algorithms

MDPI

Article

A Fault Diagnosis Method Based on Graph Convolutional Attention Networks for Complex Coupled Systems

Xiaojing Cao [1,*], Kaihe You [2], Xuguang Chai [2], Zetong Sun [3] and Ying Li [1]

[1] School of Cross-border E-commerce, Guangdong University of Finance & Economics, Guangzhou 510320, China
[2] Department of Information Engineering, Hebei Vocational University of Technology and Engineering, Xingtai 054000, China
[3] School of Physics and Optoelectronic Engineering, Jinan University, Guangzhou 511442, China
* Correspondence: caoxiaojing@gdufe.edu.cn

Abstract

Large-scale units, transmission lines, and distributed integrated equipment in Industry 5.0 exhibit complex characteristics, including multiple components, extensive sensor networks, and strong signal coupling. Conventional fault diagnosis methods typically rely exclusively on time-series and time-frequency features, overlooking the physical topological relationships among devices. This limitation leads to challenges in identifying complex faults, reduced accuracy in fault tracing, and poor interference resistance. To overcome these issues, this paper introduces a combined graph convolutional attention-based fault diagnosis method (CGCA), which synergizes Graph Convolutional Networks (GCN) with Graph Attention Networks (GAT). Initially, the industrial physical system is modeled as a weighted topological graph, with components and sensors acting as nodes. By constructing a weighted adjacency matrix based on equipment connections and signal coupling strengths, prior knowledge of the system topology is effectively embedded. Subsequently, a graph convolutional network aggregates global node coupling features, which are further processed by a multi-head graph attention network to adaptively assign node weights. This mechanism not only enhances the representation of fault source features but also enables simultaneous fault classification and localization. Extensive comparative experiments and noise robustness tests are performed on the CWRU bearing dataset and the PU gearbox dataset. The results indicate that, relative to benchmark algorithms, the proposed method delivers superior performance in fault classification and source localization under typical industrial operating conditions, consistently surpassing reference models in both classification and localization accuracy. Quantitatively, on the CWRU bearing dataset, CGCA reaches 98.6% classification accuracy, 98.2% F1-score, and 96.9%/99.1% Top-1/Top-2 fault-source localization; on the PU gearbox dataset it reaches 97.8% accuracy and 95.8%/98.7% Top-1/Top-2 localization and retains 82.5% accuracy at 0 dB noise.

Keywords: complex coupled system; fault diagnosis; fault tracing; graph convolutional network; graph attention network

Check for updates

Academic Editor: Carolina Del-Valle-Soto

Received: 27 August 2026
Revised: 4 November 2026
Accepted: 5 November 2026
Published: 19 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Intelligent Fault Diagnosis (IFD) of rotating machinery serves as a core technology for ensuring the operational safety and stable performance of key industrial equipment. In recent years, deep learning-based diagnostic methods have become the mainstream research paradigm due to their powerful end-to-end feature extraction capabilities; however,

several critical challenges remain unresolved in practical industrial scenarios. Most existing studies excessively prioritize diagnostic accuracy while neglecting model calibration under distribution shifts, leading to overconfident predictions by deep models where softmax probabilities fail to align with actual prediction reliability—particularly when encountering out-of-distribution (OOD) samples and distribution-shifted data caused by varying operating conditions [1]. Moreover, generic deep learning models are typically designed through empirical trial-and-error, which leads to limited interpretability of the network architecture and unclear physical meaning of the feature extraction process, thereby restricting their deployment in safety-critical industrial applications [2]. In addition, acquiring a large number of labeled fault samples is difficult in real-world industrial settings, and the collected vibration signals are frequently contaminated by strong noise. Traditional lightweight diagnostic models struggle to strike a balance among model complexity, feature extraction capability, and noise robustness, and thus fail to maintain stable diagnostic performance under small-sample and high-noise conditions [3]. Furthermore, continuous speed fluctuations under variable-speed operation render mechanical vibration signals inherently nonlinear and time-varying, and conventional convolutional neural networks (CNNs) cannot effectively capture the discriminative fault features in such circumstances, which further degrades diagnostic performance [4]. To address these issues, researchers have conducted exploratory studies from the perspectives of model calibration, interpretable network design, lightweight anti-noise modeling, and variable-speed adaptive feature learning, laying the foundation for the development of trustworthy, interpretable, and robust intelligent fault diagnosis technologies.

However, the practical deployment of intelligent fault diagnosis systems remains hampered by multiple data- and sensor-related bottlenecks in real industrial scenarios. First, the inherent data imbalance—characterized by an overwhelming number of healthy samples yet extremely scarce fault samples, owing to the short-lived faulty operation of rotating machinery—severely restricts the feature learning capability of deep diagnostic models and leads to biased and unstable classification results [5]. Although generative adversarial network (GAN)-based data augmentation methods have been developed to address this issue, traditional GANs suffer from poor generalization, insufficient diversity of generated fault features, and heavy reliance on complex data preprocessing, which considerably limits their effectiveness in extremely data-scarce fault scenarios [6]. Second, multi-sensor fusion diagnosis, widely adopted for comprehensive health monitoring of large-scale industrial equipment, faces the critical challenge of abrupt channel loss caused by sensor damage or harsh operating conditions. Conventional multi-channel diagnostic models rely entirely on complete input data and thus undergo severe performance degradation once partial sensor signals are lost, greatly compromising the reliability of online diagnosis [7]. Furthermore, when training data are limited, data-driven diagnostic models are plagued by out-of-distribution (OOD) generalization difficulties and dual-source uncertainty—including aleatoric uncertainty arising from measurement noise and epistemic uncertainty stemming from insufficient sample distribution coverage—making it difficult to distinguish genuine mechanical faults from model prediction deviations [8,9]. In response to these challenges, researchers have proposed targeted solutions such as generalization-regularized generative augmentation models, multi-scale feature recovery-based multi-channel networks, and uncertainty-aware consistency diagnosis frameworks. Nevertheless, jointly addressing data imbalance, incomplete sensor signals, and model uncertainty remains an urgent priority for the development of robust and reliable intelligent fault diagnosis systems.

As a pivotal technique for ensuring the health and safety of rotating machinery, Intelligent Fault Diagnosis (IFD) faces substantial barriers to industrial implementation. In

addition to the calibration-, interpretability-, and data-related challenges reviewed above, existing data-driven models are typically designed around a single diagnostic objective and operate directly on time-series or time-frequency features, without explicitly encoding the physical assembly and coupling relationships among components and sensors. As a consequence, they struggle to trace a fault back to its originating component and offer limited interpretability of which nodes actually contribute to the fault. To bridge this gap, this paper proposes a combined graph convolutional attention-based fault diagnosis method (CGCA) that (i) models the industrial system as a weighted topological graph in which components and sensors act as nodes, (ii) fuses time-frequency node features through a graph convolutional network (GCN) and a multi-head graph attention network (GAT) for adaptive fault-node weighting, and (iii) jointly outputs both the fault category and the Top-1/Top-2 fault-source localization. The method is validated on the CWRU bearing dataset and the PU gearbox dataset under both standard and noisy conditions.

2. Related Work

This section reviews recent advances in intelligent fault diagnosis and positions the proposed method relative to the existing literature. To address the obstacles identified in the Introduction, the research community has pursued a variety of complementary directions—including privacy-preserving collaborative learning, open-set recognition, few-shot learning, interpretable and physics-informed modeling, robust feature extraction, imbalance learning, and lightweight networks suitable for edge deployment [10].

To address the challenges of label scarcity, distributed data privacy, and unseen faults in fault diagnosis, privacy-preserving collaborative learning and open-set recognition have emerged as promising research directions. In scenarios where decentralized multi-source data cannot be directly shared due to privacy policies, federated learning has been widely adopted to enable collaborative model training without centralizing raw local data, thereby facilitating knowledge sharing while safeguarding data privacy. To tackle cross-domain diagnosis tasks involving unseen faults and domain drift, zero-shot learning has been introduced into fault diagnosis by integrating semantic knowledge with feature transfer mechanisms, enabling the model to recognize fault categories that remain unseen during training [11].

While progress has been made in privacy-preserving collaborative training and zero-shot recognition, the field still faces critical hurdles regarding few-shot fault diagnosis in data-scarce environments and effective graph-based feature modeling. Consequently, research has shifted towards graph contrastive learning and the rigorous validation of deep learning methods. Addressing the challenge of limited labeled data, Yang et al. [12] proposed a generalized graph contrastive learning framework that utilizes spectral features to build K-nearest neighbor graphs. By designing positive–negative graph pairs for unsupervised pre-training and semi-supervised fine-tuning, this method significantly improves diagnostic performance and generalization capabilities. In a related study, Jiang et al. [13] conducted a comprehensive benchmark of eight mainstream deep learning models. Their analysis revealed that, contrary to expectations, complex novel models do not yield significant accuracy gains compared to basic CNNs and LSTMs; instead, they suffer from higher computational overhead and poorer stability in small-sample, variable-condition scenarios. These insights expose the practical bottlenecks of current deep learning approaches, offering a strong foundation for developing lightweight, label-efficient, and highly generalized fault diagnosis solutions.

Furthermore, to improve federated learning performance under non-IID data and complex multi-working conditions, Zhang et al. [14] proposed a multiscale recursive attention-gated federated framework, which adaptively weights client contributions through attention

gates and integrates local multiscale features with globally aggregated information, effectively enhancing the accuracy and robustness of multi-condition fault diagnosis. For health monitoring of disturbances caused by lubricant contamination, Rameshkumar et al. [15] established statistical correlations linking vibration, acoustic emission, and sound features to solid particle contaminants in gearboxes, demonstrating that feature-level multi-sensor fusion achieved over 99% accuracy in lubrication condition recognition, thereby providing a reliable multi-sensor fusion scheme for lubrication-related mechanical health assessment. Additionally, to address data scarcity and class imbalance, Sifat et al. [16] conducted a systematic review of GAN-based data augmentation for rolling bearing fault diagnosis. They concluded that WGAN-GP and cGAN are the most effective variants, identified data imbalance, training instability, and insufficient samples as core challenges, and provided standardized guidance for developing high-performance data augmentation strategies in small-sample fault diagnosis.

In an effort to overcome the physical inconsistency and lack of interpretability associated with purely data-driven approaches, Lu et al. [17] introduced a physics-informed temporal convolutional network (PITCN) for detecting partial discharge in overhead covered conductors. By embedding physical constraints for noise suppression and pulse identification into a specialized physics-data loss function, the model ensures predictions align with underlying fault mechanisms while yielding interpretable PD pulses. To tackle fault diagnosis in nonstationary and high-noise environments, Li et al. [18] proposed a joint threshold learning convolutional network (JT-LCN). This framework features a plug-and-play joint thresholding module for adaptive denoising, leverages depthwise separable convolution for a lightweight structure, and incorporates dynamic self-knowledge distillation to boost generalization efficiently. Furthermore, to handle imbalanced fault diagnosis with scarce fault samples in offshore wind turbines, Lei et al. [19] developed an adaptive weighted cost-sensitive dense convolutional network (CSIDNN). This approach combines large-kernel convolution with interactive dense connections for robust feature extraction, uses self-normalizing activation to stabilize training, and employs an adaptive weighted class balance loss to dynamically adjust misclassification costs, resulting in a substantial increase in the detection accuracy of minority fault classes.

To further enhance privacy security and aggregation robustness in federated fault diagnosis, Ma et al. [20] proposed a multi-level federated learning framework based on cloud-edge-client collaboration, which enables layer-by-layer restricted sharing of model parameters to strengthen data privacy protection, uses the Euclidean metric to identify outliers, adopts centralized Kalman filtering for outlier-tolerant fusion, and employs sequential Kalman filtering to achieve asynchronous aggregation for handling transmission delay and packet loss. For cross-domain fault diagnosis involving domain shift, data scarcity, and unseen faults, Abbas et al. [21] developed an enhanced transfer learning framework that integrates the Universal Adaptation Network (UAN), Spectral-normalized Neural Gaussian Process (SNGP), WideResNet, and attention mechanisms. This framework enables reliable domain adaptation, open-set recognition, and uncertainty quantification, thereby improving diagnostic robustness under variable operating conditions. To address the shortage of labeled data and domain shift in rotating machinery diagnosis, Iqbal et al. [22] presented a Triplex Transfer LSTM (TTLSTM) method, which combines empirical mode decomposition (EMD) and the Pearson correlation coefficient (PCC) for feature extraction and selection, and incorporates L2 regularization during transfer fine-tuning to mitigate overfitting, thereby enabling high-precision fault identification with limited labeled samples in the target domain.

