and T time-series length, $W_q \in \mathbb{R}^{T\times d}$ is the learnable projection matrix, and $b_q \in \mathbb{R}^d$ is the bias term. This operation preserves key temporal features while reducing computational complexity from $\mathcal{O}(C^2T^2)$ to $\mathcal{O}(C^2d^2)$, significantly alleviating the computational burden for subsequent graph construction.

The Graph Generation Layer (GGL) module then establishes topological connections between sensors in an unsupervised fashion. Considering the interaction relationships among nodes in sensor networks, pairwise node similarity is computed as:

$$s_{ij} = \exp(w_e) \cdot (H_i H_j^\top), \tag{8}$$

where $H_i \in \mathbb{R}^d$ and $H_j \in \mathbb{R}^d$ represent feature vectors of the ith and jth sensor nodes respectively, $s_{ij} \in \mathbb{R}$ denotes the raw similarity score, and $w_e \in \mathbb{R}$ is a learnable scaling parameter. This computation captures linear correlations between node features while ensuring positive similarity scores through $\exp(w_e)$, providing a quantitative basis for graph construction.

To form an effective sensor network topology, the computed similarities are transformed into concrete graph representations. A gather operation selects each node's k-nearest neighbors to form the neighbor feature matrix:

$$N_e = \text{gather}(H, \text{topk\_indices}), \tag{9}$$

where $N_e \in \mathbb{R}^{C\times k\times d}$ contains neighbor features, k is the preset neighbor count, and topk_indices $\in \mathbb{R}^{C\times k}$ stores indices of top-k similar neighbors. The gather operation functions as an index selection: $N_{i,j} = H_{\text{topk\_indices}_{i,j}} \forall i \in [1, C], j \in [1, k]$, aggregating the most similar k neighbors.

Simultaneously, edge weights are normalized via a sigmoid function to represent connection strengths: $E = \sigma(s)$, where $E \in \mathbb{R}^{C\times k\times 1}$ is the edge weight matrix and $\sigma(\cdot)$ denotes the sigmoid function. This normalization confines edge weights to (0,1), preserving relative similarity relationships while ensuring numerical stability. The entire process operates unsupervised, automatically building sensor network topology based solely on node feature similarities.

This graph construction mechanism dynamically adjusts connection strength through learnable w_e while employing sparse connectivity ($k \ll C$) to maintain critical topological relationships. The computational complexity remains manageable at $\mathcal{O}(Ckd')$, enabling adaptive capture of fault signal propagation characteristics across sensor networks. The resulting graph structure provides a reliable foundation for subsequent feature aggregation and fault diagnosis.

Based on the unsupervised sensor topology graph, the model proceeds to the node feature aggregation stage. First, neighbor features are fused with edge weights, encoding both neighbor node information and their connection strengths:

$$F = \text{LeakyReLU}(\text{concat}[N_e, E]W_f + b_f), \tag{10}$$

where $W_f \in \mathbb{R}^{(d+1)\times d'}$ is the transformation matrix, and $\mathbf{b}_f \in \mathbb{R}^{d'}$ is the bias term. In railway switch multi-sensor networks, edge weights reflect connection reliability or signal strength between sensor nodes. Incorporating edge information better models node dependencies while avoiding topology information loss from using neighbor features alone.

Next, a feature alignment matrix is constructed:

$$H_{\text{expand}} = H.\text{unsqueeze}(1).\text{expand}(-1, k, -1), \tag{11}$$

$$A = \text{LeakyReLU}(\text{concat}[H_{\text{expand}}, F]W_a + b_a), \tag{12}$$

where $W_a \in \mathbb{R}^{(d+d')\times d'}$ is the alignment matrix, and $b_a \in \mathbb{R}^{d'}$ is the bias term. This step computes feature correlations between central nodes and their neighbors, establishing interaction strength metrics. The alignment operation distinguishes neighbor contributions and captures complex feature interaction patterns, providing the discriminative basis for subsequent attention computation.

Normalized attention weights are obtained via the softmax function, transforming neighbor importance scores into probability distributions for competitive selection along the neighbor dimension: $\boldsymbol{\alpha} = \text{softmax}(A)$. The $\boldsymbol{\alpha} \in \mathbb{R}^{C\times k\times d'}$ indicates neighbor importance weights, ensuring comparable contribution metrics. Feature aggregation is then performed:

$$F_{\text{transformed}} = FW_t + b_t, \tag{13}$$

$$C = \sum(\boldsymbol{\alpha} \odot F_{\text{transformed}}, \text{dim} = 1), \tag{14}$$

where $W_t \in \mathbb{R}^{d'\times d'}$ denotes the linear transformation matrix projecting neighbor features to the aggregation space, $b_t \in \mathbb{R}^{d'}$ represents the corresponding bias vector, and $C \in \mathbb{R}^{C\times d'}$ contains aggregated contextual features. This weighted summation enables selective feature aggregation, enhancing important signals while suppressing noise, particularly effective for handling irrelevant or redundant information in sensor networks.

Finally, feature fusion is achieved through GRU units:

$$z = \sigma(W_z[\text{concat}[H, C]] + b_z), \tag{15}$$

$$r = \sigma(W_r[\text{concat}[H, C]] + b_r), \tag{16}$$

$$\hat{h} = \tanh(W_h[\text{concat}[r \odot H, C]] + b_h), \tag{17}$$

$$H_{\text{new}} = (1 - z) \odot H + z \odot \hat{h}, \tag{18}$$

where $W_z, W_r, W_h \in \mathbb{R}^{(d+d')\times d}$ are GRU parameter matrices, $b_z, b_r, b_h \in \mathbb{R}^d$ are bias terms, $z \in \mathbb{R}^{C\times d}$ is the update gate, $r \in \mathbb{R}^{C\times d}$ is the reset gate, $\hat{h} \in \mathbb{R}^{C\times d}$ represents candidate states, and $H_{\text{new}} \in \mathbb{R}^{C\times d}$ contains updated node features. This gated mechanism balances historical and neighbor information more effectively than direct feature replacement, enabling more stable feature updates in sensor network scenarios.

Following node feature aggregation, the model proceeds to the graph classification stage, which aims to transform local node features into graph-level representations with global characterization capability. Considering the heterogeneity of node features in railway switch multi-sensor systems, a simple averaging operation is adopted to obtain global features, avoiding additional parameter bias while establishing a stable baseline for subsequent fine-grained feature fusion: $h_G = \frac{1}{C}\sum_{i=1}^{C} h_i$.

To establish interactions between global and local features, the model employs a feature expansion strategy. The global feature h_G is expanded along the node dimension to match the original feature matrix H, preserving global context while establishing node–global correspondences:

$$h_G^{\text{expand}} = h_G.\text{unsqueeze}(0).\text{expand}(C, -1). \tag{19}$$

Through concatenation, the expanded global features and original node features form a composite representation containing multi-scale information. This dual-path fusion mechanism captures both the overall network state and local details, providing rich information for attention weight computation:

$$F_{\text{concat}} = \text{concat}[h_G^{\text{expand}}, H]. \tag{20}$$

Building upon feature interactions, the model introduces learnable attention mechanisms to quantify each node's contribution to fault diagnosis. Nonlinear transformations and normalization generate physically meaningful node importance weights, enabling adaptive focus on sensors most relevant to current fault patterns:

$$\boldsymbol{\beta} = \text{softmax}(W_\beta F_{\text{concat}} + b_\beta). \tag{21}$$

Attention-weighted feature fusion produces enhanced global representations, implementing soft selection in feature space that emphasizes critical nodes while suppressing irrelevant noise:

$$\tilde{h}_G = \sum(\boldsymbol{\beta} \odot H, \text{dim} = 0). \tag{22}$$

To mitigate potential information loss from simple weighted fusion, we retain the GRU gating mechanisms to adaptively integrate the initial global feature vector h_G with the attention-fused feature vector $\tilde{h}_G$. The GRU processing utilizes h_G as the state vector and $\tilde{h}_G$ as the input vector, where the update and reset gates dynamically regulate feature blending proportions while maintaining gradient stability. The enhanced feature representation h_G^{final} is obtained through this gated fusion process.

Finally, the optimized global features feed into fully-connected layers for fault type probability prediction. The end-to-end training process fully leverages multi-sensor cooperative monitoring advantages, automatically learning mappings from raw signals to fault categories:

$$p(y|X) = \text{softmax}(W_c h_G^{\text{final}} + b_c), \tag{23}$$

where $W_c \in \mathbb{R}^{d\times m}$ denotes the weight matrix of the classification layer mapping the final graph representation to class logits, $b_c \in \mathbb{R}^m$ represents the bias vector, and m is the number of fault categories.

This hierarchical feature refinement mechanism — from global averaging through attention fusion to gated updates — progressively extracts fault characteristics from sensor networks, significantly improving recognition accuracy for complex failure modes.

3. Experimental study and results analysis

3.1. Experimental setup

3.1.1. Cross-category small-sample experimental setup

This study designs 2-class, 3-class, and 4-class testing scenarios to construct a stringent cross-category evaluation framework, simulating the actual diagnostic scenarios of railway multi-sensor monitoring environments. This experimental setup is consistently applied across all datasets evaluated in this work, ensuring direct comparability of results between different data sources. The framework requires the model to rapidly tune its parameters during the testing phase using only k small samples of each new class ($k \in \{3, 5, 10\}$), evaluate its robustness using all remaining samples of each class, and achieve full-process data fusion through nine-channel multi-sensor measurements. The experiments adopt a three-level category isolation protocol, where 16 fault classes are first divided into base classes (corresponding to 14/13/12 classes for 2/3/4-class scenarios, respectively) and testing

Table 1
Experimental task design for cross-category small-sample evaluation.

Task	Testing scenarios	Small samples per testing class	Evaluation samples (testing set)	
			Case I	Case II
T1	C, e	3	83	79
T2	C, e	5	79	75
T3	C, e	10	69	65
T4	B, b, h	3	159	159
T5	B, b, h	5	153	153
T6	B, b, h	10	138	138
T7	a, B, c, D	3	191	192
T8	a, B, c, D	5	183	184
T9	a, B, c, D	10	163	164
T10	A, a, b, c, g	3	262	283
T11	A, a, b, c, g	5	252	273
T12	A, a, b, c, g	10	227	248

classes (2/3/4 classes). The base classes are further split into strictly disjoint training and validation sets, where the number of validation set classes matches the number of testing classes. This dedicated validation partition serves exclusively for hyperparameter optimization during model development, ensuring the testing set remains untouched for final unbiased evaluation. During the training phase, the model's adaptability is developed through dynamically constructed meta-tasks. For each meta-task, N classes ($N \in \{2,3,4\}$, strictly corresponding to the testing scenarios) are randomly sampled from the training set using stratified random sampling, which ensures diversity in the selected classes. This sampling mechanism naturally introduces class shifts across tasks, with the overlap probability of consecutive task classes following $P \sim \mathrm{Bin}(N, N/|Y_{\mathrm{train}}|)$ distribution. Each meta-task consists of $N \times K$ support samples ($K \in \{3,5,10\}$) and $N \times 15$ query samples, adhering to the standard few-shot learning settings [28]. To ensure the reliability of training, the support set and query set samples are mutually exclusive, the sample pool is reset at the beginning of each training epoch, and stratified random sampling is applied. The model is trained over 30 epochs (each epoch includes 20 tasks) using the MAML framework (with 2-step inner-loop gradient updates) and the Adam optimizer ($\beta = 10^{-3}$). The testing protocol involves adapting the model using k small samples of each testing class and evaluating diagnostic performance with all remaining samples of each class, where the adaptation process is performed using the Adam optimizer ($\beta = 10^{-4}$). The validation phase fully replicates the testing protocol, including the number of classes and sample configurations, but uses disjoint fault classes. Specifically, the model is fine-tuned on the k-shot support set of the validation set and then evaluated on all remaining samples of each class. The resulting performance metrics are used as criteria for hyperparameter optimization. The complete configurations of the twelve experimental tasks (T1-T12) on the two case datasets are detailed in Table 1, including the specific fault categories randomly sampled through stratified random sampling (e.g., two particular fault categories randomly selected for the 2-class testing scenario) and the corresponding sample sizes for small-sample adaptation/large-sample evaluation. This systematic design ensures all experimental tasks strictly adhere to the predefined random sampling protocols, thereby establishing a rigorous benchmark for evaluating the model's cross-category small-sample learning capability in practical railway maintenance scenarios.

3.1.2. Compared methods

To rigorously validate the superiority of our proposed GMSIFN method, we conducted comprehensive comparisons with three categories of state-of-the-art approaches, covering generic graph/meta-learning baselines, recent advanced RTA and related industrial small-sample fault diagnosis methods, and representative continual learning frameworks relevant to evolving industrial fault scenarios. The compared methods are categorized and described as follows:

(1) Graph representation learning and meta-learning paradigms relevant to small-sample classification:

Graph Attention Network (GAT) [29]: Leverages masked self-attentional layers to dynamically compute adaptive node weights. This architecture assigns importance to neighboring nodes through attention mechanisms, enabling discriminative feature aggregation without explicit graph constraints.

Graph Transformer (GraphTrans) [30]: Integrates Transformer-based self-attention with GNNs to model long-range dependencies. Employs a global readout token to optimize pairwise node interactions, enhancing structural awareness for graph-level tasks.

Model-Agnostic Meta-Learning (MAML) [25]: Gradient-based framework that learns transferable model initializations. Optimizes parameters for rapid adaptation to new tasks through minimal gradient updates on limited support data.

Meta-Molecular Graph Neural Network (Meta-MGNN) [31]: Combines pre-trained GNNs with meta-learning and self-supervised auxiliary tasks to strengthen small-sample generalization capabilities.

Meta-Graph Attention Network (Meta-GAT) [32]: Extends GAT with bilevel meta-learning to transfer knowledge from auxiliary tasks, reducing sample complexity through gradient-based optimization.

Meta-Learning with Discriminant Space Optimization (MLDSO) [33]: Enhances metric spaces using discriminant boundary and clustering losses, improving robustness under noisy and variable conditions.

(2) Recent advanced RTA fault diagnosis and related industrial methods relevant to small-sample fault diagnosis:

Fault Diagnosis Considering Sensor Abnormality Network (FD-CSANet) [34]: Utilizes a Sensor Information Aggregate Module (SIAM) to integrate multi-channel sensor signals, coupled with a DBEM backbone for feature extraction and classification, enabling accurate RTA fault diagnosis.

Input Spatial Over-parameterized Network (ISONet) [35]: Employs input spatial over-parameterization to enhance feature learning capacity, achieving high accuracy in industrial small-sample fault diagnosis tasks such as aero-engine inter-shaft bearing fault classification.

(3) Representative continual learning Methods:

Continually Evolved Classifier (CEC) [36]: Adopts a decoupled training strategy to separate representation learning and classifier learning, and employs an attention-based adaptation module to update classifiers with global contextual information, enabling small-sample continual learning without catastrophic forgetting.

Elastic Weight Consolidation (EWC) [37]: Implements a parameter regularization mechanism that selectively slows updates to weights important for previous tasks by using the Fisher information matrix to measure parameter importance, thereby mitigating catastrophic forgetting in continual learning.

All compared methods retained their original configurations, with input layers modified solely for data compatibility. For generic graph/meta-learning baselines, graph-based models (GAT and GraphTrans) converted raw vibration signals into K-Nearest Neighbor (KNN) graphs, where the neighborhood size k was optimized for each method via grid search on validation sets. Non-meta baselines were trained using standard batch learning without episodic splitting, and meta-learning baselines (MAML, Meta-MGNN, Meta-GAT, MLDSO) strictly followed canonical episodic training protocols. For recent advanced RTA and related industrial small-sample fault diagnosis methods, the original multi-channel RTA signals were directly fed into the dimension-adjusted multi-channel input layers without any additional processing. For representative continual learning frameworks relevant to evolving industrial fault scenarios, although the experiments did not explicitly specify multi-stage continual learning phases, the employed diagnostic procedure—initial training on base classes followed by fine-tuning on novel classes (testing classes)—can be naturally interpreted as a two-stage continual learning workflow, and multi-stage algorithms were adapted accordingly by treating the process as consisting of a base-class training phase and a subsequent novel-class fine-tuning phase

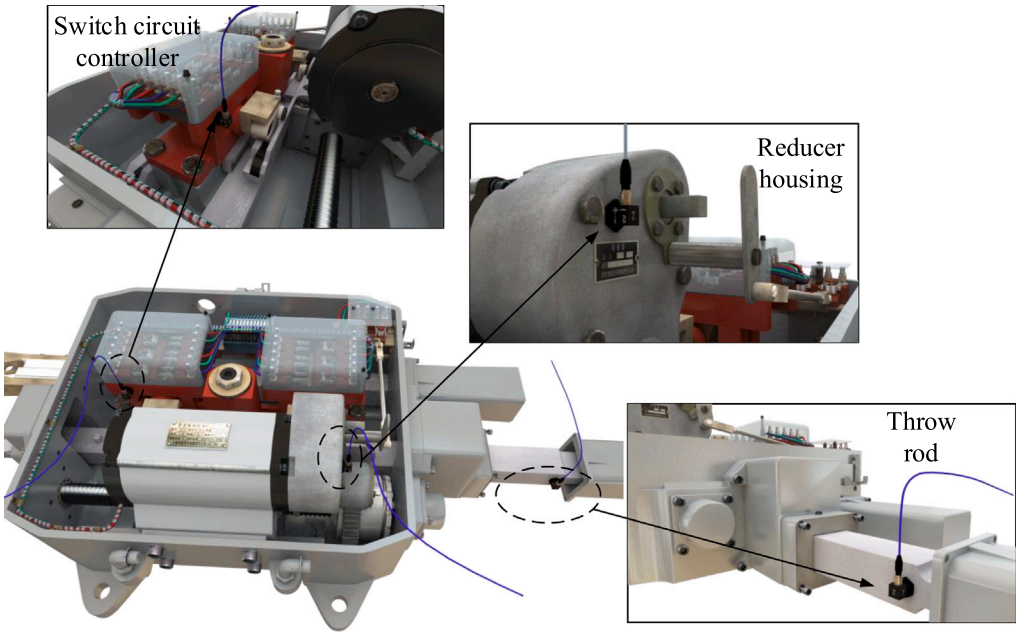

Fig. 5. Dataset from ZDJ9-RTA deployed on a dedicated railway test line.

prior to evaluation. Hyperparameters for all methods were tuned using validation sets, and final performances were reported after completing small-sample fine-tuning on the test samples. This comparative setup, including all baseline implementations and hyperparameter tuning strategies described above, is consistently applied across all datasets used in this work to ensure fair and direct performance comparison.

3.2. Case I: Railway test line ZDJ9-RTA dataset

This case employs a multi-sensor vibration dataset from ZDJ9 railway turnout actuators (collected from the railway test line provided by the Xiangyang Signaling and Communication Section, Wuhan Railway Bureau) to address the limitations of single-sensor paradigms in small-sample diagnostic research. As shown in Fig. 5, a synchronized triaxial sensor array was deployed at three critical mechanical interfaces (reducer housing, throw rod side, and switch controller side) to acquire nine-channel vibration signals at a 512 Hz sampling rate, yielding 5120 temporal samples per channel during each 10-second operational cycle. As shown in Table 2 and Fig. 6, the dataset contains 16 fault categories derived from 8 primary fault states with bidirectional variants, including three force-dependent anomalies (3 kN underdriving, 5 kN/6 kN/8 kN overdriving), two mechanical failures (unlocking failure and sudden indication loss during approach movement), and no-load operation conditions. The 804 carefully annotated samples (40–60 per category) demonstrate directional characteristics due to asymmetric mechanical loading during normal-to-reverse and reverse-to-normal transitions. The dataset offers two key advantages: (1) nine-channel synchronous acquisition enables spatiotemporal correlation capture for graph neural network-based multi-sensor fusion, and (2) the 16 diverse fault categories provide an ideal platform to evaluate small-sample generalization across varying cross-category scenarios and novel fault conditions.

3.2.1. Hyperparameter selection for the GMSIFN

To optimize the hyperparameters of the proposed GMSIFN, we employed a two-stage process combining Bayesian [38] optimization and grid search. First, Bayesian optimization was used to determine preliminary search ranges for two critical hyperparameters: d, which defines the dimensionality for signal compression, and k, the preset neighbor count used for graph construction. Ranges identified through Bayesian optimization were subsequently refined using grid search over $d \in \{120, 240, 360, 480, 600\}$ and $k \in \{2, 3, 4, 5, 6, 7, 8\}$, both being integers.

Table 2
Detailed description of railway test line ZDJ9-RTA dataset.

RTA state	Normal-to-reverse	Number	Reverse-to-normal	Number
Underdriving (3 kN)	A	55	a	55
Fault-free (4 kN)	B	59	b	59
Overdriving (5 kN)	C	49	c	49
Overdriving (6 kN)	D	40	d	40
Overdriving (8 kN)	E	40	e	40
No load	F	50	f	50
Unlocking failure	G	60	g	59
Sudden indication loss after approaching movement	H	49	h	50

To ensure the robustness of the optimal hyperparameter selection, the results presented in Fig. 7 represent the average class-wise accuracy across five experiments for each task. Although the testing set's fault categories were fixed, the validation and training sets were randomly sampled during each iteration. This randomness in sampling helped assess the adaptability of the d-k combinations under diverse cross-category diagnostic conditions.

The hyperparameter sensitivity analysis clearly demonstrates that model performance varies significantly across tasks, depending on the selected combinations of d and k, where for example Task T1 achieved its best accuracy (82.56%) with $d = 360$ and $k = 4$, while Task T6 reached an optimal performance (99.86%) with $d = 360$ and $k = 2$, and similarly Task T9 peaked at $d = 360$ and $k = 5$ with 97.11% accuracy whereas Task T12 achieved 83.91% at $d = 360$ and $k = 7$, with these task-specific variations revealing systematic patterns where excessively large d values (e.g., $d = 600$) introduce redundant representations that dilute key signal features while increasing computational complexity, whereas overly small d (e.g., $d = 120$) leads to inadequate feature extraction that limits temporal information representation, and similarly overlarge k values (e.g., $k = 8$) aggregate noise from distant neighbors that deteriorates diagnostic accuracy while excessively small k (e.g., $k = 2$) restricts neighborhood representation causing loss of valuable local relationships.

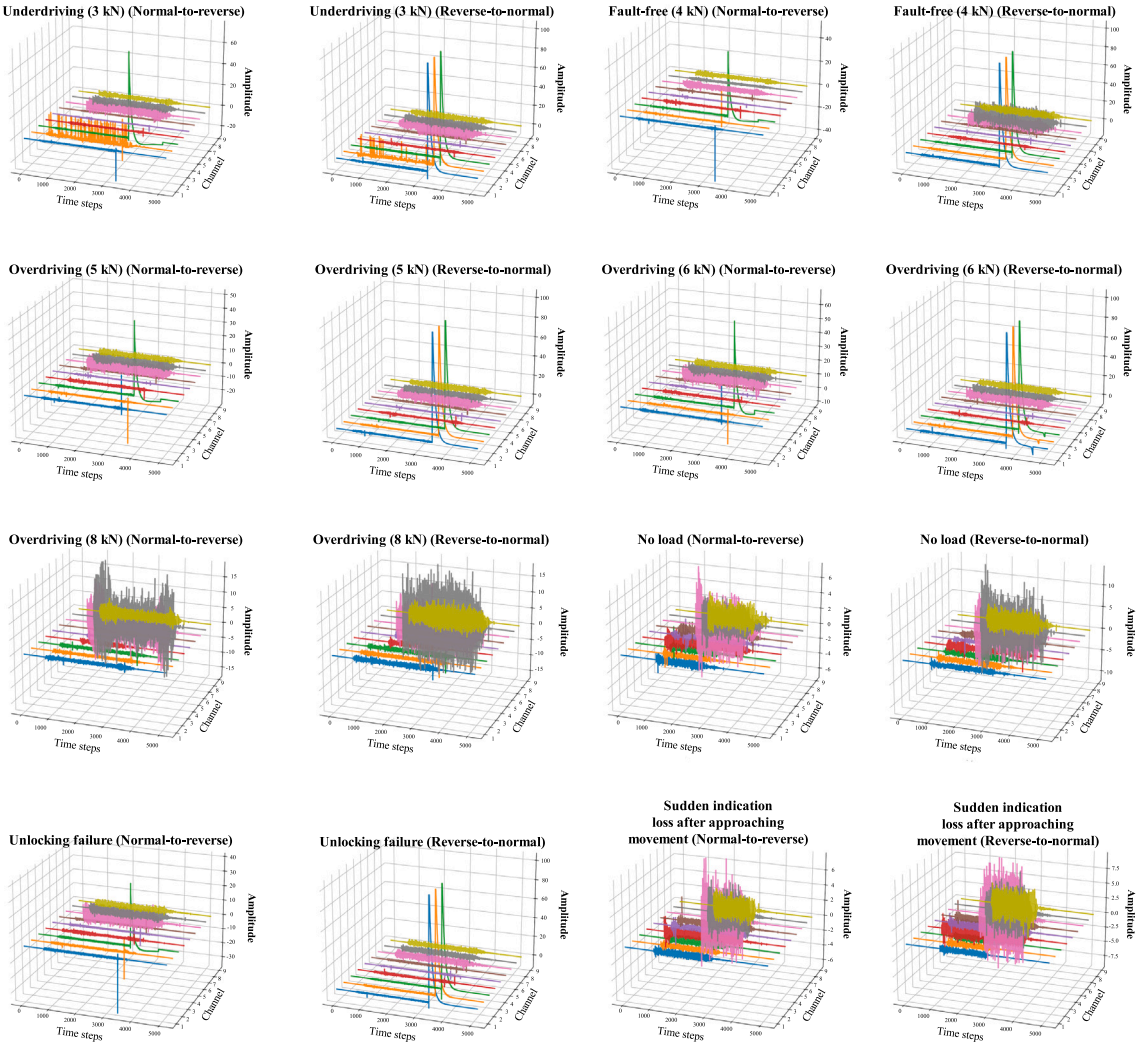

Fig. 6. Visualization of nine-channel 16 types of railway test line ZDJ9-RTA fault signals.

In conclusion, the hyperparameter sensitivity results confirm that the selection of d and k significantly affects GMSIFN's diagnostic performance. By incorporating randomness during validation dataset construction across five experiment iterations, we ensured that the optimal d-k configurations exhibit excellent cross-category adaptability. This systematic approach underscores the robustness and effectiveness of the proposed model, achieving competitive performance in complex multi-sensor fault diagnosis environments.

3.2.2. Comparative studies

Comprehensive evaluation across 12 cross-category small-sample tasks in our RTA dataset demonstrates the superior performance of our proposed GMSIFN against three categories of state-of-the-art approaches — generic graph/meta-learning baselines, recent advanced RTA and related industrial small-sample fault diagnosis methods, and representative continual learning frameworks relevant to evolving industrial fault scenarios — using three established metrics: Accuracy for overall diagnostic correctness, F1-score [39] to harmonize precision and recall, and Area Under the ROC Curve (AUROC) [40] evaluating class-separation capacity, hereafter abbreviated as ACC, F1, and AUC for conciseness. All results in Tables 3 and 4 reflect averages and standard deviations from ten independent runs, where each run dynamically resampled training categories to simulate real-world task variability while fixing test categories, ensuring statistically robust validation under authentic railway maintenance conditions. The mean performance metrics across all tasks are additionally visualized in Fig. 8, providing a comprehensive performance overview.

Analysis reveals two consistent trends: diagnostic performance declines as task complexity escalates, exemplified when transitioning from T1 (2 novel faults) to T10 (5 novel faults), where GAT's accuracy drops from 70.22% to 47.52% and Meta-MGNN declines from 83.73% to 46.43%, highlighting the intrinsic challenge of generalizing to expanding novel fault patterns with small samples; conversely, enlarging support sets consistently enhances performance, as our method achieves 96.01% ACC in T6 with 10 samples per testing class adaptation compared to 90.56% in the 3-sample T4 task, underscoring sample adequacy as critical to small-sample success.