In the domain of signal-level interpretable feature extraction and transformer-based diagnosis, Yi et al. [23] introduced a hybrid approach combining wavelet packet analysis

with the Vision Transformer (ViT) for diagnosing wind tunnel valve leakage. By extracting the maximum component energy and its variation rate as sensitive indicators, the proposed ViT model demonstrated superior accuracy and stability over traditional CNNs and ResNets. Addressing the challenges of noise interference and sample scarcity in rotating machinery, Li et al. [24] proposed MFSFormer. This framework incorporates multiscale convolutional embedding and fuse-shuffle attention to capture multiscale features and channel-window dependencies, thereby significantly enhancing diagnostic robustness in noisy and few-shot environments. Additionally, Xu et al. [25] presented a multi-receptive field denoising residual convolutional network (MF-DRCN) featuring adaptive feature integration. This design improves discriminative feature extraction while effectively suppressing noise, resulting in excellent anti-noise performance and diagnostic accuracy even under strong noise and non-stationary operating conditions.

In an effort to overcome the dual hurdles of limited annotations and lack of interpretability, Wang et al. [26] proposed a multiview contrastive shapelet learning (MCSL) framework. By synergizing supervised and self-supervised contrastive learning, this method extracts discriminative shapelet features from scarce labeled data, providing both intuitive interpretability and robust diagnostic performance for rotating machinery. To facilitate speed-independent fault detection in variable operating scenarios, Kannan et al. [27] presented an enhanced quadratic discriminant analysis (QDA) method that incorporates vibration–sound signals and mRMR feature selection, ensuring stable multi-component fault identification across untrained rotational speeds. Additionally, targeting noise-resilient diagnosis within IoT-enabled sensing–computing environments, Jain et al. [28] developed a CNN–LSTM architecture activated by an enhanced exponential linear unit (EELU). This design is capable of preserving negative signal features while suppressing strong noise interference, resulting in exceptional accuracy, robustness, and real-time capability under heavy Gaussian noise conditions.

For multi-signal fault diagnosis in rotating machinery, Kumar et al. [29] conducted a comparative evaluation of supervised deep feed-forward neural networks (DFFNN) and unsupervised autoencoder models utilizing vibration, sound, and acoustic emission signals. The study demonstrated that the stacked autoencoder (SAE) offers a superior trade-off, achieving higher classification accuracy while significantly reducing computation time by 40% and lowering memory usage, thus facilitating automated real-time monitoring. To tackle the issue of data scarcity in electromagnetic pulse valves, Wang et al. [30] introduced a transfer attention fusion convolutional neural network (TAFCNN) leveraging dynamic simulation data. This framework effectively bridges the sim-to-real gap through parameter fine-tuning and employs an attention mechanism to fuse multi-sensor features, resulting in over 93% diagnostic accuracy using only about 8% real data. Additionally, Feng et al. [31] presented an unsupervised feature extraction framework based on GraphSAGE. By mining sample relationships through K-nearest neighbor fault graphs, this method ensures robust diagnostic performance even in scenarios with extremely limited labeled samples, surpassing the capabilities of conventional CNN and LSTM models.

In an effort to tackle missing sensor data and facilitate edge-deployable fault diagnosis for rotating machinery, Yu et al. [32] introduced GRUDMU-DSCNN. This framework synergizes a gated recurrent unit with decay and maskless update (GRUDMU) to adaptively impute missing values under interval missing and missing-completely-at-random (MCAR) patterns, while employing depthwise separable convolution (DSCNN) for lightweight feature extraction. Consequently, the model delivers superior accuracy and faster inference on edge devices, all while preserving robustness against incomplete time-series data. Addressing fault detection and diagnosis (FDD) in chemical processes, Kai et al. [33] conducted a systematic investigation into the impact of temporal subsequence length on

various supervised deep learning models, such as MLP, CNN, LSTM, transformer, and industrial process optimization-vision transformer (IPO-ViT). The findings demonstrate that IPO-ViT achieves optimal FDD performance on CSTR and TEP benchmarks, particularly when labeled training data is scarce, owing to its multi-head attention mechanism, which excels at capturing long-range temporal dependencies. Additionally, to handle nonlinear industrial processes characterized by high-dimensional data, Li et al. [34] presented a sparse Gaussian feature extractor (SGFE) incorporating latent label consistency. By projecting raw data into a fault-labeled orthogonal feature space and constraining features to follow multiple sparse multivariate Gaussian distributions, this approach significantly enhances linear separability and interpretability, leading to improved diagnostic accuracy for elusive faults on the Tennessee Eastman (TE) process dataset.

To directly extract discriminative features from raw vibration signals and enhance noise robustness and domain adaptation capabilities, Zhang et al. [35] developed the Wide First-layer Kernels Deep Convolutional Neural Network (WDCNN), which employs wide kernels in the first convolutional layer to suppress high-frequency noise and integrates Adaptive Batch Normalization (AdaBN) to achieve effective cross-working-condition domain adaptation without manual feature engineering. For few-shot fault diagnosis with scarce labeled samples, Xu et al. [36] proposed an attention-guided residual spatiotemporal network (AR-CLSTM) that fuses dual attention mechanisms, residual connections, LSTM, and label smoothing regularization, enabling robust spatiotemporal feature learning and high-precision fault identification in small-sample scenarios. To address the challenges of unlabeled multi-sensor data and temporal heterogeneity, Lu et al. [37] designed a temporal heterogeneous relationship-enhanced deep clustering network (THRDCN), which embeds heterogeneous graph convolution into a gated recurrent unit (GRU) to construct the HG-GRU module for modeling temporal heterogeneous relationships, and achieves end-to-end unsupervised fault diagnosis via a relationship-enhanced KL divergence loss.

Despite significant strides in tackling specific issues such as label scarcity, privacy preservation, and noise robustness, existing methods are predominantly designed around a single diagnostic objective and typically process time-series or time-frequency features in isolation, without explicitly encoding the physical assembly topology and the coupling relationships among components and sensors. As a result, such methods cannot simultaneously identify the fault category and trace the root cause to its originating component in complex, coupled electromechanical systems, and they provide limited interpretability regarding which nodes actually contribute to the fault. To bridge this gap, this paper develops a unified graph-based fault diagnosis framework that embeds physical topology priors through a weighted adjacency matrix and jointly performs fault classification and fault-source localization [38,39].

3. Overall Architecture of the CGCA Network

Diagnostic tasks for complex, coupled electromechanical systems in Industry 5.0 are hindered by issues such as fault coupling, propagation, and the difficulty of decoupling compound faults. Conventional approaches relying on temporal or time-frequency models (e.g., CNNs, LSTMs) are limited by their exclusive focus on data-dimensional feature extraction. This prevents the incorporation of physical assembly topology and component coupling priors, resulting in insufficient robustness against noise and an inability to accurately trace fault sources or identify subtle anomalies [40,41]. In response, this study constructs a unified Graph Convolutional Attention (CGCA) framework that synergizes the global topological aggregation advantages of GCNs with the adaptive weight allocation characteristics of GATs. During feature construction, time-frequency multi-domain information is fused with node coupling correlation weights. In the encoding stage, GCNs

are deployed to reveal global transmission coupling patterns, while GATs are utilized to adaptively amplify the representation weights of faulty nodes. This approach enables the simultaneous output of fault diagnosis and component localization, effectively realizing end-to-end modeling from raw sensor data to intelligent identification and root cause tracing.

Figure 1 illustrates the overall architecture of the proposed CGCA, a graph convolutional attention-based fault diagnosis network. Following a full-chain computational pipeline, it takes raw multi-sensor time-series vibration signals as input and passes them successively through the input layer, the graph feature encoding layer, the global pooling layer, the diagnostic output layer, and the source tracing and localization layer, finally producing both the fault classification result and the fault-source localization result. This workflow aligns with the fault diagnosis mechanism of multi-component coupled equipment in Industry 5.0. The internal node feature generation submodule corresponding to the input stage is detailed separately in Figure 2.

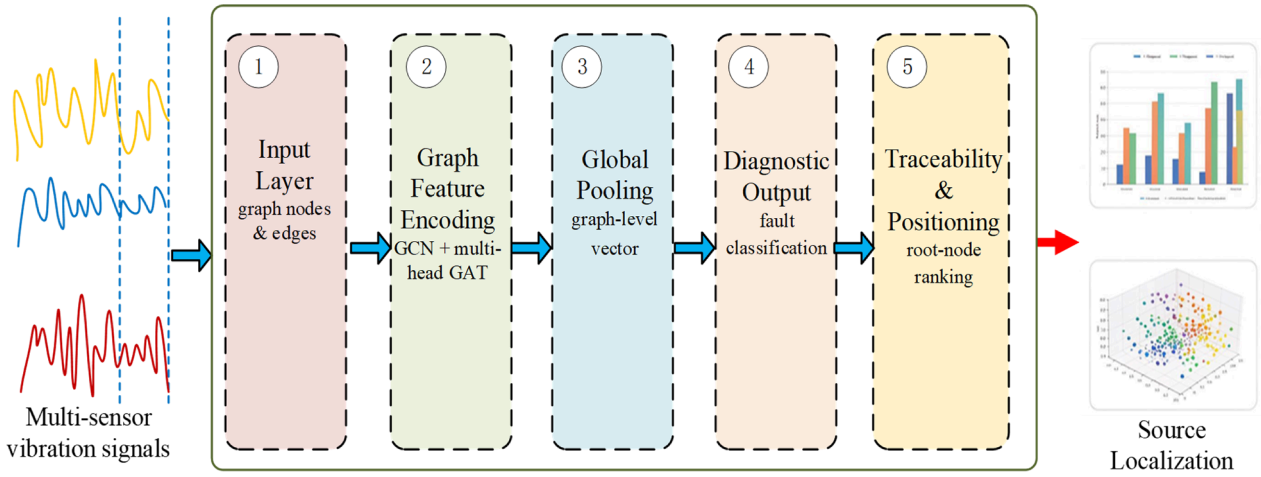

Figure 1. Overall architecture of fault diagnosis framework for complex coupled systems.

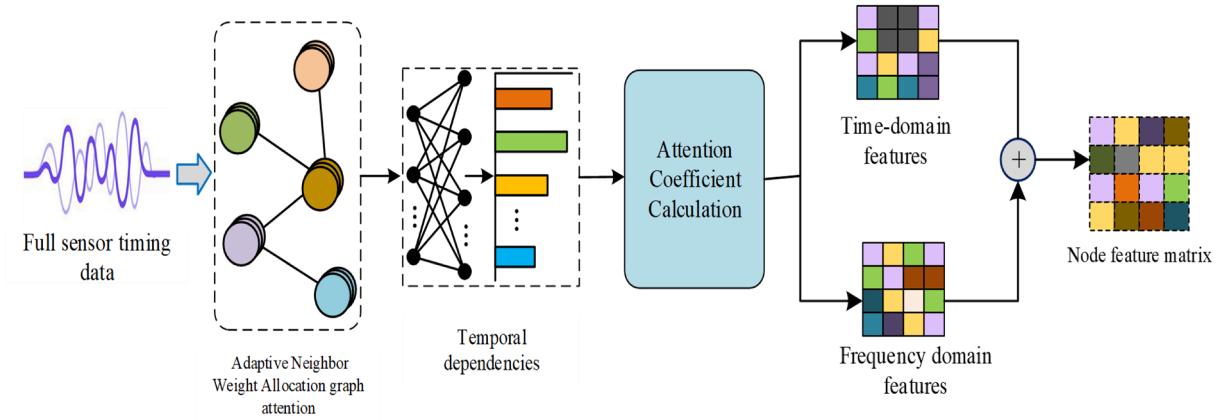

Figure 2. Graph attention-based node feature generation: adaptive neighbor weight allocation and time–frequency domain feature fusion.