Compared to generic graph/meta-learning baselines, GMSIFN exhibits decisive advantages. Against non-meta methods GAT and GraphTrans—which train conventionally on all training categories before fine-tuning—we achieve significant accuracy improvements, with gains exceeding 20% against GraphTrans in all 12 tasks and against GAT in 75% of tasks (9/12). Specifically, maximum gains reach 46.78% (T1) against GraphTrans and 26.73% (T1) against GAT, though T7-T12 show more moderate gains (6.44%–17.45%). This superiority stems from our category-shift meta-learning strategy, which explicitly simulates dynamic fault distributions during training to enable rapid cross-category adaptation, a capability absent in static-training paradigms. Even against meta-graph baselines Meta-GAT and Meta-MGNN, GM-

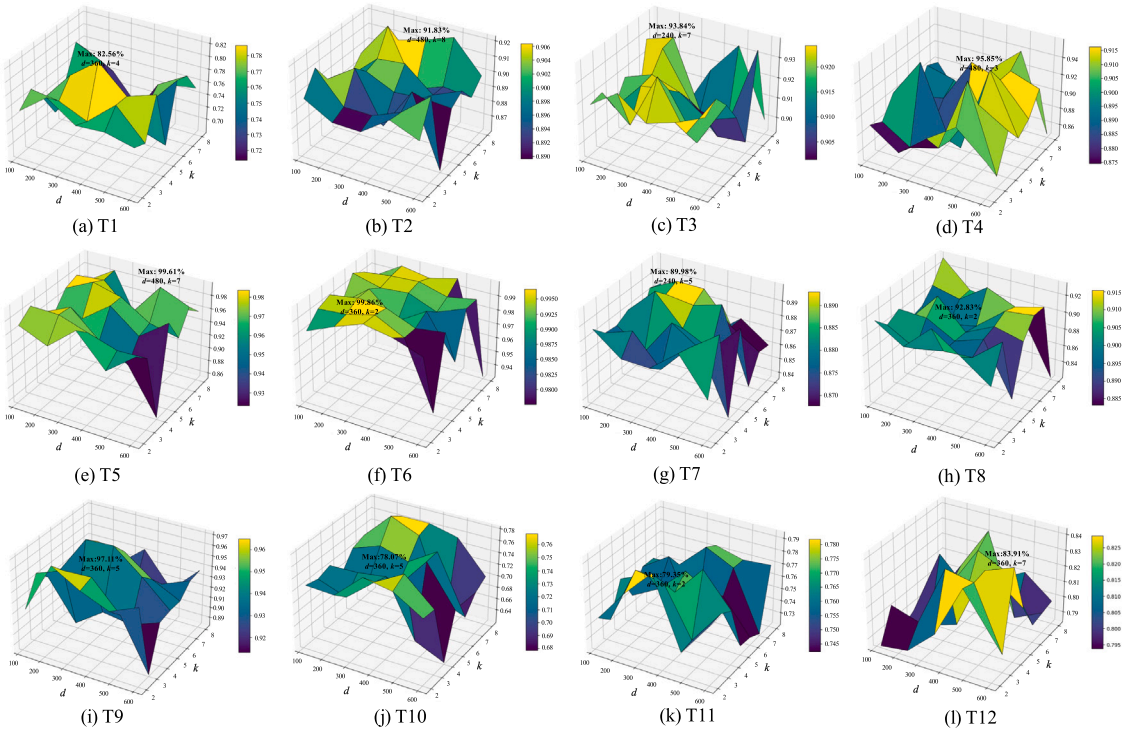

Fig. 7. Hyperparameter sensitivity analysis of the GMSIFN model.

Table 3
Experimental results with generic graph/meta-Learning baselines on RTA dataset (%).

Task No.	Metric	GAT	GraphTrans	Meta-GAT	Meta-MGNN	MAML	MLDSO	Ours
T1	ACC	70.22 ± 4.35	50.17 ± 3.88	77.82 ± 14.88	83.73 ± 7.84	86.99 ± 1.31	39.21 ± 7.89	**96.95 ± 0.12**
	F1	70.71 ± 3.84	33.33 ± 2.56	77.26 ± 15.53	90.14 ± 5.32	85.18 ± 1.63	22.60 ± 8.80	**96.93 ± 0.14**
	AUC	73.45 ± 1.03	57.94 ± 8.48	78.75 ± 18.11	81.67 ± 4.23	90.23 ± 1.22	32.17 ± 4.61	**98.79 ± 0.29**
T2	ACC	76.02 ± 2.84	58.67 ± 5.63	79.18 ± 9.85	91.14 ± 5.97	88.14 ± 1.01	41.20 ± 8.03	**98.28 ± 0.26**
	F1	75.11 ± 1.25	45.89 ± 4.86	77.79 ± 8.62	90.58 ± 6.52	86.93 ± 0.76	36.03 ± 6.53	**98.26 ± 0.28**
	AUC	83.26 ± 4.02	65.30 ± 3.80	85.27 ± 6.47	89.93 ± 7.31	95.57 ± 0.63	43.22 ± 3.87	**97.26 ± 0.38**
T3	ACC	88.55 ± 2.42	60.86 ± 2.87	81.45 ± 5.01	93.67 ± 5.06	93.55 ± 3.99	57.23 ± 5.24	**99.95 ± 0.02**
	F1	84.08 ± 1.62	53.55 ± 3.02	79.76 ± 8.76	93.64 ± 5.72	92.73 ± 2.84	55.48 ± 3.04	**99.63 ± 0.17**
	AUC	90.59 ± 2.23	69.38 ± 4.84	90.13 ± 6.06	92.91 ± 4.53	96.42 ± 3.08	66.89 ± 2.18	**99.98 ± 0.01**
T4	ACC	72.65 ± 17.93	32.39 ± 2.83	66.35 ± 7.23	74.84 ± 13.21	72.19 ± 7.14	29.88 ± 6.16	**90.56 ± 1.89**
	F1	71.72 ± 18.95	15.92 ± 2.43	62.77 ± 9.96	76.85 ± 13.71	71.33 ± 6.12	27.43 ± 5.87	**90.46 ± 1.78**
	AUC	86.95 ± 5.33	39.09 ± 1.94	84.83 ± 3.98	88.29 ± 6.17	88.27 ± 5.87	30.13 ± 4.63	**94.71 ± 1.22**
T5	ACC	78.11 ± 5.56	35.35 ± 2.94	68.48 ± 4.71	72.22 ± 2.29	80.72 ± 2.75	35.85 ± 4.05	**93.80 ± 3.59**
	F1	78.16 ± 5.87	17.89 ± 2.52	70.48 ± 5.79	73.58 ± 5.84	79.57 ± 3.83	31.77 ± 3.54	**92.56 ± 2.87**
	AUC	89.18 ± 4.87	50.53 ± 2.26	87.42 ± 7.93	90.87 ± 4.23	90.05 ± 1.98	45.87 ± 4.15	**95.03 ± 1.56**
T6	ACC	85.14 ± 11.24	42.25 ± 3.26	73.85 ± 6.54	80.07 ± 1.81	85.87 ± 4.51	45.88 ± 3.66	**96.01 ± 0.37**
	F1	84.76 ± 11.64	25.91 ± 2.44	73.98 ± 6.61	81.59 ± 1.75	68.48 ± 4.71	45.50 ± 3.80	**96.04 ± 0.39**
	AUC	90.17 ± 8.92	59.52 ± 3.58	90.11 ± 4.42	91.36 ± 3.74	93.24 ± 3.96	55.94 ± 0.74	**99.23 ± 0.14**
T7	ACC	56.55 ± 10.48	24.35 ± 4.98	58.91 ± 2.12	54.19 ± 6.02	54.45 ± 5.24	26.86 ± 1.63	**73.03 ± 6.55**
	F1	54.43 ± 12.37	9.78 ± 3.51	54.57 ± 1.79	55.30 ± 10.10	48.57 ± 6.62	26.84 ± 1.62	**71.58 ± 6.14**
	AUC	81.43 ± 5.76	38.17 ± 14.92	83.32 ± 5.45	80.31 ± 5.94	76.33 ± 4.81	36.91 ± 1.55	**86.88 ± 4.58**
T8	ACC	57.59 ± 8.53	26.32 ± 5.19	63.11 ± 3.56	55.46 ± 6.29	62.30 ± 3.90	27.12 ± 0.71	**75.04 ± 6.19**
	F1	54.10 ± 8.80	10.80 ± 3.33	62.62 ± 3.33	59.67 ± 8.59	61.34 ± 2.91	26.09 ± 0.23	**72.36 ± 6.90**
	AUC	79.22 ± 5.81	42.09 ± 7.38	83.87 ± 0.97	84.50 ± 5.86	83.14 ± 3.01	38.25 ± 0.28	**90.95 ± 1.37**
T9	ACC	73.32 ± 7.06	26.99 ± 3.07	71.04 ± 6.74	62.46 ± 2.32	71.78 ± 9.82	27.99 ± 0.42	**79.76 ± 3.69**
	F1	71.34 ± 8.61	11.57 ± 2.33	71.49 ± 7.24	62.31 ± 3.06	70.87 ± 9.94	27.97 ± 0.41	**79.59 ± 3.57**
	AUC	90.01 ± 2.19	48.91 ± 2.19	82.31 ± 6.48	89.92 ± 5.42	81.19 ± 8.48	38.76 ± 0.37	**93.34 ± 1.08**
T10	ACC	47.52 ± 9.73	20.61 ± 0.76	45.44 ± 9.73	46.43 ± 5.16	37.93 ± 13.55	18.15 ± 0.85	**55.54 ± 5.92**
	F1	42.15 ± 11.26	7.05 ± 0.48	41.88 ± 10.68	47.41 ± 3.44	37.99 ± 12.18	18.13 ± 0.86	**52.44 ± 6.18**
	AUC	71.79 ± 5.85	32.26 ± 1.15	67.25 ± 7.05	74.98 ± 2.86	48.63 ± 9.97	23.19 ± 1.82	**78.68 ± 3.01**
T11	ACC	50.99 ± 2.18	19.45 ± 1.99	49.65 ± 7.98	50.78 ± 6.89	43.77 ± 9.73	18.95 ± 0.32	**58.96 ± 4.17**
	F1	47.15 ± 1.82	6.38 ± 1.19	49.47 ± 6.90	49.01 ± 5.73	41.60 ± 12.31	18.92 ± 0.34	**57.24 ± 5.59**
	AUC	71.48 ± 3.24	49.44 ± 4.12	72.97 ± 1.02	70.92 ± 6.96	69.07 ± 8.62	25.43 ± 0.28	**79.73 ± 3.61**
T12	ACC	55.07 ± 3.53	19.39 ± 2.21	55.07 ± 1.77	53.30 ± 0.44	50.66 ± 9.26	19.95 ± 0.12	**64.08 ± 4.81**
	F1	50.06 ± 7.39	6.35 ± 1.31	55.09 ± 3.50	51.87 ± 1.43	44.05 ± 10.68	19.93 ± 0.14	**60.22 ± 2.96**
	AUC	70.74 ± 4.13	53.01 ± 6.10	77.09 ± 2.33	78.20 ± 3.46	67.74 ± 7.03	28.13 ± 0.39	**83.01 ± 2.74**

Table 4
Performance comparison of recent advanced RTA-related methods and continual learning methods on RTA dataset (%).

Task No.	Metric	FDCSANet	ISONet	CEC	EWC	Ours
T1	ACC	87.21 ± 3.67	78.35 ± 2.57	91.43 ± 2.81	92.17 ± 3.01	**96.95 ± 0.12**
	F1	85.41 ± 3.05	75.91 ± 5.83	90.89 ± 3.82	92.04 ± 3.15	**96.93 ± 0.14**
	AUC	89.32 ± 2.88	88.38 ± 2.81	91.61 ± 1.77	94.57 ± 2.51	**98.79 ± 0.29**
T2	ACC	91.54 ± 1.63	91.77 ± 1.88	93.57 ± 4.43	95.73 ± 1.02	**98.28 ± 0.26**
	F1	89.66 ± 2.43	91.09 ± 1.93	94.49 ± 4.52	94.74 ± 0.98	**98.26 ± 0.28**
	AUC	92.27 ± 1.96	95.04 ± 1.77	96.61 ± 3.39	96.28 ± 0.21	**97.26 ± 0.38**
T3	ACC	94.78 ± 1.78	92.31 ± 3.89	95.65 ± 4.35	98.28 ± 0.73	**99.95 ± 0.02**
	F1	95.41 ± 2.14	91.32 ± 3.25	95.57 ± 4.44	98.13 ± 0.69	**99.63 ± 0.17**
	AUC	96.28 ± 2.16	96.26 ± 2.56	97.75 ± 2.26	99.46 ± 0.03	**99.98 ± 0.01**
T4	ACC	48.43 ± 13.21	64.21 ± 8.24	82.78 ± 4.23	78.49 ± 8.57	**90.56 ± 1.89**
	F1	45.20 ± 9.85	59.41 ± 13.09	81.99 ± 3.26	76.22 ± 9.09	**90.46 ± 1.78**
	AUC	73.05 ± 8.38	88.99 ± 1.34	90.62 ± 2.73	83.41 ± 7.05	**94.71 ± 1.22**
T5	ACC	66.67 ± 9.81	78.43 ± 3.92	85.48 ± 3.02	86.05 ± 5.77	**93.80 ± 3.59**
	F1	63.81 ± 13.39	75.76 ± 5.46	84.48 ± 3.61	85.36 ± 4.53	**92.56 ± 2.87**
	AUC	74.88 ± 4.12	92.16 ± 3.37	92.13 ± 4.24	92.17 ± 3.55	**95.03 ± 1.56**
T6	ACC	69.57 ± 8.69	87.32 ± 5.44	89.46 ± 2.28	91.67 ± 3.26	**96.01 ± 0.37**
	F1	65.12 ± 10.76	86.97 ± 5.75	90.32 ± 3.22	91.59 ± 3.31	**96.04 ± 0.39**
	AUC	84.79 ± 2.85	94.82 ± 4.12	98.24 ± 1.55	96.39 ± 1.93	**99.23 ± 0.14**
T7	ACC	37.96 ± 8.12	54.71 ± 4.45	67.51 ± 2.25	64.66 ± 2.36	**73.03 ± 6.55**
	F1	29.31 ± 5.63	52.58 ± 5.28	63.08 ± 1.95	61.97 ± 2.45	**71.58 ± 6.14**
	AUC	65.76 ± 14.57	81.19 ± 6.08	77.52 ± 2.41	79.35 ± 1.51	**86.88 ± 4.58**
T8	ACC	63.93 ± 7.65	64.21 ± 1.82	72.36 ± 2.91	72.12 ± 3.81	**75.04 ± 6.19**
	F1	53.59 ± 8.04	63.23 ± 2.24	70.76 ± 3.35	67.73 ± 6.54	**72.36 ± 6.90**
	AUC	75.60 ± 11.36	85.53 ± 3.86	84.39 ± 4.82	84.84 ± 2.63	**90.95 ± 1.37**
T9	ACC	64.51 ± 6.84	71.17 ± 6.75	75.87 ± 2.81	78.84 ± 2.15	**79.76 ± 3.69**
	F1	58.77 ± 13.52	64.08 ± 6.85	74.37 ± 4.23	75.12 ± 1.17	**79.59 ± 3.57**
	AUC	81.55 ± 5.35	88.36 ± 3.13	87.51 ± 3.16	90.61 ± 0.84	**93.34 ± 1.08**
T10	ACC	26.79 ± 8.54	37.67 ± 8.44	47.19 ± 4.94	31.87 ± 7.06	**55.54 ± 5.92**
	F1	17.14 ± 11.25	35.41 ± 7.93	45.62 ± 6.99	20.68 ± 6.75	**52.44 ± 6.18**
	AUC	61.27 ± 11.28	69.04 ± 5.44	71.38 ± 6.17	65.84 ± 7.89	**78.68 ± 3.01**
T11	ACC	32.45 ± 14.54	43.89 ± 4.49	55.47 ± 7.03	36.25 ± 6.88	**58.96 ± 4.17**
	F1	23.94 ± 17.91	41.21 ± 5.39	50.93 ± 6.35	27.37 ± 6.44	**57.24 ± 5.59**
	AUC	68.61 ± 5.22	72.73 ± 8.21	75.73 ± 3.52	74.51 ± 3.56	**79.73 ± 3.61**
T12	ACC	58.59 ± 2.64	49.78 ± 12.78	62.02 ± 4.59	55.76 ± 4.57	**64.08 ± 4.81**
	F1	57.66 ± 2.74	45.24 ± 12.21	59.84 ± 3.25	48.39 ± 3.08	**60.22 ± 2.96**
	AUC	71.26 ± 2.31	75.57 ± 5.79	78.42 ± 3.18	80.72 ± 2.12	**83.01 ± 2.74**