3.1. Top-Level Architecture

The overall network consists of a five-level serial forward inference process, divided into five sequential modules: the input layer, graph feature encoding layer, global pooling layer, diagnostic output layer, and source tracing and localization layer.

(1) Input Layer: Input of multi-sensor time-series signals and graph structures

The multiple colored curves on the left represent raw vibration time-series signals collected from sensors at different measurement points on the unit or production line. These signals are segmented along the time dimension_$_t$ (segment length C). After preprocessing, they are combined with the physical topological relationships of the equipment to construct graph input data: components and sensors serve as graph nodes, while adjacency relationships and coupling weights form topological edges, creating standardized graph samples that are fed into the network.

(2) Graph Feature Encoding Layer (G_1~G_n): GCN + GAT Joint Feature Extraction

This layer is the core of the model, corresponding to the algorithm's graph convolutional + multi-head graph attention-based feature learning. represents a stacked multi-layer CGCA structure: GCN aggregates global topological coupling features using a weighted adjacency matrix, while GAT adaptively assigns node association weights via attention coefficients. The fusion yields deep node features that account for both global device associations and local fault characteristics, achieving the integration of topological prior knowledge with data features.

(3) Global Pooling Layer ($F_1 \sim F_n$)

Global average pooling is performed on the encoded graph-wide node features, compressing the node feature matrix of dimension $N \times D$ into a one-dimensional global feature vector. This achieves the transformation from "node-level features" to "graph-level global features," providing aggregated features for subsequent fault classification.

(4) Diagnostic Output Layer ($F_1' \sim F_n'$)

Composed of a fully connected layer and a Softmax layer, this layer performs fault classification based on the global graph features, outputting probability distributions for various fault types (normal, single fault, composite fault), corresponding to the fault diagnosis classification results shown in the bar chart on the right.

(5) Root Cause Analysis Layer: Fault Source Localization Based on Attention Weights

Leveraging the node attention coefficients retained from the GAT training phase, a node association weight mapping network is constructed. By calculating the fault contribution scores S_i for each node, all nodes are ranked by contribution, and the top 1 or top 2 root fault nodes are identified. Finally, the three-dimensional scatter plot of fault traceback and localization results on the right is output, enabling precise localization of the faulty component.

3.2. Sub-Block Diagram of Lower-Level Amplification

As shown in Figure 2, this submodule provides a complete description of the computational steps from raw sensor data to the multi-domain fusion node feature matrix. It represents the generation process of the graph model input, as well as the refinement process of the node feature matrix and the internal implementation of the graph feature encoding layer.

(1) Acquisition and Preprocessing of Full-Sensor Time-Series Data

Raw vibration data from all sensor points is imported at the input end. Signal acquisition, segmentation, and standardization/normalization preprocessing are performed sequentially to remove abnormal noise and outliers, resulting in a standardized time-series sequence.

(2) Attention-Based Adaptive Allocation of Neighbor Weights and Modeling of Temporal Dependencies

Leveraging the GAT attention mechanism, neighbor node weights are adaptively assigned based on node coupling strength quantified by Pearson correlation coefficients; simultaneously, sensor temporal dependencies S are mined and integrated with temporal dynamic patterns to provide temporal prior information for attention coefficient calculation.

(3) Solving for Attention Coefficients α and Extracting Temporal Features S_t

Normalized attention coefficients between nodes α are calculated using the attention formula, and a time-domain feature matrix S_t is obtained through weighted aggregation based on these coefficients, representing the time-domain statistical characteristics of the signal (kurtosis, peak, mean, variance, etc.).

(4) Extraction of Frequency-Domain Features F_t and Multi-Domain Weighted Fusion

A spectral transform is applied to the raw signal to extract the frequency-domain feature matrix F_t; a weighted fusion strategy is adopted, using α and $1 - \alpha$ as the fusion weights for the time-domain and frequency-domain features, respectively, to perform weighted summation: $X = \alpha \cdot S_t + (1 - \alpha) \cdot F_t$. This ultimately generates the node feature matrix, which serves as the raw node input feature X for the upper-layer graph feature encoding layer CGCA.

Here, the attention coefficients that serve as the time-frequency fusion weights are produced by the same GAT attention mechanism used in the graph encoding stage: for each node, the coupling strength toward its neighbors (quantified by the Pearson correlation coefficients between the corresponding sensor signals) is mapped through the attention scoring function and Softmax normalization to yield the fusion weights. These weights are not fixed heuristics; they are computed in the forward pass and are optimized jointly with the network through end-to-end backpropagation of the joint loss (classification cross-entropy plus attention regularization) described in Section 4.4.

The entire framework implements a dual-driven modeling approach that combines physical topology mechanisms with multi-domain signal features: on the one hand, it constructs a weighted topological graph based on equipment assembly relationships to incorporate prior knowledge of industrial equipment structures; on the other hand, it optimizes node feature representations through multi-domain spatiotemporal feature fusion and adaptive attention-based weighting. Specifically, the GCN uncovers global component coupling patterns, while the GAT identifies the attention weights of critical fault-related nodes, thereby concurrently accomplishing two major tasks: fault classification and diagnosis, as well as fault source tracing and localization.

4. Graph Construction and Implementation Details

4.1. System Topology Construction

The industrial coupling system is abstracted as an undirected weighted graph G = (V,E,A,X), where V is the set of nodes, V = $\{v_1, v_2, \ldots, v_N\}$, and N represents the total number of system nodes. Nodes are divided into two categories: component nodes (physical units such as unit bearings, motors, gearboxes, and pump bodies) and sensor nodes (vibration, temperature, current, and pressure sensors); a single component may be equipped with multiple sensors. E is the set of edges, where $(v_i, v_j) \in E$ indicates a physical connection or signal coupling between nodes v_i and v_j. Edges are categorized as: component-to-component edges (mechanical transmission, piping connections), component-to-sensor edges (sensor installation relationships), and sensor-to-sensor edges (signal coupling relationships). $A \in R^{N \times N}$: Adjacency matrix, describing node connections and coupling strength; weighted adjacency better reflects industrial reality: $A_{ij} = \begin{cases} w_{ij}, & (v_i, v_j) \in E \\ 0, & Others \end{cases}$ $w_{ij} \in (0,1]$ represents edge weights, determined by equipment coupling density and signal correlation coefficients; if no connection exists, the weight is 0. $X \in R^{N \times F}$: Node feature matrix, F represents the feature dimensions of a single node. Each row corresponds to a node, and features consist of sensor time-series statistical characteristics (mean, variance, kurtosis, frequency-domain energy, time-domain peak, etc.).

Topology construction rules are established based on the actual structure of industrial equipment. First, a component topology skeleton is built according to physical assembly logic; for example, a chain-like structure is constructed in which the motor, coupler, gearbox, and load are connected in sequence. Second, each sensor is associated with its corresponding component to establish subordinate connection edges between components and sensors. Third, for sensors whose signals share the same origin or exhibit significant coupling interference, additional inter-sensor association edges are added. Finally, the weight of each connection edge is calculated using the Pearson correlation coefficient to quantitatively characterize the signal coupling strength between nodes. Pearson correlation encodes statistical co-variation rather than strict physical causality; it serves as a practical coupling prior that is further refined end-to-end by attention weights, and the resulting root-cause ranking should be read as an attention-assisted suggestion rather than a causal proof. For the CWRU dataset the graph connects the drive-end, fan-end and base sensor nodes to the bearing component node along the shaft load path (4 nodes); for the PU dataset the graph follows the motor-clutch-gearbox-load chain, with accelerometer nodes attached to the corresponding gear, shaft and bearing component nodes.

4.2. Fault Diagnosis Model Integrating Graph Convolutional Networks and Graph Attention

The model consists of five major modules: the input layer, the graph feature encoding layer (GCN layer + GAT layer), the global pooling layer, the diagnostic output layer, and the source tracing and localization layer. $X(l) \in R^{N \times D_l}$: the graph feature matrix of the l layer, where D_l is the feature dimension of that layer; $\tilde{A}$: the normalized adjacency matrix; W: the network's learnable weight matrix; $\sigma(\cdot)$: the nonlinear activation function (ReLU); N: the total number of graph nodes.

GCNs uncover global topological correlations to aggregate features from neighboring nodes, learn global coupling patterns among components and sensors, and address the challenge of extracting interdependent signal features across multiple nodes.

(1) Adjacency Matrix Normalization

To avoid gradient explosion in deep neural networks, symmetric normalization is employed.

$$\tilde{A} = I_N + A \quad \hat{A} = D^{-\frac{1}{2}} \tilde{A} D^{-\frac{1}{2}} \tag{1}$$

I_N is the identity matrix (adding self-loops to preserve the node's own features); D is the degree matrix, $D\text{ii} = \sum_{j=1}^{N} \tilde{A}_{ij}$, a diagonal matrix.

(2) GCN Forward Propagation Formula

Feature update in a single-layer GCN.

$$X^{(l+1)} = \sigma\left(\hat{A} X^{(l)} W^{(l)}\right) \tag{2}$$

Stacking two layers of GCNs performs first- and second-order neighbor feature aggregation, outputting intermediate features X_{gcn} that integrate topological relationships. Two GCN layers are chosen because the tested bearing and gearbox systems are shallow, chain-coupled structures in which a fault typically propagates over at most two physical hops; deeper stacking (three or more layers) was examined but produced marginally lower accuracy and heavier over-smoothing on these small graphs.

$$X_{gcn} = GCN_2\left(GCN_1\left(X, \hat{A}\right)\right) \tag{3}$$

4.3. Graph Attention Network: Enhancing the Weights of Critical Fault Nodes

In multi-fault scenarios, the impact of a faulty node on its neighbors varies. The GAT leverages an attention mechanism to adaptively assign weights to neighboring nodes,

thereby highlighting the features of the fault source node and providing a solid basis for fault tracing.

(1) Calculation of Attention Coefficients

For node i and its neighbors j, the original attention scores.

$$e_{ij} = \text{LeakyReLU}(a^T[Wx_i \| Wx_j]) \tag{4}$$

$\|$ denotes feature concatenation; a is the attention weight vector, and x_i is the feature vector of node i.

Apply Softmax normalization to the scores to obtain normalized attention coefficients (representing the strength of the influence between nodes):

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in N(i)} \exp(e_{ik})} \tag{5}$$

N(i) is the set of first-order neighbors of node i.

(2) GAT Node Feature Update

Aggregate neighbor features by weighting them with the attention coefficients.

$$x_i' = \sigma\left(\sum_{j \in N(i)} \alpha_{ij} W x_j\right) \tag{6}$$

The model employs multi-head attention to enhance robustness, with the outputs from the K heads concatenated.

$$x_i^{gat} = \|_{k=1}^{K} \sigma\left(\sum_{j \epsilon N(i)} \alpha_{ij}^k W^k x_j\right) \tag{7}$$

Finally, the GCN output features and GAT output features are fused through element-wise addition to obtain the graph-wide deep features: Element-wise addition is preferred over concatenation because the GCN and GAT branches share the same node feature dimension; concatenation would double the feature width and enlarge the classifier parameters without adding topology, whereas a learnable weighted fusion was tested but gave negligible gain at extra training cost. Addition therefore keeps the model lightweight while retaining both global coupling and local attention signals.

$$X_{fusion} = X_{gcn} + X_{gat} \tag{8}$$

The feature map Z is processed through a Softmax layer to produce three weight coefficients: a_i, b_i, and c_i. In the CGCA framework, these three coefficients are the attention-derived fusion weights for the time-domain feature stream, the frequency-domain feature stream, and the graph-encoded feature stream, respectively; they are produced by the Softmax layer from the fused graph-level features and are learned jointly with the network during training rather than being hand-set constants. The corresponding mathematical expressions are as follows:

$$a_i = \frac{e^{A_i Z}}{e^{A_i Z} + e^{B_i Z} + e^{C_i Z}} \tag{9}$$

$$b_i = \frac{e^{B_i Z}}{e^{A_i Z} + e^{B_i Z} + e^{C_i Z}} \tag{10}$$

$$c_i = \frac{e^{C_i Z}}{e^{A_i Z} + e^{B_i Z} + e^{C_i Z}} \tag{11}$$

$$a_i + b_i + c_i = 1 \tag{12}$$

$$V = a_i \times U_1 + b_i \times U_2 + c_i \times U_3 \tag{13}$$

In the formula: a_i, b_i, and c_i represent the soft attention vector weights for U_1, U_2, and U_3, respectively; $e^{A_i Z}$ denotes the weight response index for the ith convolutional kernel; A_i

denotes the ith row of the linear transformation matrix A, with B_i and C_i following the same logic; V denotes the final output feature map. For clarity, these weights are the learnable parameters of the graph feature fusion layer; they linearly transform the fused node features and are trained jointly with the GCN and GAT parameters in an end-to-end manner.