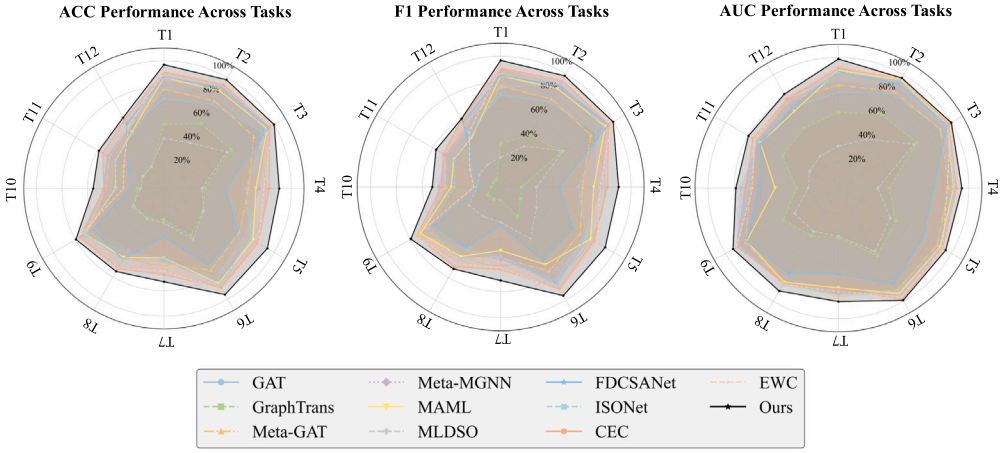

Fig. 8. Visualization of small-sample diagnostic results on railway test line ZDJ9-RTA dataset.

SIFN sustains >10% margins (e.g., 99.95% vs. 81.45% in T3) due to two innovations: our unsupervised GGL dynamically constructs sensor topologies from raw signals, eliminating reliance on predefined graphs misaligned with electromechanical systems, while the GRU-gated graph attention hierarchically fuses neighbor and central node features—preserving critical fault-propagation paths overlooked by Meta-MGNN's mean-pooling and Meta-GAT's flat aggregation. Pure meta-learning baselines MAML and MLDSO exhibit severe limitations despite sharing our training paradigm, as MAML lags by 12.19% accuracy in T4 (72.19% vs. 90.56%), while MLDSO collapses catastrophically (F1 ≤27.97% in T9), exposing their inability to model spatial relationships.

Compared to recent advanced RTA and related industrial small-sample fault diagnosis methods such as FDCSANet and ISONet, our GMSIFN addresses key shortcomings inherent to these designs. FDCSANet is tailored for RTA diagnostics and incorporates a Sensor Information Aggregate Module to mitigate abnormal sensor readings, yet it lacks explicit modeling of multi-sensor spatial dependencies and does not address rapid adaptation to previously unseen fault categories,

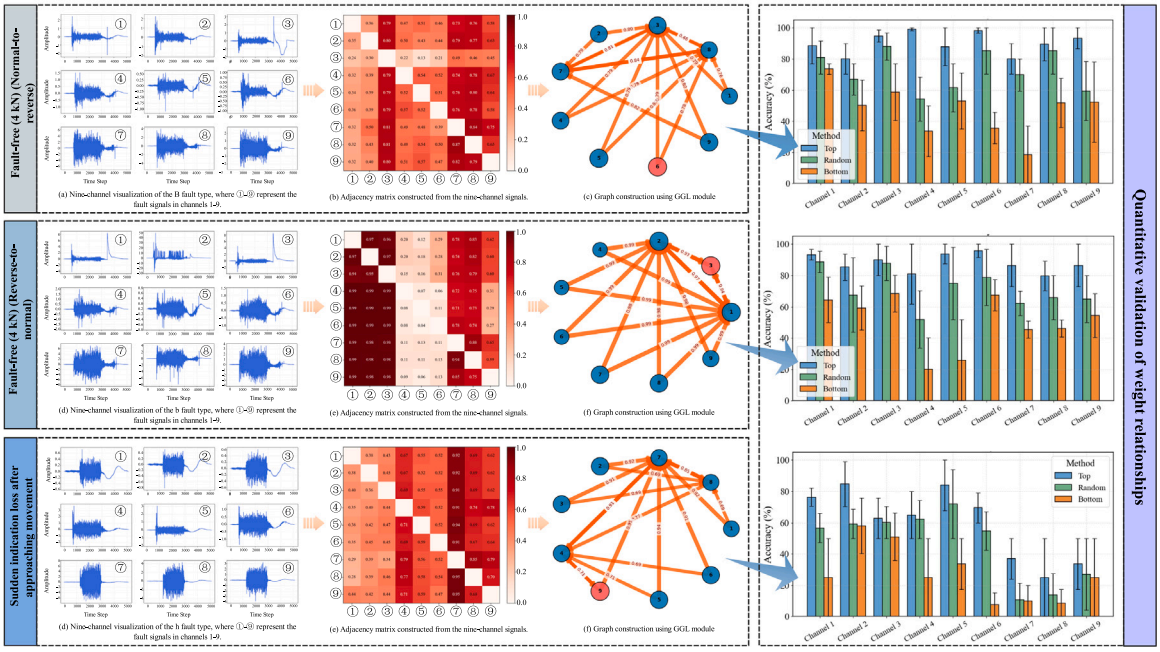

Fig. 9. Quantitative validation and visualization of graph data construction in T6 for Fault-free (4 kN) (Normal-to-reverse), Fault-free (4 kN) (Reverse-to-normal), and Sudden indication loss after approaching movement (Reverse-to-normal) fault types.

resulting in performance degradation under cross-category conditions (e.g., T4 ACC: 48.43% vs. our 90.56%). ISONet, while effective for small-sample fault diagnosis in aero-engine scenarios through input spatial over-parameterization, similarly omits spatial topology modeling among sensors and assumes static fault category sets, thereby limiting adaptability to novel RTA faults. Across all tasks, GMSIFN consistently outperforms these methods in ACC, F1, and AUC; for instance, in T6 we achieve 96.01% ACC compared to FDCSANet's 69.57% and ISONet's 87.32%, validating the benefit of fusing dynamic graph construction with meta-learning for small-sample rapid adaptation.

Compared to representative continual learning frameworks such as CEC and EWC, our approach focuses exclusively on fast adaptation to novel classes without retaining knowledge of previously seen classes. Continual learning methods are designed to balance preservation of old-class accuracy with acquisition of new-class performance, inherently introducing trade-offs that slow adaptation to novel faults. EWC employs parameter regularization via the Fisher information matrix to retain important weights, but this retention constrains parameter updates, reducing flexibility for new-class optimization (e.g., T9 ACC: 78.84% vs. our 79.76%). CEC leverages a graph-based adaptation of classifiers with frozen feature representations to mitigate forgetting, yet its classifier-update process is optimized for incremental extension rather than immediate adaptation, resulting in lagged performance under small-sample cross-category conditions (e.g., T4 ACC: 82.78% vs. our 90.56%). In contrast, GMSIFN's meta-learning initialization discards old-class constraints entirely, allowing parameter space to be fully optimized for novel-class recognition—yielding consistently higher metrics across the twelve tasks.

As shown in Fig. 2, conventional convolution operations applied to multi-sensor vibration signals rely on point-wise kernels and thus fail to capture the structural dependencies among measurement points of the RTA. In contrast, our proposed graph convolution leverages an unsupervised similarity learning mechanism in the GGL framework to adaptively assign edge weights, enabling effective aggregation of features from mechanically coupled sensors. For Task T6, we select three representative fault categories-Normal (4 kN) (normal-to-reverse), Normal (4 kN) (reverse-to-normal), and "loss of indication after near action" (reverse-to-normal)-as illustrated in Fig. 9. For each fault type, the adjacency matrix is computed as the sample-wise mean over all training data to suppress noise from individual samples and reveal stable connection patterns. The nine channels correspond to three-dimensional vibration components *(x, y, z)* from three physical locations: Channels 1-3 from the switch circuit controller (contact seat sensor), Channels 4-6 from the throw rod, and Channels 7–9 from the reducer housing. In the Normal (4 kN) (normal-to-reverse) condition, Channel 1 (switch controller, x-axis) exhibits the highest edge weights with Channel 3 (switch controller, z-axis, weight 0.79) and Channel 8 (reducer housing, y-axis, weight 0.76), indicating strong coupling between axial vibrations in the switch controller and the torque response in the reducer housing during mid-process operation. In the Normal (4 kN) (reverse–to–normal) condition, Channel 4 (throw rod, x-axis) has uniformly high edge weights (0.99) with Channels 1-3 (all from the switch controller), reflecting synchronous displacement of the rod and impact at the contact seat during reverse switching. In the "loss of indication after near action" fault, Channel 7 (reducer housing, x-axis) shows strong connections with Channel 4 (throw rod, x-axis, weight 0.79), Channel 8 (reducer housing, y-axis, weight 0.85), and Channel 9 (reducer housing, z-axis, weight 0.79), consistent with the mechanism in which abnormal torque transmission inside the reducer affects the throw rod's final stroke, thereby disrupting indication switching at the switch controller. To quantitatively validate whether these high-weight connections indeed capture meaningful neighbor relationships associated with faults, we design a fixed-channel neighbor selection experiment: In each run, we fix one channel as the primary channel and select k neighbor channels from the remaining eight using three strategies: Top—select the two neighbors with the highest mean edge weights relative to the primary channel; Bottom—select the two neighbors with the lowest mean edge weights; Random—randomly select two neighbors from the remaining channels. Based on prior ablation results in Task T6, $k = 2$ is set as the optimal neighbor number. The primary channel and its two selected neighbors are concatenated and fed into a simple CNN classifier without any graph structure (two convolutional layers, ReLU activation, max pooling, and a fully connected output layer). This process is repeated for all nine channels (i.e., nine independent experiments) to assess the effect of neighbor selection on classification performance directly. The validation logic of this scheme is as follows: if the graph weights truly reflect meaningful

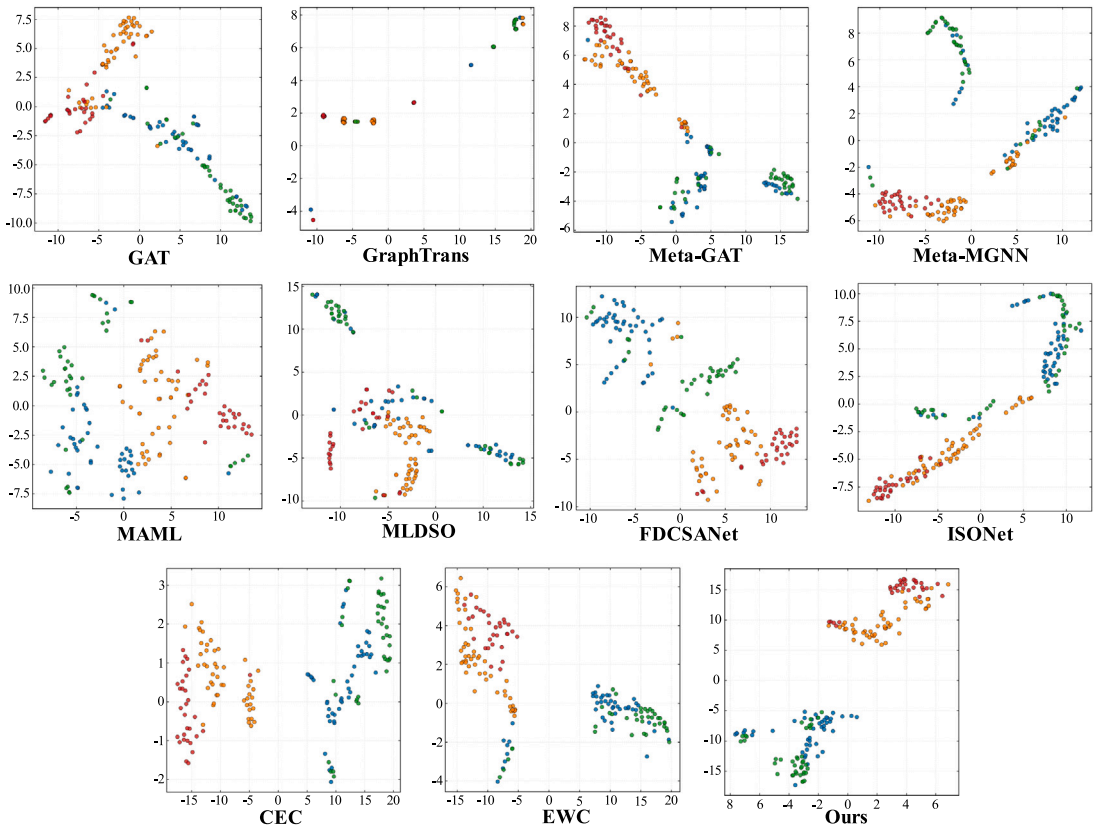

Fig. 10. The feature visualization results on the railway test line ZDJ9-RTA dataset for T9 are presented, where blue represents the Underdriving (3 kN) (Reverse-to-Normal) fault type, orange denotes the Fault-Free (4 kN) (Normal-to-Reverse) fault type, green indicates the Overdriving (5 kN) (Reverse-to-Normal) fault type, and red signifies the Overdriving (6 kN) (Normal-to-Reverse) fault type.

neighborhood relationships, then these neighbors should contain sufficient fault information, and using them for diagnosis should yield a significant improvement in accuracy. The experimental results support this hypothesis: for example, in Normal (4 kN) (reverse-to-normal), fixing Channel 6 (throw rod, z-axis) yields an accuracy of 95.76%±4.24% with the Top strategy, 78.83%±17.92% with Random, and 67.44%±9.88% with Bottom; in the "loss of indication after near action" case, fixing Channel 4 (throw rod, x-axis) results in 65.01% ± 14.92 with Top, 62.19% ± 12.09 with Random, and only 25.00% ± 25.00% with Bottom. These results demonstrate that high-weight neighbors possess stronger discriminative power in fault diagnosis.

Qualitative evidence further confirms GMSIFN's superiority, as Fig. 10's t-SNE visualization for the challenging T9 task shows learned features forming maximally separated clusters where Underdriving (3 kN) (Reverse-to-Normal) and Fault-Free (4 kN) (Normal-to-Reverse) occupy distinct peripheral regions, while Overdriving (5 kN) (Reverse-to-Normal) and Overdriving (6 kN) (Normal-to-Reverse) cluster centrally with minimal overlap (<5%). Competitors exhibit entangled distributions–Meta-MGNN conflates 3 kN and 6 kN faults, while MAML merges fault-free and 5 kN samples–underscoring our method's discriminative power from synergistic adaptive topology modeling and meta-learned generalizability.

Beyond feature discriminability, the adaptation dynamics derived from the convergence curves in Fig. 11 and detailed in the supplementary data reveal that our GMSIFN achieves near-optimal cross-category accuracy within the first fine-tuning epoch in the majority of tasks. For example, in T1, GMSIFN attains 96.00% ACC at epoch 0, converging to stability thereafter, whereas GAT only reaches 16.67% ACC by epoch 3 and requires 5 epochs to exceed 80%. GraphTrans achieves 66.67% at epoch 6 but drops at later epochs due to unstable training. Meta-GAT and Meta-MGNN require at least 3–4 epochs for comparable performance, while MAML remains below 1% ACC during early epochs due to a lack of spatial modeling. When compared with recent RTA-oriented and industrial small-sample methods, the superiority is even more pronounced: FDCSANet in T4 remains at 0% ACC until epoch 4 and peaks at only 66.67%, while ISONet achieves 55.56% by epoch 3 and 88.89% by epoch 8—both substantially trailing our 90.56% single-epoch result. Across tasks T1–T12, FDCSANet and ISONet typically require 6–9 epochs to approach their maxima, reflecting the absence of dynamic cross-category optimization. Finally, relative to continual learning frameworks such as EWC and CEC, our method avoids the trade-off between retaining past knowledge and adapting to new categories. For instance, in T6, EWC needs 7 epochs to reach 96.67% ACC, whereas GMSIFN obtains 89.47% at epoch 0 and stabilizes above 99% immediately after. In T11, EWC takes 9 epochs to reach 96%, but our approach delivers 98.96% at epoch 0 and sustains >98.9% thereafter. Note that CEC does not involve any fine-tuning process for new classes in the small-sample diagnosis stage; instead, it directly normalizes the extracted features of the small-sample data to form class prototypes and supplements them as weight vectors for the new classes. As there is no training process for the new classes, CEC naturally has no convergence curve and is therefore excluded from Fig. 11. This accelerated convergence arises from two complementary innovations: (1) The category-shift meta-learning framework explicitly optimizes for dynamic fault distributions through episodic simulation, producing robust initialization for immediate cross-category transfer; and (2) The graph-guided feature aggregation in GGL constructs informative sensor topologies before adaptation, enabling single-step knowledge extraction even from sparse samples. By contrast, static pre-training paradigms (GAT/GraphTrans), RTA-specific but graph-free networks (FDCSANet/ISONet), and old-class regularized continual learning models (EWC/CEC) all require multiple epochs to approach suboptimal performance.

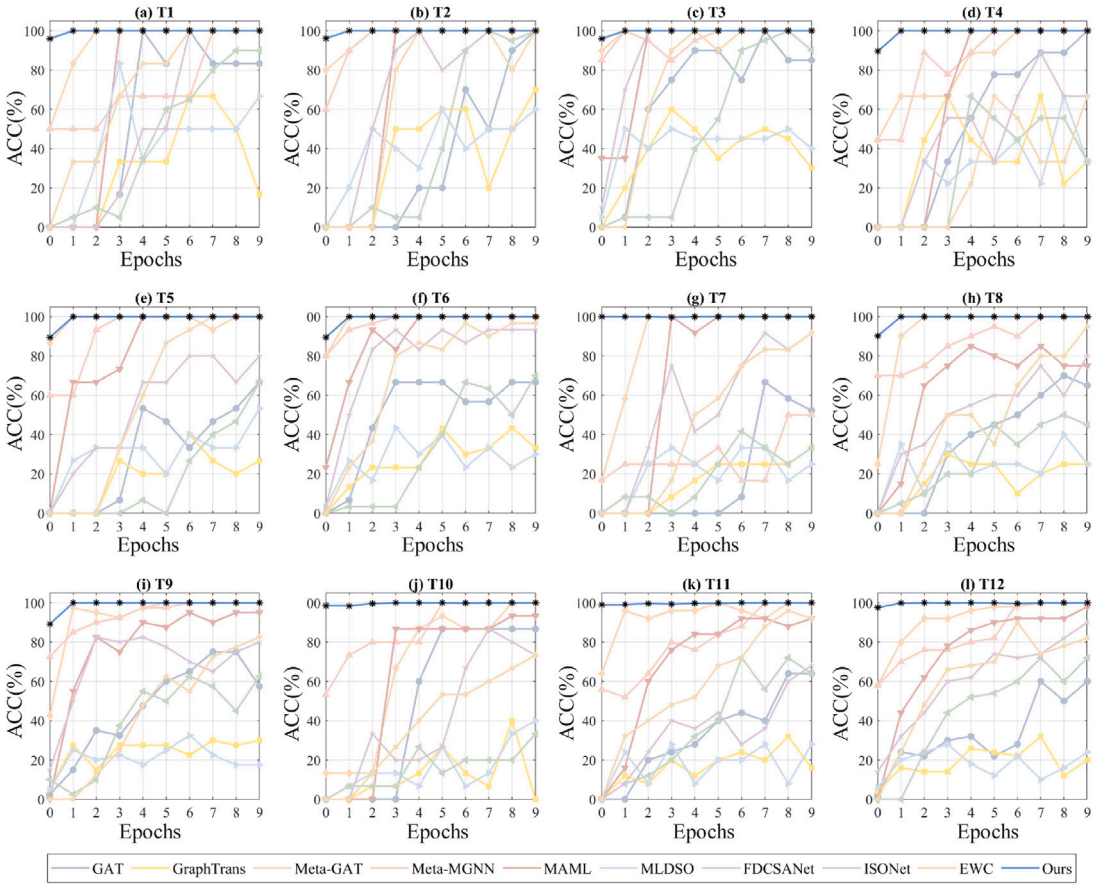

Fig. 11. The accuracy trend on the railway test line ZDJ9-RTA dataset after fine-tuning the model for 10 epochs on a small sample from the testing set, following successful training on the training set.

Collectively, these results establish our framework as the new state-of-the-art for railway turnout fault diagnosis. The unified architecture — combining unsupervised graph construction for spatial integrity, GRU-augmented feature fusion for hierarchical learning, and dynamic meta-learning for cross-category adaptation — addresses critical gaps in conventional approaches, proving vital for evolving railway systems where emergent faults demand rapid diagnosis with small samples.

3.2.3. Ablation experiments

To comprehensively validate the contribution of each component in our framework, six ablation variants are designed:

Ablation_1 (Verification of cross-category meta-training strategy): The GMSIFN model is trained conventionally on all training categories and fine-tuned with small-sample samples during testing, eliminating the meta-learning adaptation mechanism.

Ablation_2 (Effect of linear compression): Removing the linear compression layer in the GGL module, directly constructing graphs from raw 5120-length signals via KNN.

Ablation_3 (Role of GRU in node feature aggregation): Discarding the GRU unit in node feature aggregation, using GAT alone for neighbor feature propagation (which replaces central node features with aggregated neighbors, discarding original signal patterns).

Ablation_4 (global–local feature fusion): Replacing the GRU-based feature fusion with direct classification—aggregating all node features into a global representation vector (via sum-pooling and dimension expansion), then feeding solely this global vector to linear layers, abandoning the integration of fine-grained local node features;

Variant_1 (Graph construction method): Substituting KNN with PathGraph-based [41] topology construction while maintaining identical neighbor counts;

Variant_2 (GRU vs. LSTM): Replacing all GRUs with LSTMs in both feature aggregation and classification modules.

Variant_3 (Meta-learning framework): Replacing the MAML-based meta-optimization process with the Reptile [42] framework to investigate the influence of different gradient-based meta-learning strategies on cross-category adaptation.

Experimental results in Table 5 and Fig. 12 demonstrate the consistent superiority of our full model across all 12 tasks (T1–T12). The overall performance degradation observed in ablation models highlights the synergistic necessity of each component. Specifically, Ablation_1 exhibits the most severe deterioration (e.g., T1 ACC drops from 96.95% to 87.95%), confirming that the cross-category meta-learning strategy is fundamental for generalizing to novel fault categories. This is particularly evident in complex multi-class scenarios, when fault categories increase from 2 to 4 (T10), Ablation_1's ACC plunges to 45.80% versus 55.54% for our full model. Ablation_2 causes moderate performance loss (T4 ACC: 90.56% → 82.39%) due to inefficient modeling of high-dimensional raw signals, validating the critical role of feature compression in distilling discriminative patterns from nine-channel sensor measurements.

The GRU hierarchies prove indispensable for robust representation learning. In Ablation_3, discarding GRU gates causes direct feature replacement by neighbor aggregation (via GAT), whereas our GRU mechanism selectively preserves original node features through its reset/update gates, enabling hierarchical fusion of local and neighbor characteristics. This enriches feature information diversity, explaining the significant F1 declines (T6 F1: 96.04% → 89.22%). Ablation_4 shows unstable AUC performance (T3 AUC decreases to 98.98% versus 99.98% for the full model), confirming that bypassing global–local

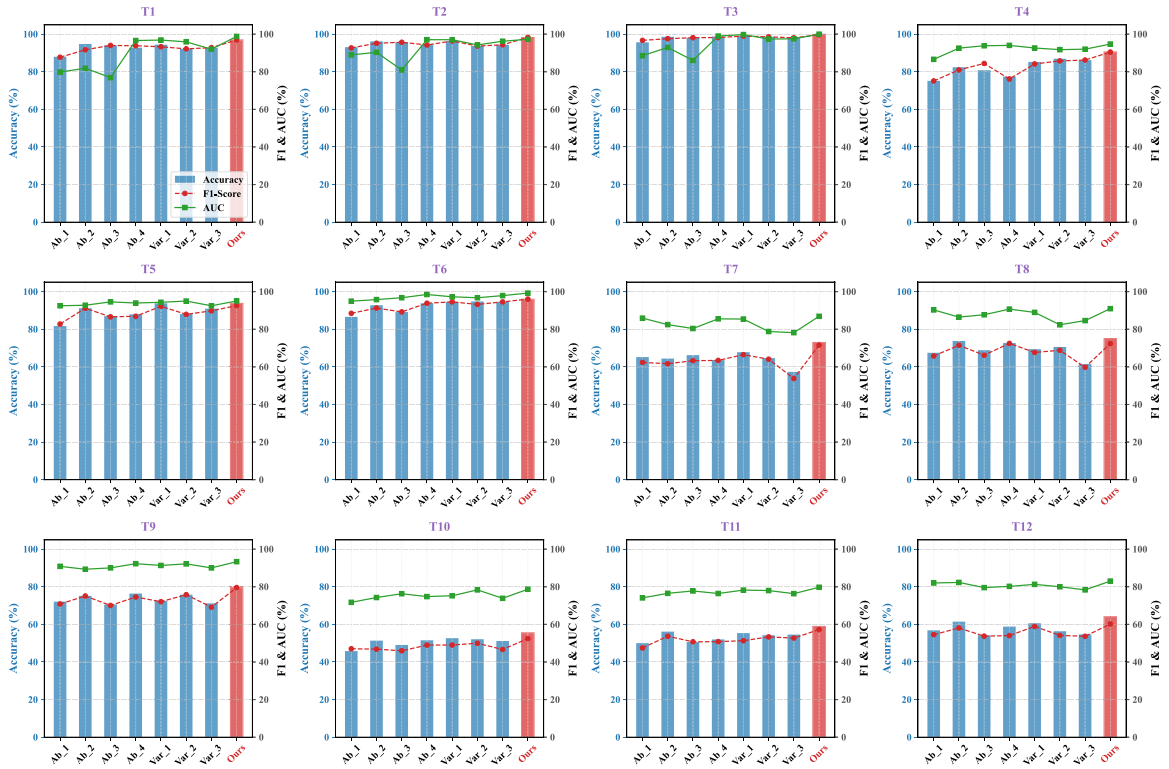

Fig. 12. Comprehensive visualization of ablation experiment outcomes.

feature fusion impairs discrimination of critical vibration patterns—particularly distinguishing force-dependent anomalies (e.g., Underdriving (3 kN) vs. Overdriving (8 kN)) from mechanical failures (e.g., Unlocking failure). Variant_1 demonstrates comparable AUC in simple tasks (T3: 99.76% vs. 99.98%) but suffers in multi-class scenarios (T4 ACC: 85.22% vs. 90.56%), revealing KNN's superiority in capturing sensor correlations. Variant_2 exhibits higher variance (T12 AUC std: 5.35 vs. 2.74), underscoring GRU's advantage in modeling sequential vibration signatures across the reducer housing, throw rod, and switch controller interfaces.