4.4. Fault Diagnosis Branch

Perform global average pooling (GAP) on the fused features to map the graph-level features to a global vector.

$$h_{global} = GAP\left(X_{fusion}\right) \tag{14}$$

A fully connected layer with a Softmax activation function is then applied to perform fault classification (normal, fault 1, fault 2, ..., multiple composite faults).

$$y = Softmax\left(h_{global}W_{fc} + b_{fc}\right) \tag{15}$$

The output y is a probability vector for each class; the class with the highest probability is selected as the diagnostic result.

Fault Traceback and Node Localization Branch

Fault tracing is performed using GAT attention coefficients. When a fault occurs, the fault source node generates significantly higher attention weights for its neighboring nodes; the average attention contribution of all nodes in the entire graph is then calculated.

$$S_i = \frac{1}{|N(i)|}\sum\nolimits_{j\in N(i)} \alpha_{ji} \tag{16}$$

where S_i let this be the fault contribution score for node i; the higher the score, the greater the probability that the node is the root fault node (fault source). Sort all nodes in descending order by S_i; the Top-1/Top-2 nodes are identified as the faulty components/sensors, thereby achieving end-to-end fault tracing.

Loss Function

Cross-entropy loss is used for fault classification training, combined with an attention regularization term to constrain tracing accuracy.

$$L = L_{ce} + \lambda L_{reg} \tag{17}$$

The classification cross-entropy loss is as follows.

$$L_{ce} = -\frac{1}{M}\sum\nolimits_{m-1}^{M}\sum\nolimits_{c=1}^{C} y_{m,c}\log\left(\hat{y}_{m,c}\right) \tag{18}$$

where M is the number of samples, C is the number of fault categories, $y_{m,c}$ is the true label, and $\hat{y}_{m,c}$ is the predicted probability. Attention regularization term: constrains the discriminative power of fault node scores, whereλ is the balancing coefficient.

4.5. Algorithm Execution Steps

The overall algorithm is divided into five execution steps: Step 1: System topology modeling and graph structure initialization; Step 2: Preprocessing of multi-source sensor data and construction of node time-frequency features; Step 3: Deep encoding of graph features using CGCA; Step 4: Fault classification inference; Step 5: Fault root cause analysis and localization based on attention weights. These steps fully implement the end-to-end process from raw data input from the industrial site to fault category determination and root cause component localization.

Step 1: System Topology Modeling and Weighted Graph Initialization

Based on equipment mechanical assembly and sensor installation layout, an undirected weighted topological graph (G = (V,E,A,X)) is initialized according to established topological construction rules:

(1) Division of the node set V: The system's physical units are categorized into two types of nodes—mechanical component nodes (motors, bearings, gearboxes, drive shafts, etc.) and sensor nodes (vibration, temperature, and current measurement points). All nodes are aggregated to form the universal node set $V = \{v_1, v_2, ..., v_N\}$, where N represents the total number of system nodes;

(2) Basic Topology Edge ε Construction: ① Based on the physical assembly and transmission relationships of the equipment, establish a skeletal connection between components (e.g., motor → coupling → gearbox → load) to generate component-to-component connection edges; ② Establish sensor-to-component dependency edges based on the physical installation hierarchy of sensors; ③ For sensors with common measurement points or significant signal coupling interference, add sensor-to-sensor association edges to complete the construction of the full system edge set ε;

(3) Assigning values to the weighted adjacency matrix A: Calculate edge weights $w_{ij} \in (0, 1]$ using the Pearson correlation coefficients of the raw time-series signals from the sensors corresponding to each node. Assign a value of $A_{ij} = w_{ij}$ to positions where a connection exists $(v_i, v_j) \in E$, and set the corresponding positions of nodes without physical coupling connections to $A_{ij} = 0$. This ultimately generates a weighted adjacency matrix that quantifies the strength of component/sensor coupling.

Step 2: Preprocessing of multi-sensor time-series data and generation of the node feature matrix X

For raw vibration time-series data collected in real-time from multiple sensors in industrial settings, construct a standardized input feature matrix $X \in R^{N \times F}$ for each node (where F represents the feature dimensions for a single node):

(1) Raw signal preprocessing: Apply wavelet threshold denoising to all sensor channels to remove Gaussian noise from the industrial environment, eliminate outliers caused by sensor anomalies, and use $x^* = \frac{x - \mu}{\sigma}$ standardization to eliminate dimensional differences; along the time axis, segment long time series into fixed-length samples to unify the number of time points per sample;

(2) Domain-specific feature extraction: For the preprocessed single-node time-series signals, time-domain features (statistical metrics such as mean, variance, kurtosis, peak, and waveform factor, constituting the time-domain feature matrix S_t) and frequency-domain features (frequency-domain metrics such as FFT spectral energy and fundamental frequency amplitude, constituting the frequency-domain feature matrix F_t) are extracted separately;

(3) Time-frequency adaptive weighted fusion: Using the GAT attention coefficients α as fusion weights, the time-frequency features of each node are fused using $x_i = \alpha \cdot S_{t,i} + (1 - \alpha) \cdot F_{t,i}$. The features of all N nodes are concatenated to obtain the model's input node feature matrix X.

Step 3: CGCA Hierarchical Graph Feature Encoding (Core forward computation of the model, corresponding to the graph feature encoding layer in the structural graph)

Using the weighted adjacency matrix A and node features X as inputs, this stage implements deep feature learning through a hierarchical approach combining global topological aggregation and locally attention-weighted learning.

GCN global topological feature extraction:

(1) Add self-loops to the original adjacency matrix and perform symmetric normalization: $\tilde{A} = I_N + A$, $= D^{-\frac{1}{2}} \tilde{A} D^{-\frac{1}{2}}$, D are the degree matrices corresponding to $\tilde{A}$;

(2) Stack two layers of GCN to aggregate first- and second-order neighbor coupling features layer by layer, followed by a forward pass: $X^{(l+1)} = \sigma\left(X^{(l)} W^{(l)}\right)$, $\sigma = ReLU$ are the activation functions, $W^{(l)}$ are the trainable weights for the l layer, and the output is a feature that integrates global topological correlations: X_{gcn}.

Multi-head GAT Local Fault Feature Enhancement: Based on GCN output features, multi-head attention weighting is performed to adaptively amplify the feature weights of faulty nodes:

(1) Calculate the raw attention scores between nodes i and j: $e_{ij} = LeakyReLU\left(a^T\left[Wx_i \| Wx_j\right]\right)$, where $\|$ denotes feature concatenation;

(2) Apply Softmax normalization to the scores to obtain attention coefficients $\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k\epsilon N(i)} \exp(e_{ik})}$ and $N(i)$, representing the first-order neighbors of node i;

(3) Set up parallel computation for the $K = 8$ head attention, and concatenate the multi-head outputs to obtain the GAT feature X_{gat};

Feature Fusion:

Element-wise addition is used to implement two-branch feature fusion: $X_{fusion} = X_{gcn} + X_{gat}$, yielding deeply fused features that balance global device coupling and local fault characteristics.

Step 4: Global Pooling + Fault Classification Output (corresponding to the global pooling layer and diagnostic output layer in the structural diagram)

(1) Global Average Pooling: Perform global average pooling (GAP) on the node-fused features X_{fusion} from $N \times D$, compressing them into a one-dimensional grid-level global feature vector h_{global}, thereby achieving a dimensionality transformation from node-level features to system-wide operational condition features;

(2) Fault Classification Prediction: The global features are fed into a fully connected layer with a Softmax activation function: $y = Softmax\left(h_{global}W_{fc} + b_{fc}\right)$, where y represents the probability distribution vectors for various fault categories (normal, single fault, and multiple composite faults). The index with the highest probability is selected as the fault category diagnosis result for the current sample.

Step 5: Fault Traceback and Localization Based on Attention Coefficients

Based on the normalized attention coefficients obtained from GAT training, the fault contribution of each node is quantified to achieve precise fault source localization:

(1) Calculation of Node Fault Contribution Score: Define the node fault contribution metric $s_i = \frac{1}{|N(i)|}\sum_{j\in N(i)} \alpha_{ji}\ (i)$. A higher value of S_i indicates a greater probability that the node is the root cause of the fault;

(2) Ranking and localization of faulty nodes: Sort all nodes in the system in descending order based on S_i. The Top-1 and Top-2 nodes are identified by the model as the faulty component or sensor nodes;

In complex failure scenarios, multiple actual failure source nodes simultaneously exhibit high contribution scores, whereas secondary failure nodes—affected by fault propagation—exhibit significantly lower scores. This distinction effectively separates the original failure source from abnormal measurement points induced by fault propagation, thereby addressing the missed-detection problem that traditional methods face in complex failure scenarios.

The network's full set of parameters was optimized using a joint loss function: $L = L_{ce} + \lambda L_{reg}$, where L_{ce} represents the cross-entropy loss for fault classification, and λL_{reg} represents the attention regularization term, which enhances the discriminative power of fault node scores. The Adam optimizer was selected to iteratively update all weights of the GCN, GAT, and fully connected layers. Early stopping was triggered when the validation set accuracy showed no improvement over multiple consecutive iterations, at which point the model was saved upon convergence for subsequent online fault diagnosis and root cause analysis.

5. Simulation Experiments

The evaluation metrics adopted in this study include: Accuracy: The proportion of samples correctly predicted, measuring the model's overall classification capability; Precision: The proportion of samples correctly classified as a specific fault type among those predicted to belong to that class, reflecting the model's misclassification rate; Recall: The proportion of samples that actually belong to a specific fault type that are successfully identified, reflecting the model's omission rate; F1 Score: The harmonic mean of precision and recall, providing a comprehensive evaluation of the model's classification performance; Top-1/Top-2 Root Cause Accuracy: Metrics specific to fault root cause analysis. Top-1 represents the probability that the model's top-priority node is the actual fault source; Top-2 represents the probability that the actual fault source is among the top two high-risk nodes identified by the model. Higher values indicate more precise fault localization. Each input window contains 1024 sampling points; the hidden dimension of the GCN and GAT layers is 32, and the number of attention heads is 4; the final fused node feature vector is 32-dimensional (16 time-domain statistics concatenated with 16 frequency-domain spectral amplitudes); training uses the Adam optimizer with a learning rate of $1 \times 10^{-3} \times 10^{-3}$, batch size 64, and 100 epochs. All experiments are repeated five times, and the mean results are reported. For reproducibility, samples from every operating condition are randomly split into training, validation, and test subsets with a fixed random seed, and the identical partition is used across all compared models. Each experiment is repeated five times with different seeds, and the mean $\pm$ standard deviation over the five runs is reported. The network is trained with the Adam optimizer (learning rate $1 \times 10^{-3} \times 10^{-3}$, batch size 64, up to 100 epochs) with early stopping based on the validation accuracy.