For Variant_3, replacing the MAML-based meta-optimization with the first-order Reptile algorithm led to consistent performance degradation across multiple tasks. For example, in T9 the ACC decreased by 8.59%, F1 dropped by 10.44% . Similar declines were observed in T4 (ACC: 90.56% → 86.16%), T5 (F1: 92.56% → 89.79%), and T6 (AUC: 99.23% → 97.88%). These reductions reflect Reptile's update mechanism — parameter interpolation towards task-specific directions without explicit second-order gradient information — which, fails to capture the local curvature of the task loss landscape that is crucial for our cross-category small-sample diagnosis. In such scenarios, fault distributions differ significantly between training and testing, and multi-sensor vibration sequences possess strong spatio-temporal coupling. MAML's bi-level optimization with explicit inner- and outer-loop backpropagation encodes these cross-sensor and cross-category structural priors within meta-initialization parameters, enabling rapid and stable adaptation under small-sample conditions. Conversely, Reptile's coarse parameter interpolation lacks curvature-sensitive adaptability, increasing feature inseparability among easily confused fault categories and reducing generalization robustness. Under strict small-sample fine-tuning and full-sample evaluation protocols, its overall diagnostic performance falls notably short of MAML, further justifying MAML as the baseline meta-learning framework in our study.

To further elucidate model behaviors, Fig. 13 visualizes the four-class T9 task outcomes (confusion matrices and ROC curves). The confusion matrices reveal that Ablation_3 misclassifies 27% of Fault-free (4 kN) samples as Overdriving (5 kN), while Ablation_4 shows confusion between Underdriving (3 kN) and Fault-free conditions. Our full model maintains 93% diagnostic consistency. The Receiver Operating Characteristic (ROC) curves demonstrate our method's precision: achieving 0.93 AUC with 85% true positive rate at 10% false alarm level. In the confusion matrix of Variant_3, there are clear misclassifications between Underdriving (3 kN) and Overdriving (5 kN), as well as between Overdriving (5 kN) and Overdriving (6 kN), leading to significantly lower discriminability and reduced AUC for these categories compared to our method. This demonstrates that MAML's curvature-aware bi-level optimization better captures inter-class discriminative features in cross-category, small-sample diagnosis settings. These results verify that our framework enables cross-category small-sample diagnosis of railway actuators by synergizing: (1) meta-learning for novel fault adaptation, (2) vibration feature distillation through compressed graph learning, and (3) hierarchical fusion of local sensor signals and global mechanical states.

3.3. Case II: Field deployment ZDJ9-RTA dataset

To further validate the performance of GMSIFN and other state-of-the-art methods, this case employs a more challenging field deployment vibration dataset from ZDJ9 railway turnout actuators (provided by China Railway Signal & Communication Corp (Xi'an)). As shown in Fig. 14, this dataset was collected from in-service ZDJ9-RTAs during daily railway operations. Three tri-axial vibration sensors were mounted at the reducer housing, the throw rod side, and the switch controller side, identical to the configuration in Case I. Signals were synchronously acquired at a sampling rate of 512 Hz over a 10-second cycle, resulting in 5120 temporal samples per channel. As shown in Fig. 15, compared with the railway test line dataset in Case I, the field deployment dataset inherently includes complex real-world interferences, such as mechanical vibrations from adjacent train passages, transient impacts from routine maintenance, short-duration high-frequency disturbances,

Table 5
Ablation study results on RTA dataset (%).

Task No.	Metric	Ablation_1	Ablation_2	Ablation_3	Ablation_4	Variant_1	Variant_2	Variant_3	Ours
T1	ACC	87.95 ± 6.03	94.75 ± 2.62	93.98 ± 3.62	92.77 ± 2.41	94.38 ± 2.56	92.17 ± 3.01	92.77 ± 2.41	**96.95 ± 0.12**
	F1	87.71 ± 6.23	91.67 ± 5.71	93.99 ± 3.61	93.81 ± 3.65	93.27 ± 1.98	92.18 ± 3.02	92.78 ± 2.39	**96.93 ± 0.14**
	AUC	79.84 ± 4.94	81.84 ± 7.08	76.88 ± 5.08	96.59 ± 0.23	96.85 ± 1.98	95.89 ± 4.12	91.98 ± 3.11	**98.79 ± 0.29**
T2	ACC	93.05 ± 1.89	96.14 ± 1.27	95.65 ± 1.45	94.31 ± 3.17	96.37 ± 0.64	93.67 ± 6.33	94.31 ± 3.17	**98.28 ± 0.26**
	F1	92.65 ± 2.24	95.16 ± 2.07	95.67 ± 1.44	94.22 ± 3.25	96.37 ± 0.63	93.67 ± 6.33	94.32 ± 3.16	**98.26 ± 0.28**
	AUC	88.93 ± 1.85	90.34 ± 2.86	80.94 ± 14.71	97.04 ± 1.17	97.03 ± 1.74	94.26 ± 3.22	96.22 ± 2.49	**97.26 ± 0.38**
T3	ACC	95.65 ± 4.35	98.58 ± 1.02	98.10 ± 1.90	98.72 ± 1.28	99.07 ± 0.23	98.55 ± 1.45	98.48 ± 1.52	**99.95 ± 0.02**
	F1	96.69 ± 3.31	97.55 ± 1.57	98.11 ± 1.90	98.18 ± 1.83	98.96 ± 0.84	98.55 ± 1.45	98.03 ± 1.48	**99.63 ± 0.17**
	AUC	88.55 ± 4.15	92.91 ± 4.53	86.03 ± 8.42	98.98 ± 1.03	99.76 ± 0.12	97.40 ± 2.61	97.50 ± 2.12	**99.98 ± 0.01**
T4	ACC	75.16 ± 6.61	82.39 ± 6.92	80.82 ± 9.12	77.36 ± 3.78	85.22 ± 7.86	86.93 ± 8.50	86.16 ± 5.66	**90.56 ± 1.89**
	F1	75.11 ± 9.49	80.95 ± 8.46	84.40 ± 5.67	76.14 ± 9.05	84.14 ± 8.74	85.75 ± 9.76	86.21 ± 5.67	**90.46 ± 1.78**
	AUC	86.59 ± 3.71	92.54 ± 2.57	93.82 ± 5.66	94.01 ± 3.72	92.62 ± 4.21	91.75 ± 3.11	92.00 ± 4.91	**94.71 ± 1.22**
T5	ACC	81.70 ± 11.11	91.18 ± 5.56	86.96 ± 8.70	87.91 ± 6.21	93.79 ± 2.29	88.08 ± 5.03	90.95 ± 3.83	**93.80 ± 3.59**
	F1	82.76 ± 11.58	91.19 ± 5.55	86.58 ± 9.04	86.91 ± 7.22	92.21 ± 2.50	87.92 ± 5.21	89.79 ± 5.03	**92.56 ± 2.87**
	AUC	92.49 ± 2.23	92.76 ± 5.87	94.56 ± 2.46	93.95 ± 1.51	94.34 ± 1.19	94.98 ± 2.79	92.49 ± 3.43	**95.03 ± 1.56**
T6	ACC	86.59 ± 9.79	92.76 ± 2.89	89.13 ± 5.07	93.84 ± 1.09	94.57 ± 2.54	94.77 ± 3.27	94.57 ± 1.09	**96.01 ± 0.37**
	F1	88.46 ± 7.89	91.27 ± 1.47	89.22 ± 4.93	93.84 ± 1.10	94.56 ± 2.54	93.21 ± 4.06	94.54 ± 1.12	**96.04 ± 0.39**
	AUC	94.93 ± 4.03	95.77 ± 3.06	96.77 ± 2.22	98.48 ± 0.23	97.26 ± 2.43	96.75 ± 2.48	97.88 ± 1.15	**99.23 ± 0.14**
T7	ACC	65.19 ± 11.26	64.40 ± 5.76	66.23 ± 9.69	63.61 ± 7.59	67.80 ± 8.64	64.66 ± 10.73	57.33 ± 6.02	**73.03 ± 6.55**
	F1	62.38 ± 14.61	61.67 ± 8.33	63.27 ± 12.09	63.44 ± 6.66	66.47 ± 10.12	64.07 ± 11.09	53.80 ± 8.58	**71.58 ± 6.14**
	AUC	85.90 ± 5.15	82.46 ± 5.33	80.39 ± 10.40	85.54 ± 4.63	85.43 ± 6.02	78.80 ± 5.88	78.19 ± 8.21	**86.88 ± 4.58**
T8	ACC	67.49 ± 15.03	73.77 ± 8.74	68.85 ± 9.29	72.68 ± 4.37	69.40 ± 10.38	70.58 ± 4.32	61.48 ± 2.46	**75.04 ± 6.19**
	F1	65.77 ± 17.64	71.48 ± 11.06	66.20 ± 11.68	72.48 ± 4.48	67.67 ± 12.18	68.82 ± 5.27	59.76 ± 2.92	**72.36 ± 6.90**
	AUC	90.28 ± 2.69	86.46 ± 4.41	87.71 ± 4.10	90.70 ± 2.04	88.97 ± 4.53	82.38 ± 3.26	84.60 ± 4.79	**90.95 ± 1.37**
T9	ACC	72.09 ± 9.51	75.15 ± 5.92	70.55 ± 3.68	76.38 ± 8.28	72.51 ± 11.05	75.77 ± 6.44	71.17 ± 8.59	**79.76 ± 3.69**
	F1	70.84 ± 11.10	75.08 ± 6.11	70.06 ± 3.80	74.54 ± 10.14	72.06 ± 11.92	75.78 ± 6.40	69.15 ± 10.30	**79.59 ± 3.57**
	AUC	90.85 ± 4.36	89.32 ± 2.36	90.04 ± 4.45	92.27 ± 1.78	91.33 ± 2.57	92.17 ± 3.80	90.03 ± 4.16	**93.34 ± 1.08**
T10	ACC	45.80 ± 8.78	51.34 ± 4.39	49.05 ± 8.59	51.53 ± 4.97	52.67 ± 3.82	52.10 ± 6.30	51.15 ± 5.35	**55.54 ± 5.92**
	F1	47.01 ± 7.87	46.83 ± 7.32	45.94 ± 9.92	48.98 ± 6.03	49.02 ± 4.99	49.96 ± 4.98	46.65 ± 7.38	**52.44 ± 6.18**
	AUC	71.72 ± 7.56	74.28 ± 2.82	76.29 ± 3.04	74.77 ± 4.13	75.21 ± 5.21	78.34 ± 8.36	73.86 ± 5.30	**78.68 ± 3.01**
T11	ACC	50.02 ± 5.95	56.15 ± 2.58	50.60 ± 5.36	51.99 ± 3.18	55.36 ± 1.39	54.19 ± 5.29	54.56 ± 7.34	**58.96 ± 4.17**
	F1	47.47 ± 5.78	53.71 ± 2.84	50.74 ± 4.90	50.81 ± 3.52	51.30 ± 1.95	53.33 ± 5.64	52.72 ± 6.80	**57.24 ± 5.59**
	AUC	74.08 ± 5.46	76.51 ± 3.16	77.77 ± 3.74	76.42 ± 3.14	78.22 ± 2.01	77.99 ± 7.66	76.34 ± 3.65	**79.73 ± 3.61**
T12	ACC	56.83 ± 3.97	61.46 ± 4.63	54.19 ± 9.70	58.81 ± 2.42	60.58 ± 2.87	56.35 ± 9.52	54.63 ± 4.41	**64.08 ± 4.81**
	F1	54.53 ± 2.43	58.07 ± 6.71	53.73 ± 8.82	54.03 ± 3.22	58.92 ± 3.50	54.13 ± 12.45	53.67 ± 3.15	**60.22 ± 2.96**
	AUC	82.01 ± 1.75	82.30 ± 4.42	79.55 ± 6.79	80.21 ± 3.31	81.28 ± 4.77	80.01 ± 5.35	78.34 ± 2.31	**83.01 ± 2.74**

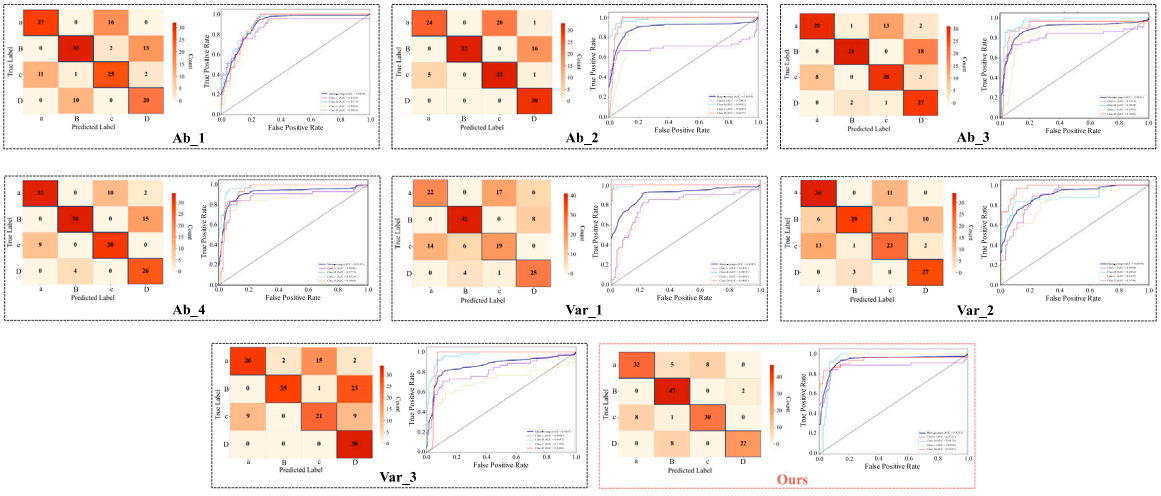

Fig. 13. Visualization of confusion matrices and ROC curves for ablated methodologies in T9.

Fig. 14. Field-collected dataset of ZDJ9-RTA from in-service railway environments.