The comparison models in the experiment include the following:

CWT-CNN [42]: A mainstream time-frequency analysis method for fault diagnosis that uses the Continuous Wavelet Transform (CWT) to convert one-dimensional time-series signals into two-dimensional time-frequency maps, combined with a convolutional neural network for classification;

ResNet [43]: A deep residual network that utilizes a residual structure to mitigate gradient vanishing in deep networks and excels at extracting deep-layer features;

CNN-SVM [44]: A hybrid model combining "deep learning and traditional algorithms," where CNN handles feature extraction and SVM performs classification;

CNN-BiGRU [45]: CNN extracts local spatial features, while bidirectional GRU uncovers temporal dependencies;

CNN-LSTM [46]: CNN extracts local spatial features, fuses spatiotemporal information, and uses bidirectional LSTM to uncover temporal dependencies;

TiRGN [47]: Time-induced Recurrent Graph Network, which models only graph relationships in the temporal dimension;

RETIA [48]: A temporal network that enhances feature representation for long sequences through a multi-scale retention mechanism;

All models use consistent dataset partitioning, signal preprocessing, sample lengths, and training hyperparameters to ensure a fair comparison; experiments are divided into two major scenarios: the CWRU bearing dataset (single component, multi-sensor, rotating machinery) and the PU gearbox dataset (multi-component strong coupling, variable-speed operating conditions, and composite failures). To further guarantee a fair comparison, all models share the same training/validation/test partitions, the same random seed for data splitting, the same sample length, and the same training hyperparameters, and every reported result is the mean of five independent runs.

The CWRU bearing dataset provides drive-end vibration signals from a motor-bearing test stand under 0–3 hp loads, with single-point faults of several diameters introduced

by electric discharge machining on the inner race, outer race and rolling element; signals are sampled at 12 kHz. The PU gearbox dataset records vibration from a motor-clutch-gearbox test rig under variable speed, covering healthy, single-component and composite gear/shaft/bearing faults. Raw signals are segmented into fixed-length windows, standardized, and split by condition into training, validation and test sets. For CWRU, the bearing component node and the three sensor nodes (drive-end, fan-end, and base) form a four-node graph; for PU, the motor–clutch–gearbox–load chain together with the associated accelerometer nodes forms a thirteen-node graph. The ground-truth fault-source labels for the Top-1/Top-2 evaluation are derived directly from the component-level fault annotations: for CWRU, the fault is always located on the driven-end bearing, so the bearing component node is the ground-truth root; for PU, single-component faults designate the corresponding component node as the root, and composite faults designate the two involved component nodes as the two ground-truth sources.

5.1. Experimental Results on the CWRU Dataset

The core challenges of the CWRU dataset include distinguishing between similar minor damage faults, resisting environmental noise, and fault localization under sensor correlation. Comparison results for classification metrics are shown in Table 1 and Figure 3.

Table 1. Comparison of Classification Metrics (CWRU Datasets).

Model	Accuracy	Precision	Recall	F1
CWT-CNN	95.1	94.8	94.5	94.6
ResNet	96.8	96.5	96.2	96.3
CNN-SVM	93.7	93.3	93.0	93.1
CNN-BiGRU	95.5	95.2	94.9	95.0
CNN-LSTM	94.9	94.6	94.3	94.4
TiRGN	96.2	95.9	95.6	95.7
RETIA	96.5	96.2	95.9	96.0
CGCA	98.6	98.3	98.1	98.2

In this paper, the CGCA model outperforms other models across all four metrics, achieving an accuracy of 98.6%. ResNet, RETIA, and TiRGN models achieve accuracies ranging from 96% to 97%, with deep networks and temporal graph networks demonstrating strong capabilities in identifying single-device failures; CNN-BiGRU, CWT-CNN, and CNN-LSTM achieve accuracy rates of 94–96%, with traditional spatio-temporal fusion and time-frequency analysis models demonstrating moderate performance; CNN-SVM achieves only 93.7% accuracy, as the traditional SVM classifier limits the overall performance ceiling.

CNN-SVM: With average feature extraction capabilities and compounded by the classification weaknesses of SVM, it lacks the ability to distinguish subtle damage features in bearings, ranking lowest across all four metrics; CWT-CNN, CNN-LSTM, and CNN-BiGRU: These models rely on signal time-frequency or pure time-series features and cannot utilize the correlation between multiple sensors. They are prone to misclassification when dealing with similar faults involving the inner ring, outer ring, and rolling elements, resulting in low recall rates. In contrast, deep neural networks and temporal networks such as ResNet, TiRGN, and RETIA can effectively learn fault features and achieve high overall accuracy; however, they focus solely on patterns within the data itself without incorporating the physical connection relationships between sensors and components. The proposed CGCA, by contrast, models the topological relationships among three sensor channels and bearing components using a graph structure. It employs GCN to aggregate multi-sensor coupled features and GAT to enhance the weights of fault-related nodes, thereby enabling the distinction of faults at different locations and with varying damage sizes while reducing false

negatives and misclassifications. The comparison results of the fault localization metrics are shown in Table 2 and Figure 4.

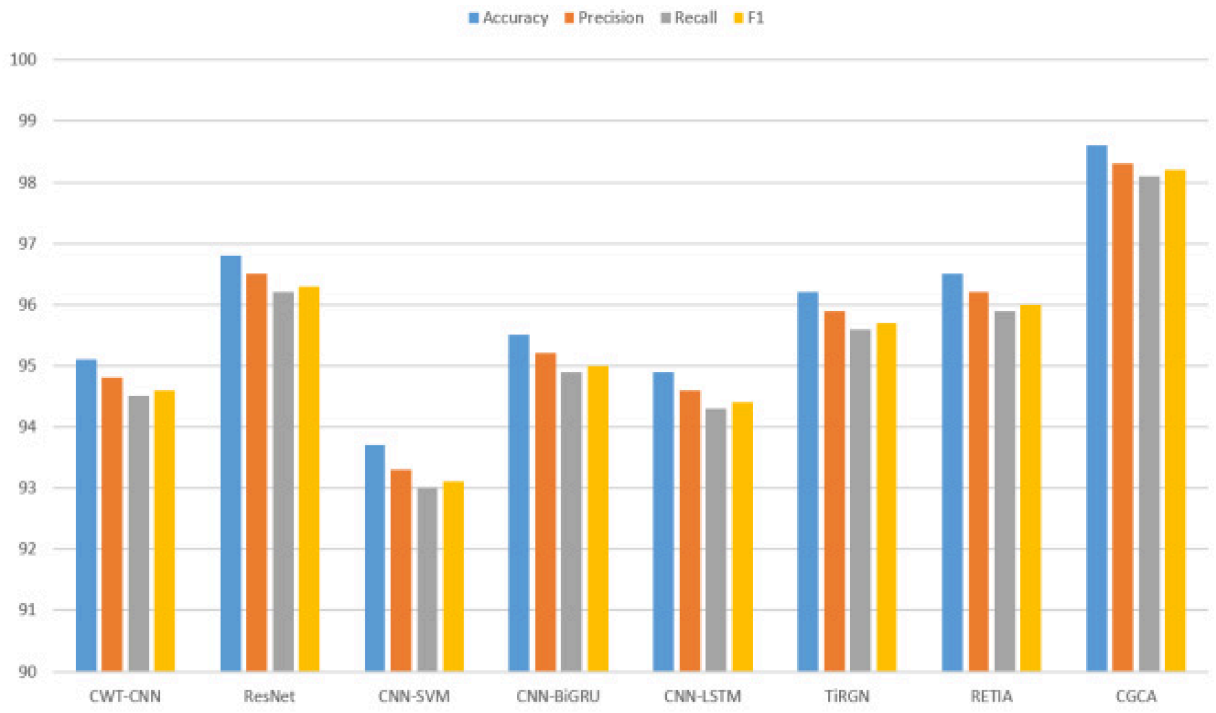

Figure 3. Comparison of Classification Metrics (CWRU Dataset).

Table 2. Comparison of Fault Localization Metrics (CWRU Datasets).

Model	Top-1	Top-2
CWT-CNN	78.2	85.5
ResNet	82.5	88.7
CNN-SVM	75.3	83.1
CNN-BiGRU	80.1	86.4
CNN-LSTM	79.5	85.8
TiRGN	88.3	94.1
RETIA	87.6	93.5
CGCA	96.9	99.1

For this dataset, the experimental setup fixes the fault source on the bearing component. The primary objective is to accurately pinpoint the root fault node within a four-node system comprising the bearing component node and its three sensor nodes. The fault is always injected on the driven-end bearing, so the ground-truth root node for the Top-1/Top-2 evaluation is the bearing component node, while the three sensor nodes serve as distractors.

Traditional architectures, such as CNN, CNN + RNN, and CWT-CNN, show weak performance in this localization task, typically yielding a Top-1 accuracy below 83%. A major limitation of these methods is their inability to differentiate between sensor faults and core component failures; they can detect that a fault exists but cannot isolate its origin. Additionally, the lack of quantification regarding node-to-node influence results in diffuse predictions. Among these, CNN-SVM yields the least satisfactory results, with a Top-1 accuracy of just 75.3%. While temporal graph networks (e.g., TiRGN) and temporal augmentation networks (e.g., RETIA) improve upon this by approaching 88% Top-1 accuracy through temporal correlation extraction, they still face limitations. Specifically, the absence of equipment physical topology in their design restricts their localization accuracy.

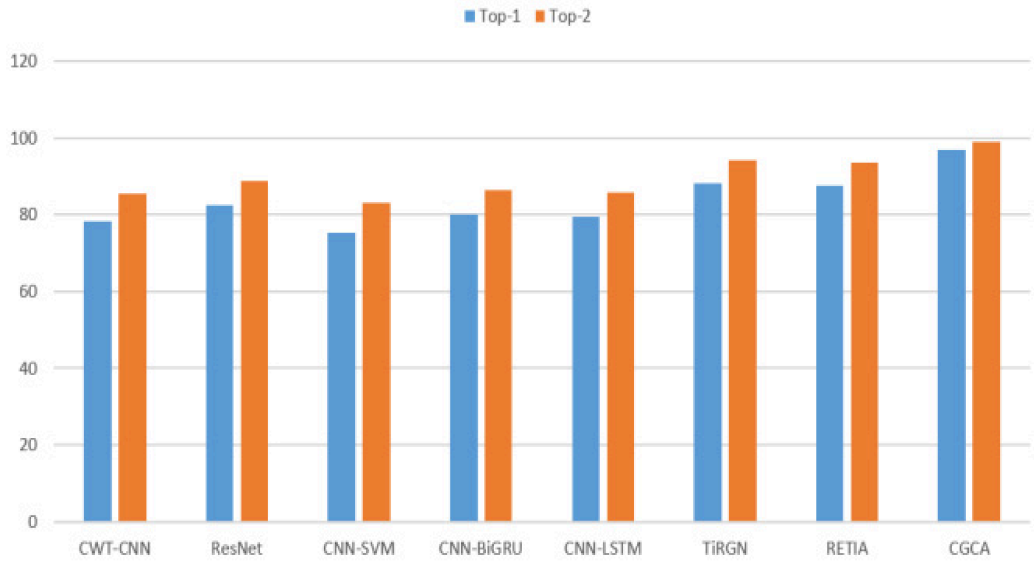

Figure 4. Comparison of Fault Localization Metrics (CWRU Dataset).

Addressing these challenges, our proposed CGCA achieves a superior Top-1 accuracy of 96.9% (Top-2: 99.1%). The model utilizes GAT attention coefficients to adaptively weigh the fault contribution of individual nodes. By assigning dominant weights to the faulty bearing node, the model ensures minimal localization bias. Crucially, CGCA demonstrates robustness against sensor noise, utilizing topological relationships to identify the true fault source with a precision that significantly exceeds existing comparison methods.

5.2. Experimental Results on the PU Dataset

The PU dataset represents a multi-component, strongly coupled system. The core challenges include transmission interference between components, drift in operating conditions during speed changes, decoupling of composite faults, and tracing multiple fault sources. The comparison of classification metrics is shown in Table 3 and Figure 5.

Table 3. Comparison of Classification Metrics (PU Dataset).

Model	Accuracy	Precision	Recall	F1
CWT-CNN	92.3	91.9	91.5	91.7
ResNet	94.1	93.7	93.3	93.5
CNN-SVM	90.5	90.1	89.7	89.9
CNN-BiGRU	93.0	92.6	92.2	92.4
CNN-LSTM	91.8	91.4	91.0	91.2
TiRGN	94.7	94.3	93.9	94.1
RETIA	94.5	94.1	93.7	93.9
CGCA	97.8	97.4	97.0	97.2

Reflecting the diagnostic difficulties posed by multi-component coupling and variable speeds, the overall accuracy for all models is lower than that of the CWRU dataset. In this context, CGCA demonstrates robust superiority with 97.8% accuracy, outperforming the second-best model (TiRGN) by 3.1%.

CNN-SVM: Yielding the lowest results (<91%), this approach is hindered by the SVM's inability to generalize effectively against the coupled noise generated by interfering gearbox signals, resulting in high rates of detection errors. CWT-CNN, CNN-LSTM, and CNN-BiGRU: Accuracy for these models ranges from 91% to 93%. They exhibit poor adaptability

to the feature drift in time-frequency and temporal domains caused by variable speeds, leading to a marked drop in sensitivity for faults like shaft bending and bearing wear. ResNet, TiRGN, and RETIA: These models perform well in isolating single-component failures (94–95%) but face a bottleneck with composite failures. Their inability to decouple the features of dual failure sources typically results in misclassifying composite faults as single failures. CGCA: This model utilizes a weighted adjacency matrix to map transmission coupling among gears, shafts, and bearings. By integrating multi-component correlations via GCN, it effectively mitigates feature drift arising from gear shifting. Furthermore, CGCA successfully identifies two distinct sets of fault features in composite scenarios, achieving the best overall results across all metrics (see Table 4 and Figure 6).

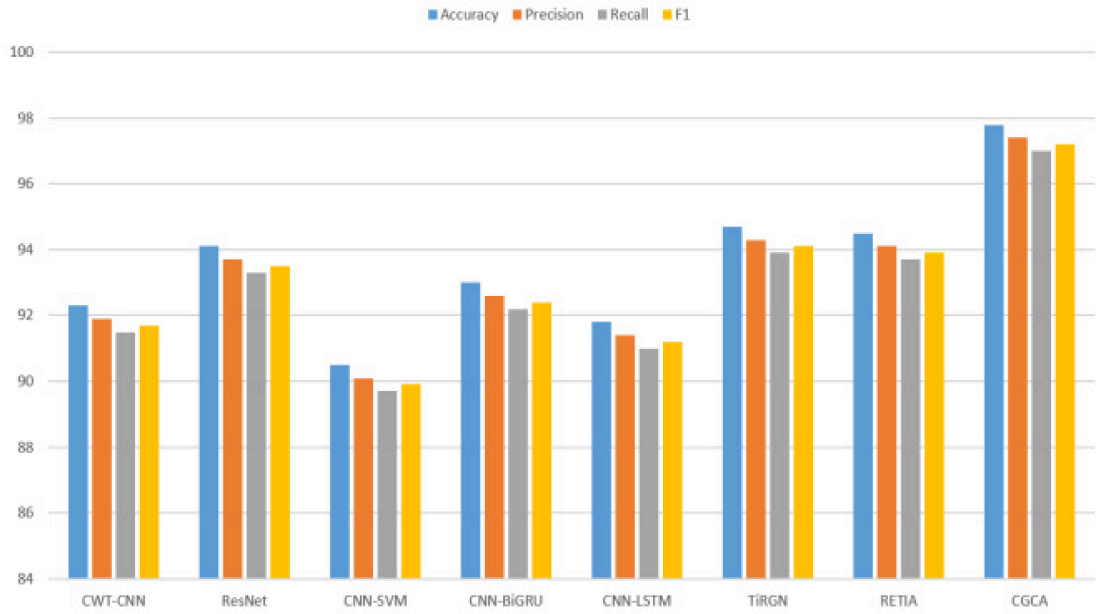

Figure 5. Comparison of Classification Metrics (PU Dataset).

Table 4. Comparison of Fault Attribution Metrics (PU Dataset).

Model	Top-1	Top-2
CWT-CNN	74.5	81.2
ResNet	79.3	86.6
CNN-SVM	71.8	79.5
CNN-BiGRU	77.6	84.3
CNN-LSTM	76.9	83.7
TiRGN	86.5	92.8
RETIA	85.9	92.1
CGCA	95.8	98.7

The PU dataset consists of 13 nodes with a chain-driven topology. It requires not only locating a single fault source but also identifying two concurrent fault sources, making fault localization significantly more challenging than in the CWRU dataset. For the PU dataset, the ground-truth fault-source nodes are obtained directly from the component-level fault labels: a single-component fault marks the corresponding component node (gear, shaft, or bearing) as the root node, whereas a composite fault marks the two involved component nodes as the two ground-truth sources. The accelerometer/sensor nodes are never ground-truth sources; they act as measurement points that propagate the fault signal.

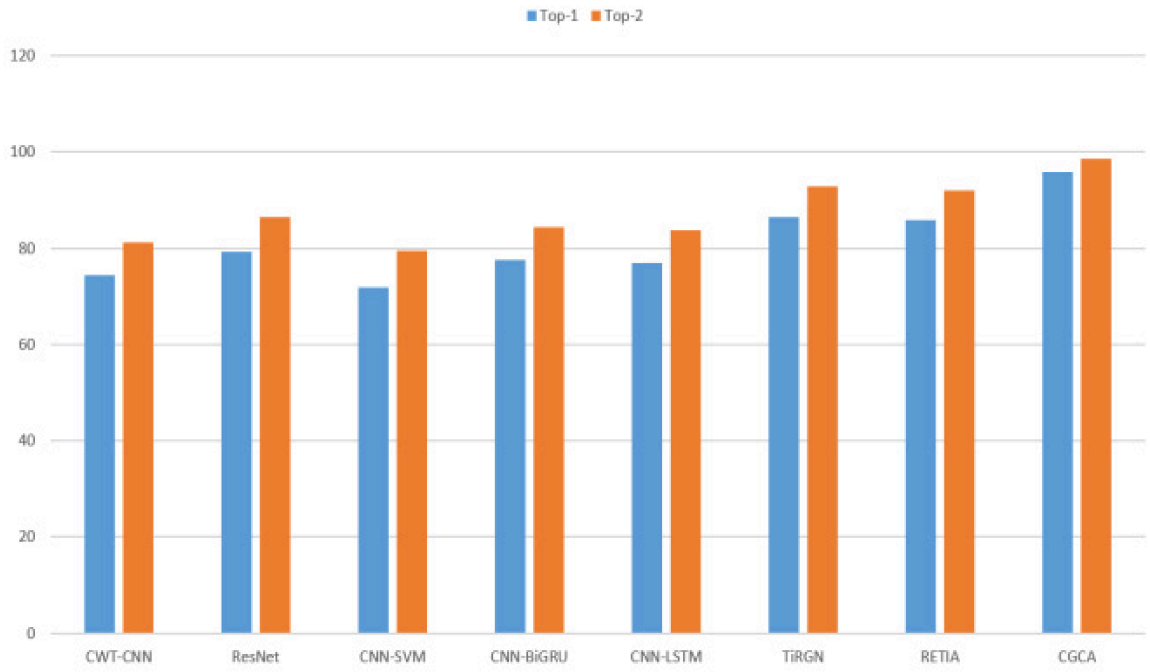

Figure 6. Comparison of fault tracing metrics for the (PU dataset).

Traditional models (CWT-CNN, CNN series, CNN-SVM): Performance in fault localization is limited, with Top-1 accuracy consistently falling below the 80% threshold. The inability to filter out interference between components hinders the separation of the actual fault source from secondary affected nodes, causing the localization results to diverge randomly. TiRGN and RETIA: While offering enhanced localization performance, these models struggle to handle composite faults effectively. Their tendency to isolate only a single fault node leads to substantial missed detections, restricting their Top-1 accuracy to a ceiling of around 86%. CGCA: This model demonstrates superior performance, reaching 95.8% in Top-1 and 98.7% in Top-2 accuracy. Its mechanism ensures that attention weights are tightly focused on the root node for single faults. In the case of dual-component composite faults, CGCA successfully highlights both true fault sources simultaneously. By differentiating between primary/secondary and source/propagated faults, it meets the rigorous fault tracing demands of multi-coupled systems.

The confusion matrices for these algorithms on the PU dataset are shown in Figures 7–14.

CWT-CNN and CNN-SVM: These models exhibit significant misclassification tendencies, particularly in gear-related faults (pitting and tooth breakage), where error rates range from 10% to 15%. A critical limitation is observed in composite fault samples, where more than 20% are misidentified as simple single faults, rendering the system unable to detect concurrent multiple faults. ResNet, TiRGN, and RETIA: Although effective for single-component failure recognition (accuracy >92%), these models face a bottleneck in composite scenarios. They typically detect only one failed component, resulting in a high false negative rate of 10–15% for composite failures. CNN-BiGRU and CNN-LSTM: Performance in these models is heavily impacted by rotational speed variations. The resulting poor feature stability causes recall rates for shaft deflection and bearing wear to drop to 85–88%. CGCA: This model demonstrates superior robustness, achieving recall rates above 98% across single-fault categories with negligible cross-category confusion. Crucially, it reduces the false negative rate for composite faults to under 3%, allowing for the precise identification of dual fault nodes. Additionally, the model utilizes attention weights to differentiate the primary fault source from secondary nodes affected by propagation, consistent with the physical fault propagation mechanisms in industrial settings.

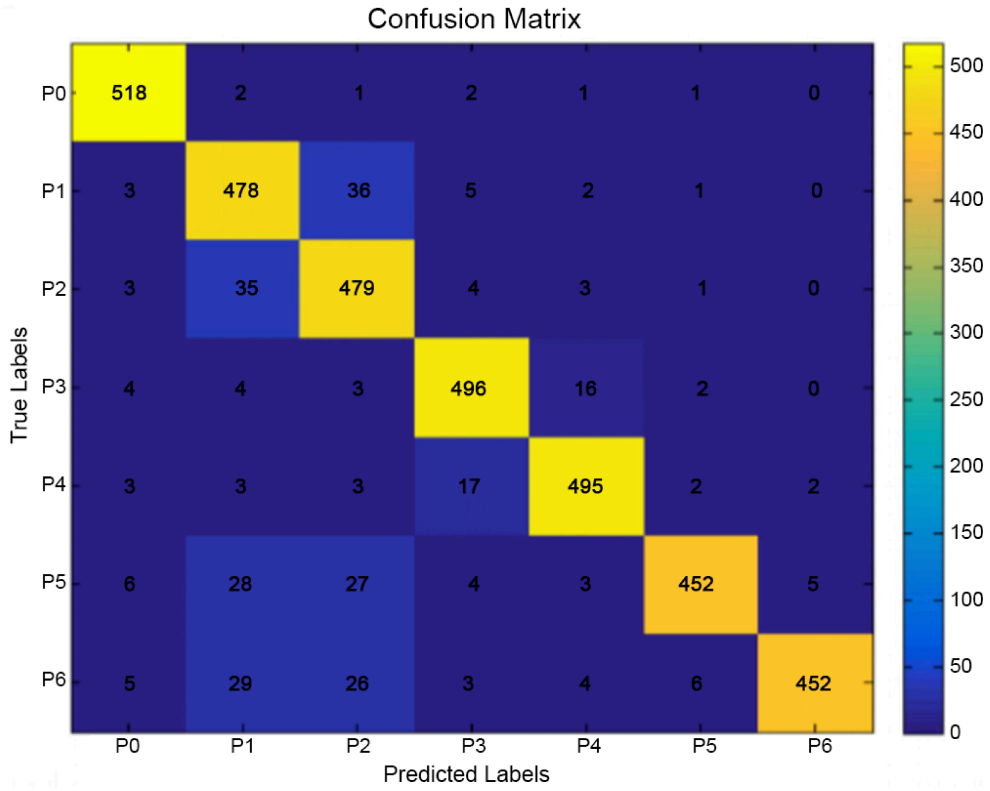

Figure 7. CWT-CNN.