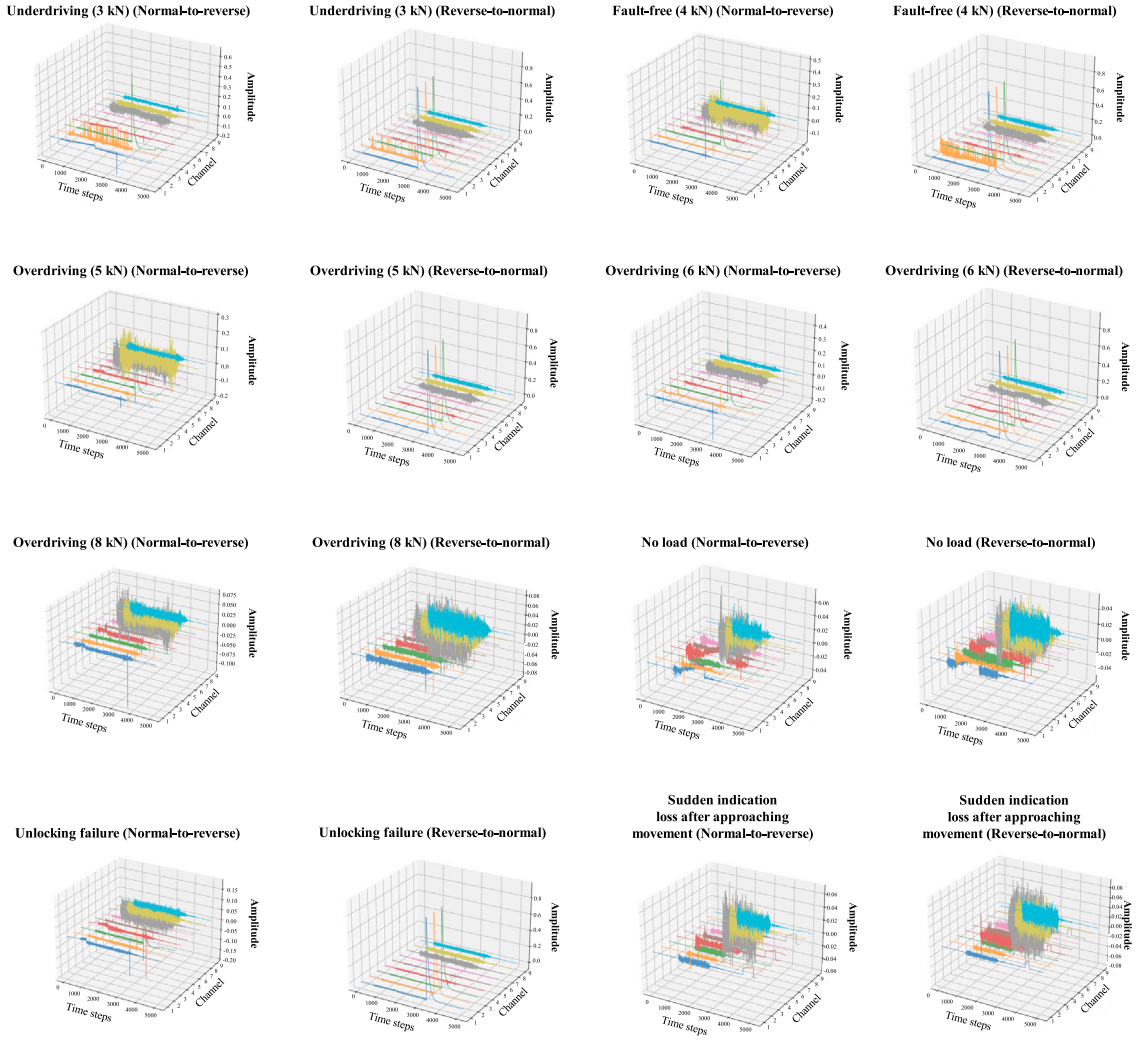

Fig. 15. Visualization of nine-channel 16 types of field deployment ZDJ9-RTA fault signals.

and excitation transmitted through steel rails by passing lead trains. These factors lead to nonstationary and noise-rich vibration patterns, making the diagnosis task considerably more challenging. The dataset maintains the same 16 fault categories as in Case I, but the number of samples per category differs, as detailed in Table 6.

3.3.1. Comparative studies

In the field-deployed ZDJ9-RTA dataset, the presence of nonstationary disturbances, transient impact noise, and mechanical vibrations from adjacent equipment significantly increased diagnostic complexity. As shown in Fig. 16, compared with the test line dataset (Case I), almost all methods exhibited decreases in ACC, F1, and AUC across the twelve cross-category small-sample tasks, with standard deviations often enlarging, indicating reduced prediction stability. For example, in Task T4, the ACC of GAT dropped from 72.65% (17.93) in Case I to 65.41% (10.69) in Case II, and its F1 decreased from 71.72% (18.95) to 62.25% (13.95) with persistent high variance, suggesting instability under noisy conditions. GraphTrans showed more pronounced degradation, e.g., its ACC in T6 declined from 42.25% (3.26) to

Table 6
Detailed description of the field deployment ZDJ9-RTA dataset.

RTA state	Normal-to-reverse	Number	Reverse-to-normal	Number
Underdriving (3 kN)	A	60	a	60
Fault-free (4 kN)	B	63	b	63
Overdriving (5 kN)	C	46	c	46
Overdriving (6 kN)	D	35	d	35
Overdriving (8 kN)	E	39	e	39
No load	F	48	f	48
Unlocking failure	G	69	g	69
Sudden indication loss after approaching movement	H	42	h	42

32.68% (3.27), reflecting reduced feature extraction efficiency in complex noise. Meta-GAT maintained moderate robustness in some tasks, such as T6 (ACC from 73.85% (6.54) to 72.05% (7.38)), but in T1 its ACC decreased from 77.82% (14.88) to 61.83% (8.37) with F1 dropping by over 15%. MAML remained stable in tasks like T3 (ACC from 93.67 to 91.55%), yet in T8 its ACC sharply fell from 62.30% (3.90) to 56.39% (6.83). MLDSO already showed relatively low performance in Case I, and its effectiveness further deteriorated under noisy interference. In contrast, GMSIFN's performance drops were minimal, such as in T11, where its ACC decreased from 58.93% (4.17) to 57.69% (3.48), maintaining clear superiority. Advanced RTA-related and industrial small-sample diagnostic methods showed similar downward trends. FDCSANet's ACC in T9 moved from 64.51% (6.84) in Case I to 64.39% (13.66), with little change in mean but larger standard deviation in Case II, indicating greater instability. ISONet's performance drops were smaller overall, yet in T7 its ACC fell from 54.71% (4.45) to 42.71% (11.98), with declines in both accuracy and recall. By comparison, GMSIFN's ACC in the same task decreased modestly from 73.03% (6.55) to 65.89% (9.64), a considerably smaller drop. Continual learning baselines were also impacted: EWC's ACC in T4 declined from 78.49% (8.57) to 60.38% (6.92), with F1 dropping by almost 21%; CEC retained high accuracy in T3 but suffered markedly in T11 (ACC from 55.47% to 45.85%), highlighting limited adaptability in multi-class noisy settings.

As shown in Fig. 17, the t-SNE visualizations corroborated these numerical findings. In Case II, most baselines displayed substantially reduced inter-class separation, particularly GAT and GraphTrans, which exhibited obvious overlap between force-related anomalies and mechanical faults. In contrast, GMSIFN preserved clear class boundaries with ACC and AUC drops typically within 3%–5%, demonstrating strong discrimination under noisy conditions.

In Case II, noisy interference significantly delayed the convergence of baseline methods. As shown in Fig. 18, most approaches required 4–8 iterations to approach peak accuracy; for example, GAT stabilized around the 6th-7th iteration in tasks T1 and T6, while GraphTrans often exceeded 8 iterations. EWC exhibited pronounced delays in high-noise tasks, and MLDSO remained below 50% accuracy in several tasks even in late iterations. In contrast, the proposed GMSIFN achieved over 99% accuracy within only 1–2 iterations in the majority of tasks (e.g., T1 and T6) and maintained optimal performance thereafter, demonstrating rapid adaptation and robustness in noisy environments.

4. Conclusion

This study establishes a novel paradigm for cross-category small-sample fault diagnosis in ZDJ9-RTA, overcoming critical limitations of conventional methods that assume homogeneous fault categories and rely on single-sensor data. To achieve this cross-category diagnostic capability while preserving multi-sensor spatial integrity, we propose a dual-innovation framework integrating dynamic category-shift meta-learning and the GMSIFN. The meta-learning strategy simulates real-world emerging faults through stochastic task construction, enabling rapid adaptation to novel fault types with small samples. Concurrently, GMSIFN dynamically infers sensor topology via signal similarity and fuses multi-sensor spatial dependencies through GRU-gated attention, explicitly modeling fault propagation paths across mechanical interfaces. Comprehensive validation on two scenarios of ZDJ9-RTA, namely the railway test line dataset and the field deployment dataset, across 12 cross-category tasks demonstrates the consistent effectiveness of the proposed framework, bridging the gap between data scarcity and evolving fault patterns, and extending small-sample diagnosis to dynamically changing industrial environments.

Despite these advancements, two limitations warrant future research: First, the framework assumes novel faults emerge as discrete categories, whereas real-world scenarios may involve continuous fault evolution. Second, GMSIFN's graph construction relies solely on data-driven sensor correlation, potentially overlooking the physical constraints of sensor networks, and has not yet explored the correspondence between the learned graph topology/edge weights and the actual physical fault propagation mechanisms within the RTA (e.g., mechanical linkages and force transmission paths), which may limit the interpretability of the graph network. Future work will address these limitations in sequence: First, incremental meta-learning will be incorporated to continuously assimilate new fault patterns without retraining, enhancing the framework's ability to handle continuously evolving faults in dynamic industrial scenarios. Second, physics-informed graph construction will be explored, for example, by integrating the kinematic chain model of the turnout actuator with hypergraph representations, to merge hierarchical mechanical structure information with multi-axis vibration signal topology, thereby achieving physically consistent topology modeling and improving the interpretability of the RTA fault propagation mechanisms for domain experts. Such enhancements could strengthen the framework's applicability to lifelong learning systems in safety-critical railway infrastructure.

CRediT authorship contribution statement

Yuhan Huang: Writing – original draft, Visualization, Software, Methodology, Conceptualization. **Xiaoxi Hu:** Writing – review & editing, Resources, Investigation, Data curation, Conceptualization. **Fei Chen:** Supervision, Software, Investigation. **Jingming Cao:** Writing – review & editing, Visualization. **Yiming He:** Writing – review & editing, Visualization, Supervision. **Qi Ming:** Writing – review & editing, Validation. **Huan Wang:** Writing – review & editing, Validation. **Jianing Wang:** Writing – review & editing. **Tao Tang:** Writing – review & editing, Supervision.

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 work was supported by the Guangxi Science and Technology Major Program, China (Guike AB22035008). The authors would also like to thank the Xiangyang Signaling and Communication Section of the Wuhan Railway Bureau for providing the RTA dataset.

Data availability

Data will be made available on request.

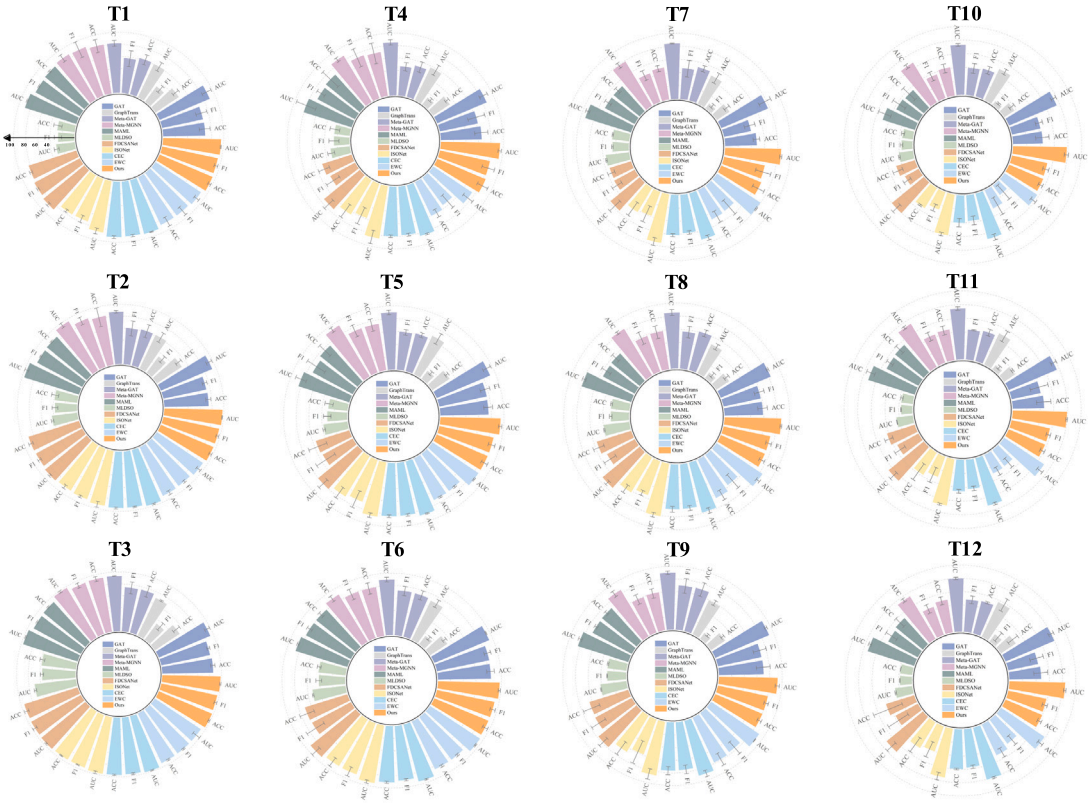

Fig. 16. Visualization of small-sample diagnostic results on field deployment ZDJ9-RTA dataset.

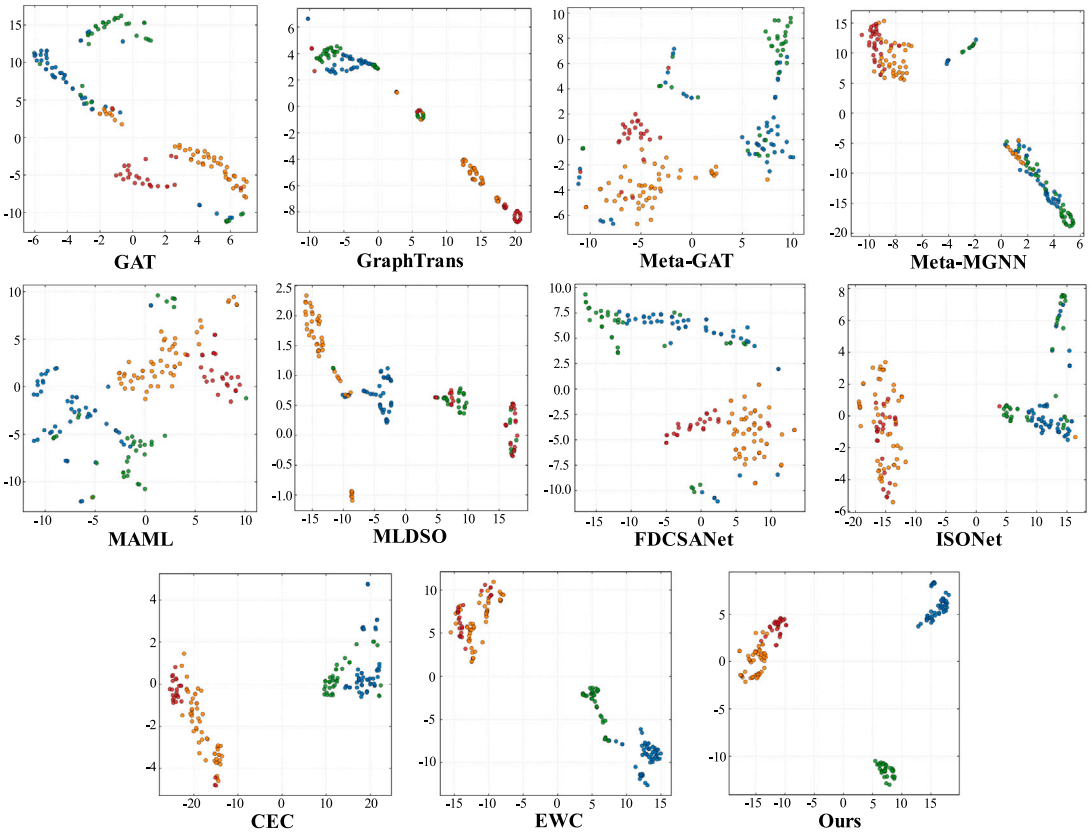

Fig. 17. The feature visualization results on field deployment ZDJ9-RTA dataset for T9 are presented, where blue represents the Underdriving (3 kN) (Reverse-to-Normal) fault type, orange denotes the Fault-Free (4 kN) (Normal-to-Reverse) fault type, green indicates the Overdriving (5 kN) (Reverse-to-Normal) fault type, and red signifies the Overdriving (6 kN) (Normal-to-Reverse) fault type.

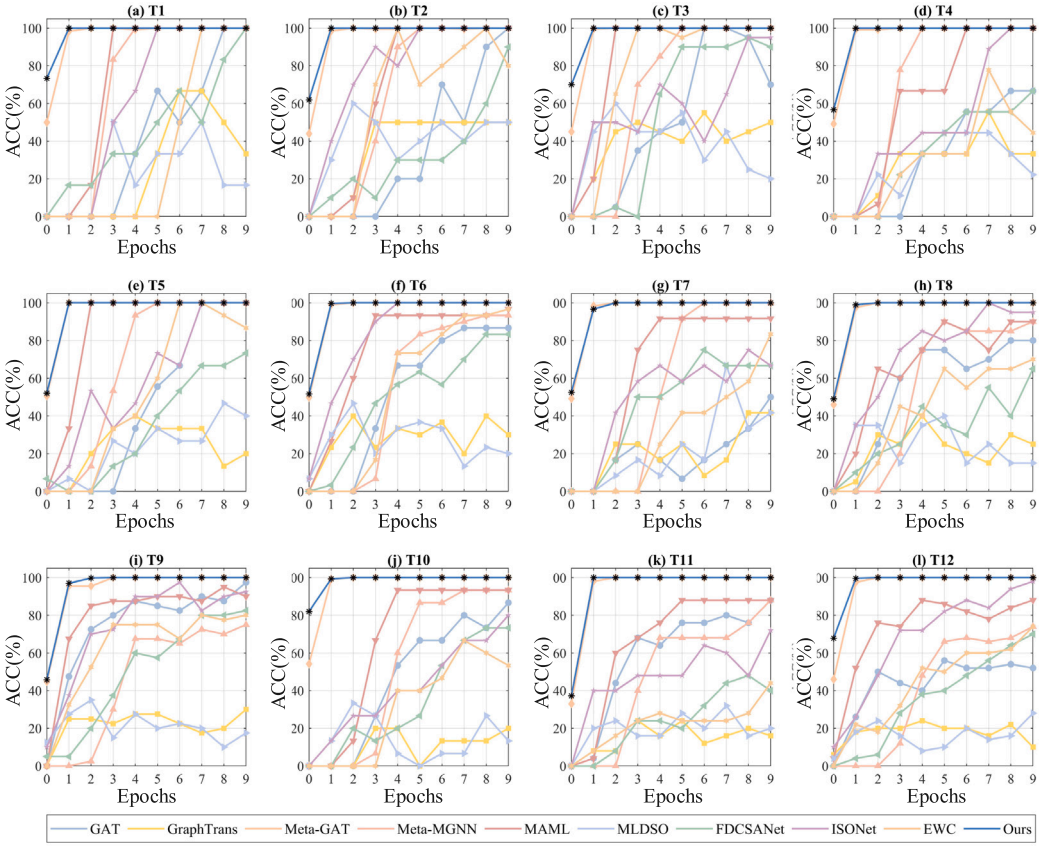

Fig. 18. The accuracy trend on the field deployment ZDJ9-RTA dataset after fine-tuning the model for 10 epochs on a small sample from the testing set, following successful training on the training set.

References

[1] J.-L. Zhou, Z.-M. Guo, A hybrid SNN-STLSTM method for human error assessment in the high-speed railway system, Adv. Eng. Informatics 60 (2024) 102408, http://dx.doi.org/10.1016/j.aei.2024.102408.

[2] C. He, H. Shi, R. Li, J. Li, Z. Yu, Interpretable modulated differentiable STFT and physics-informed balanced spectrum metric for freight train wheelset bearing cross-machine transfer fault diagnosis under speed fluctuations, Adv. Eng. Informatics 62 (2024) 102568, http://dx.doi.org/10.1016/j.aei.2024.102568.

[3] S. Li, J. Xuan, Q. Zhang, Z. Wang, L. Tang, T. Shi, Open set transfer learning for bearing defect recognition based on selective momentum contrast and dual adversarial structure, Adv. Eng. Informatics 62 (2024) 102641, http://dx.doi.org/10.1016/j.aei.2024.102641.

[4] X. Hu, Y. Cao, Y. Sun, T. Tang, Railway automatic switch stationary contacts wear detection under few-shot occasions, IEEE Trans. Intell. Transp. Syst. 23 (9) (2022) 14893–14907, http://dx.doi.org/10.1109/TITS.2021.3135006.

[5] X. Hu, Y. Cao, T. Tang, Y. Sun, Data-driven technology of fault diagnosis in railway point machines: review and challenges, Transp. Saf. Environ. 4 (4) (2022) tdac036, http://dx.doi.org/10.1093/tse/tdac036.

[6] Y. He, D. He, Z. Lao, Z. Jin, J. Miao, Z. Lai, Y. Chen, Few-shot fault diagnosis of turnout switch machine based on flexible semi-supervised meta-learning network, Knowl.-Based Syst. 294 (2024) 111746.

[7] L. Wang, G. Mi, Application of HHT and GCA-NBC in fault diagnosis of S700k switch machine, Transducer Microsyst. Technol. 40 (2021) 151–155.

[8] Z. Jin, D. He, Z. Wei, Intelligent fault diagnosis of train axle box bearing based on parameter optimization VMD and improved DBN, Eng. Appl. Artif. Intell. 110 (2022) 104713.

[9] X. Yang, A. Wang, B. An, Research on turnout fault diagnosis algorithm based on SVM, Comput. Meas. Control. 25 (07) (2017) 55–58.

[10] C. Bian, S. Yang, T. Huang, Q. Xu, J. Liu, E. Zio, Degradation state mining and identification for railway point machines, Reliab. Eng. Syst. Saf. 188 (2019) 432–443.

[11] J. Liu, T. Wen, G. Xie, Y. Cao, C. Roberts, Multi-time-scale variational mode decomposition-based robust fault diagnosis of railway point machines under multiple noises, Chin. J. Electron. 33 (3) (2024) 814–822.

[12] S. Xiao, Q. Feng, X. Hu, Y. Song, G. Cong, Z. Yao, H. Li, A novel fault diagnosis model based on deep feature fusion network under imbalanced data: towards railway dual-switch machines traction occasion, Meas. Sci. Technol. 36 (1) (2024) 016219.

[13] Y. He, D. He, Z. Lao, Z. Yao, H. Sun, C. He, Z. Yuan, A class-center fine-tuning prototypical network for few-shot fault diagnosis of turnout switch machine driven by multi-source signals, Measurement 242 (2025) 115920.

[14] X. Chen, H. Liu, Z. Duan, Railway switch fault diagnosis based on multi-heads channel self attention, residual connection and deep CNN, Transp. Saf. Environ. 5 (1) (2023) tdac045.

[15] X. Chen, X. Hu, T. Wen, Y. Cao, Vibration signal-based fault diagnosis of railway point machines via double-scale CNN, Chin. J. Electron. 32 (5) (2023) 972–981.

[16] X. Hu, X. Zhang, F. Chen, Z. Liu, J. Liu, L. Tan, T. Tang, Simultaneous fault diagnosis for sensor and railway point machine for autonomous rail system, in: 2024 IEEE 27th International Conference on Intelligent Transportation Systems, ITSC, IEEE, 2024, pp. 1011–1016.

[17] H. Shao, Y. Lai, H. Liu, J. Wang, B. Liu, LSFConvformer: A lightweight method for mechanical fault diagnosis under small samples and variable speeds with time-frequency fusion, Mech. Syst. Signal Process. 236 (2025) 113016.

[18] Z. Lao, D. He, Z. Jin, C. Liu, H. Shang, Y. He, Few-shot fault diagnosis of turnout switch machine based on semi-supervised weighted prototypical network, Knowl.-Based Syst. 274 (2023) 110634.

[19] X. Hu, T. Tang, L. Tan, H. Zhang, Fault detection for point machines: A review, challenges, and perspectives, Actuators 12 (10) (2023) http://dx.doi.org/10.3390/act12100391.

[20] Z. Lao, D. He, H. Sun, Y. He, Z. Lai, S. Shan, Y. Chen, Few-shot fault diagnosis of switch machine based on data fusion and balanced regularized prototypical network, Eng. Appl. Artif. Intell. 135 (2024) 108847.

[21] H. Zhong, D. He, Z. Lao, Z. Jin, G. Shen, Y. Chen, Improved metric-based meta learning with attention mechanism for few-shot cross-domain train bearing fault diagnosis, Meas. Sci. Technol. 35 (7) (2024) 075101.

[22] H. Ren, J. Wang, C. Shen, W. Huang, Z. Zhu, Dual classifier-discriminator adversarial networks for open set fault diagnosis of train bearings, IEEE Sensors J. 23 (18) (2023) 22040–22050.

[23] J. Wang, H. Shao, S. Yan, B. Liu, C-ECAFormer: A new lightweight fault diagnosis framework towards heavy noise and small samples, Eng. Appl. Artif. Intell. 126 (2023) 107031.

[24] C. Chen, Z. Xu, M. Mei, K. Huang, S.M. Lo, Fault diagnosis scheme for railway switch machine using multi-sensor fusion tensor machine., Comput. Mater. Contin. 79 (3) (2024).

[25] C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in: International Conference on Machine Learning, PMLR, 2017, pp. 1126–1135.

[26] T. Li, Z. Zhou, S. Li, C. Sun, R. Yan, X. Chen, The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study, Mech. Syst. Signal Process. 168 (2022) 108653.
[27] C. Liu, X. Liao, L. Zheng, Y. Huang, H. Liu, Y. Zhang, H. He, H. Huang, J. Zhou, H. Jin, L-FNNG: Accelerating large-scale KNN graph construction on CPU-fpga heterogeneous platform, ACM Trans. Reconfigurable Technol. Syst. 17 (3) (2024) 1–29.
[28] T. Hospedales, A. Antoniou, P. Micaelli, A. Storkey, Meta-learning in neural networks: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 44 (9) (2021) 5149–5169.
[29] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks, 2017, arXiv preprint arXiv:1710.10903.
[30] Z. Wu, P. Jain, M. Wright, A. Mirhoseini, J.E. Gonzalez, I. Stoica, Representing long-range context for graph neural networks with global attention, Adv. Neural Inf. Process. Syst. 34 (2021) 13266–13279.
[31] Z. Guo, C. Zhang, W. Yu, J. Herr, O. Wiest, M. Jiang, N.V. Chawla, Few-shot graph learning for molecular property prediction, in: Proceedings of the Web Conference 2021, 2021, pp. 2559–2567.
[32] Q. Lv, G. Chen, Z. Yang, W. Zhong, C.Y.-C. Chen, Meta learning with graph attention networks for low-data drug discovery, IEEE Trans. Neural Networks Learn. Syst. (2023).
[33] D. Zhang, K. Zheng, Y. Bai, D. Yao, D. Yang, S. Wang, Few-shot bearing fault diagnosis based on meta-learning with discriminant space optimization, Meas. Sci. Technol. 33 (11) (2022) 115024.
[34] X. Hu, X. Zhang, Z. Wang, Y. Chen, J. Xia, Y. Du, Y. Li, Railway switch machine fault diagnosis considering sensor abnormality scenarios, in: 2023 IEEE 26th International Conference on Intelligent Transportation Systems, ITSC, IEEE, 2023, pp. 4834–4839.
[35] Q. Xiang, X. Wang, Y. Song, L. Lei, ISONet: Reforming 1DCNN for aero-engine system inter-shaft bearing fault diagnosis via input spatial over-parameterization, Expert Syst. Appl. 277 (2025) 127248.
[36] C. Zhang, N. Song, G. Lin, Y. Zheng, P. Pan, Y. Xu, Few-shot incremental learning with continually evolved classifiers, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 12455–12464.
[37] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A.A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, et al., Overcoming catastrophic forgetting in neural networks, Proc. Natl. Acad. Sci. 114 (13) (2017) 3521–3526.
[38] P.I. Frazier, A tutorial on Bayesian optimization, 2018, arXiv preprint arXiv:1807.02811.
[39] D. Chicco, G. Jurman, The advantages of the matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation, BMC Genomics 21 (2020) 1–13.
[40] M. Friedrich-Rust, M.-F. Ong, S. Martens, C. Sarrazin, J. Bojunga, S. Zeuzem, E. Herrmann, Performance of transient elastography for the staging of liver fibrosis: a meta-analysis, Gastroenterology 134 (4) (2008) 960–974.
[41] P. Yuan, C. Xie, L. Liu, H. Jin, PathGraph: A path centric graph processing system, IEEE Trans. Parallel Distrib. Syst. 27 (10) (2016) 2998–3012.
[42] A. Nichol, J. Schulman, Reptile: a scalable metalearning algorithm 3 (2018) 4 arXiv preprint arXiv:1803.02999.