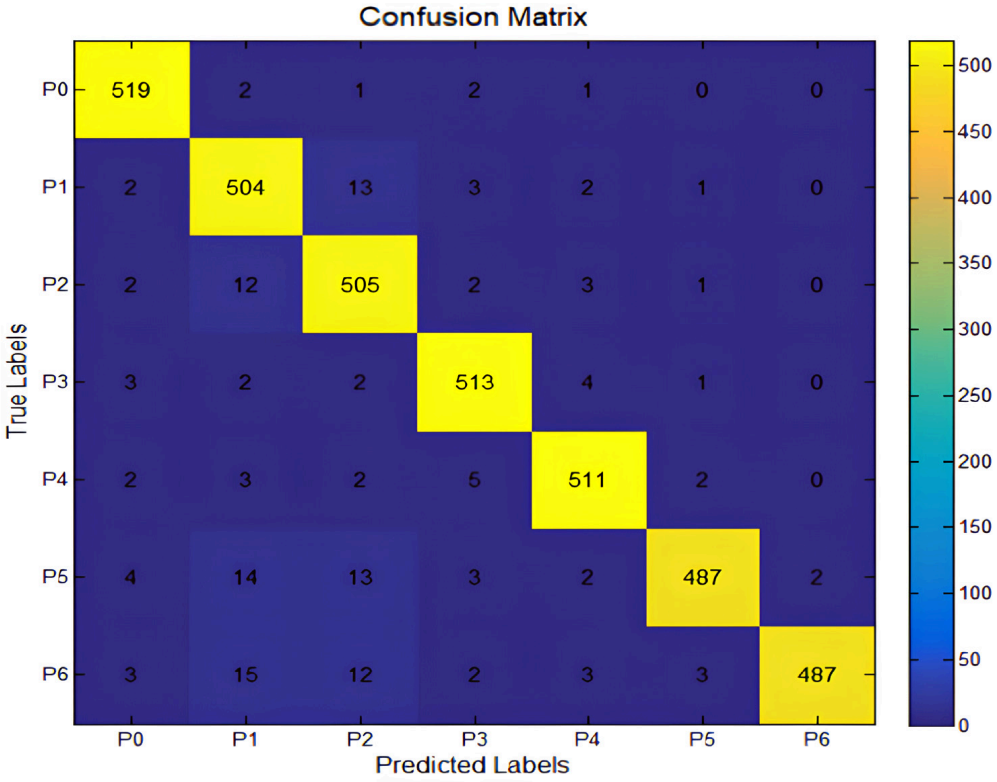

Figure 8. ResNet.

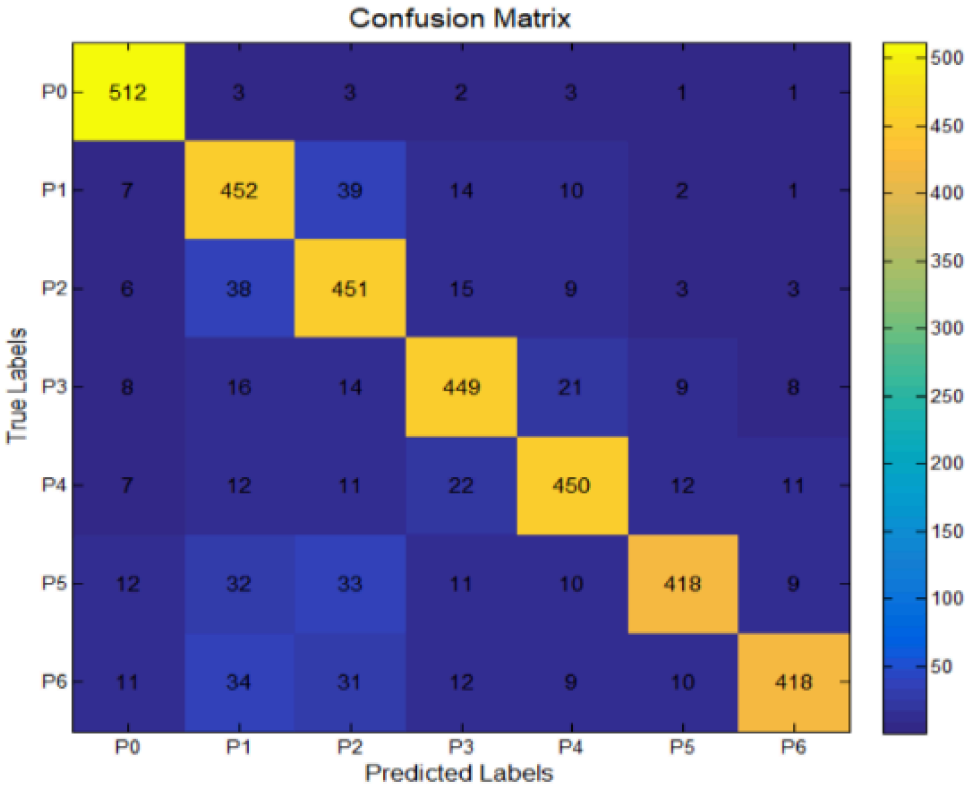

Figure 9. CNN-SVM.

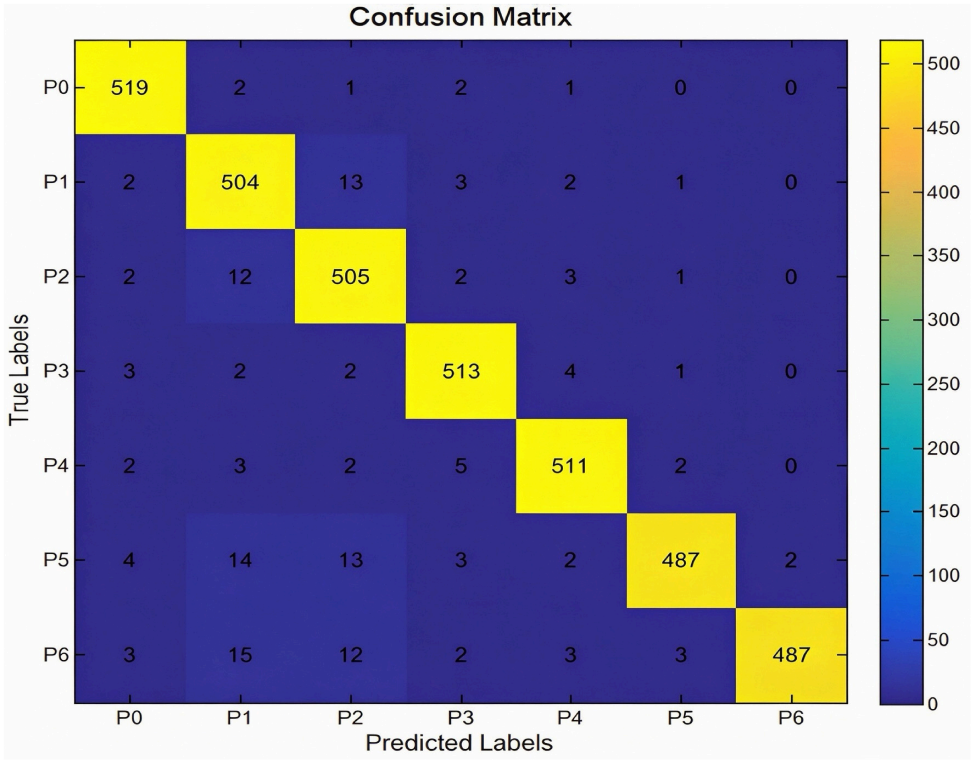

Figure 10. CNN-BiGRU.

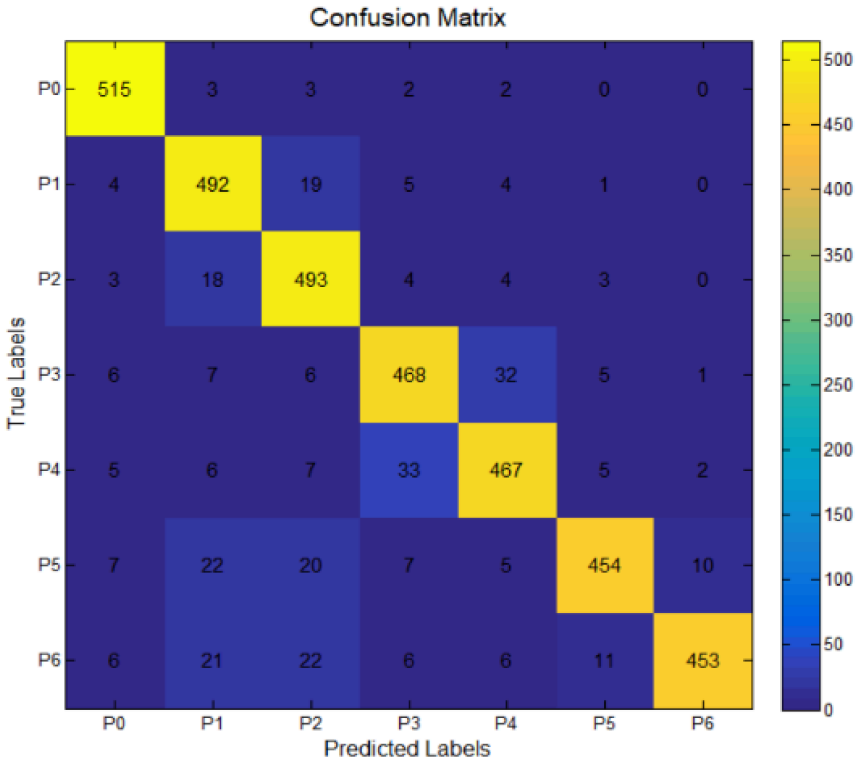

Figure 11. CNN-LSTM.

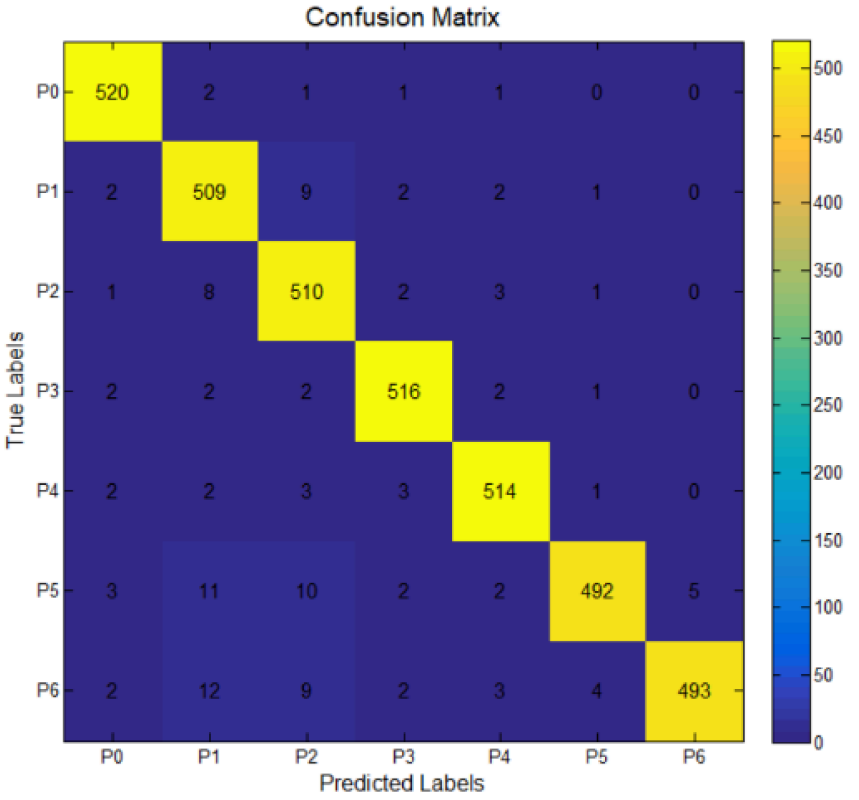

Figure 12. TiRGN.

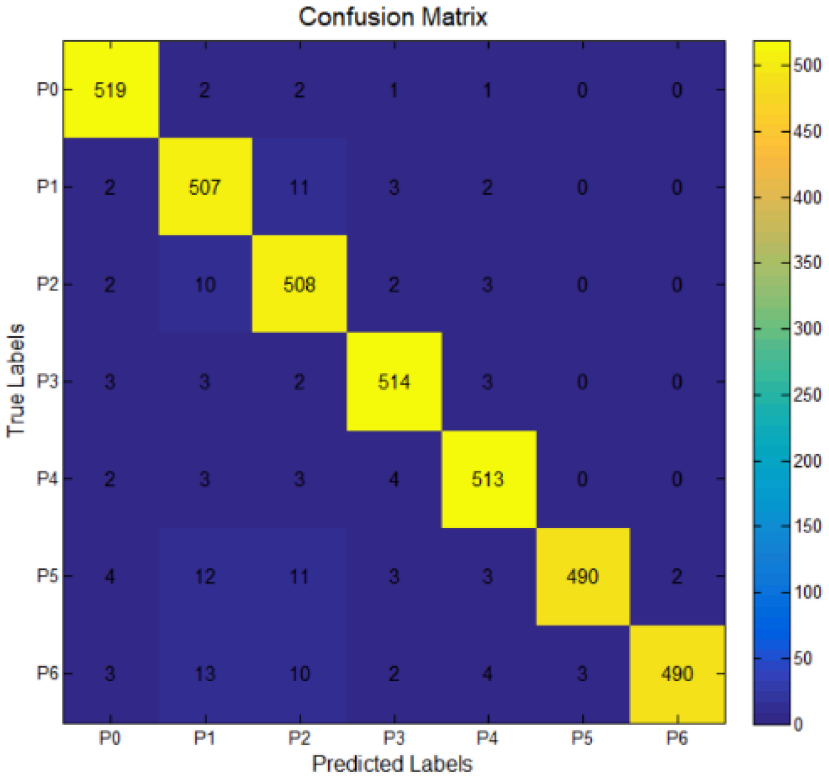

Figure 13. RETIA.

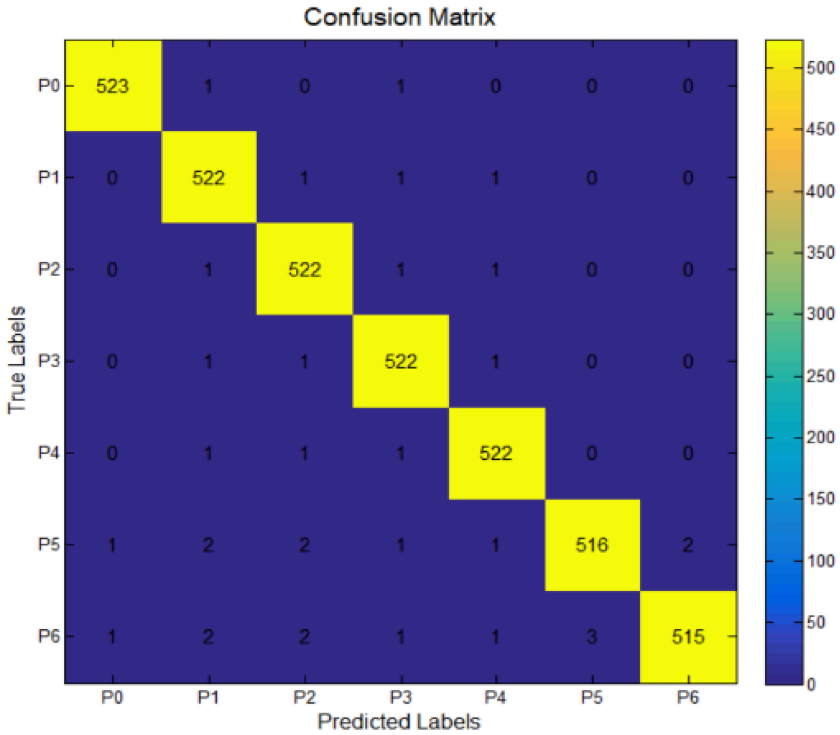

Figure 14. GGCA.

5.3. Experimental Results on Noise Robustness Comparison

Designed for complex industrial environments with strongly coupled multi-component and multi-sensor configurations (e.g., large-scale units, production lines, and distributed integrated equipment), this approach translates physical equipment topology into a graph-based framework. Here, sensors and components are modeled as nodes, and their physical

signal coupling relationships are represented as edges. Utilizing a hybrid CGCA model, the method effectively captures spatial correlations and signal coupling features across nodes, enabling accurate fault detection, classification, and root cause localization. Detailed results from the noise robustness comparison experiments on the CWRU dataset can be found in Table 5 and Figure 15.

Table 5. Comparison of Noise Robustness (CWRU dataset).

Model	10 dB	5 dB	0 dB
CWT-CNN	88.5	79.2	67.8
ResNet	90.3	82.1	71.5
CNN-SVM	85.7	76.3	65.1
CNN-BiGRU	89.1	80.5	69.3
CNN-LSTM	88.3	79.6	68.5
TiRGN	92.5	85.3	76.2
RETIA	91.8	84.6	75.5
CGCA	94.2	89.7	82.5

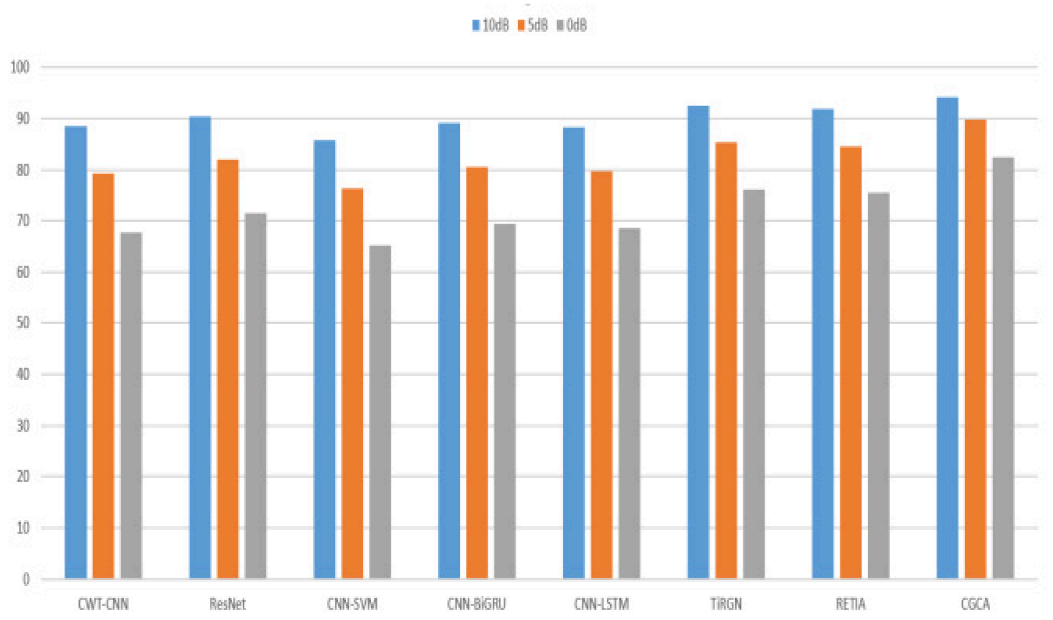

Figure 15. Comparison of Noise Robustness (CWRU Dataset).

The lower the SNR (Signal-to-Noise Ratio) value, the greater the intensity of the superimposed Gaussian noise. 10 dB represents light noise, 5 dB represents moderate noise, and 0 dB represents heavy noise, simulating complex operating conditions such as electromagnetic interference and mechanical vibration in industrial environments. The classification accuracy of all models continues to decline as noise increases, but the rate of decline varies significantly, directly reflecting the models' interference resistance.

CNN-SVM exhibits the lowest accuracy across the entire SNR range, reaching only 65.1% under heavy noise (0 dB); traditional machine learning models demonstrate extremely poor tolerance to noise; CWT-CNN, CNN-LSTM, and CNN-BiGRU experienced a precipitous drop in accuracy as noise increased, with accuracy falling below 70% under severe noise (0 dB). Models relying solely on signal features are prone to having valid fault information masked by noise; ResNet, TiRGN, and RETIA, as deep and temporal networks, possess a certain ability to filter features and can maintain an accuracy of over 75% even under severe noise. In the case of CGCA, leveraging graph topological constraints, the model

not only learns signal features but also the association patterns between devices and sensors. Noise only affects data at individual nodes and cannot disrupt the overall topological relationships. Even under 0 dB of heavy noise, the accuracy remains at 82.5%, outperforming other models and making it suitable for highly interfered industrial environments.

6. Conclusions and Future Work

This paper addresses the challenges of multiple concurrent faults, fault propagation, severe noise interference, and the lack of topological awareness in traditional models within the complex, coupled electromechanical systems of Industry 5.0. We propose a fault diagnosis and root cause analysis framework based on a Graph Convolutional Attention Fusion Network. This paper accomplishes three core tasks: First, it performs topological modeling of industrial equipment by mapping physical components and sensors to a weighted topological graph, quantifying the coupling strength between nodes, and achieving a deep integration of mechanistic knowledge with data-driven models; Second, we designed a CGCA dual-branch feature learning architecture, which leverages graph convolutions to extract global topological correlation features and employs multi-head graph attention to enhance the weights of local fault nodes, thereby achieving integrated fault identification and tracing; Third, we conducted multi-dimensional validation on public standard datasets to comprehensively evaluate the model's classification performance, fault tracing capability, noise robustness, and the effectiveness of each module.

This study still has certain limitations: First, when dealing with ultra-complex distributed systems with a very large number of nodes ($n > 100$), model inference latency increases, making it difficult to directly adapt the system to edge devices with limited computing power. Second, since the experiments in this study were conducted using offline steady-state datasets, the model's generalization capabilities under extreme variable operating conditions and dynamic topology switching scenarios require further validation. Third, edge weights derived from Pearson correlation reflect statistical co-variation rather than physical causality, so root-cause localization is an attention-assisted ranking rather than a rigorous causal analysis; the present work demonstrates classification and Top-1/Top-2 localization but does not yet provide a full fault-propagation mechanism. Fourth, all results are obtained on two laboratory benchmarks under Gaussian noise, without confidence intervals over many deployments; generalization to strongly nonlinear, dynamic-topology industrial sites remains to be verified.

Future research will focus on three directions: (1) lightweight modification of graph convolutional attention networks by designing lightweight graph modules to adapt to low-computational-power devices such as industrial edge gateways and embedded terminals; (2) introduction of a dynamic graph topology mechanism to construct a dynamic weighted adjacency matrix, thereby adapting to dynamic industrial scenarios such as changing device operating conditions and temporary component start-up or shutdown; and (3) integrating temporal graph networks to fuse spatiotemporal features, further uncovering the temporal evolution patterns of faults to enable fault trend prediction and remaining life assessment, thereby establishing an integrated intelligent operation and maintenance system encompassing "fault diagnosis, root cause analysis, and prediction."

Author Contributions: X.C. (Xiaojing Cao) and Z.S.: Conceptualization, methodology design, manuscript writing—original draft, supervision, funding acquisition. K.Y. and Y.L.: Algorithm implementation, experimental validation, data curation, manuscript revision. X.C. (Xuguang Chai): Investigation, formal analysis, visualization of experimental results, literature sorting. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Education of China Humanities and Social Science Foundation under Grant No. 25YJA910001, and in part by the Guangdong Province Philosophy and Social Science Planning Project under Grant No. GD26CYJ20.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

1. Xiao, Y.; Shao, H.; Liu, B. Evaluating calibration of deep fault diagnostic models under distribution shift. *Comput. Ind.* **2025**, *171*, 104334. [CrossRef]
2. Gazis, A.; Pappas, S.; Vavouras, T.; Ali, A.; Mastorakis, N.E. Sensing Deepfake Detection: A Survey of Detection Architectures, Adversarial Challenges, and Critical Applications in Political, Educational, and Military Domains. *ICCK Trans. Sens. Commun. Control* **2026**, *3*, 176–196. [CrossRef]
3. Zhang, J.; Zhao, Z.; Jiao, Y.; Zhao, R.; Hu, X.; Che, R. DPCCNN: A new lightweight fault diagnosis model for small samples and high noise problem. *Neurocomputing* **2025**, *626*, 129526. [CrossRef]
4. Xu, Y.; Yan, X.; Sun, B.; Liu, Z. Global contextual residual convolutional neural networks for motor fault diagnosis under variable-speed conditions. *Reliab. Eng. Syst. Saf.* **2022**, *225*, 108618. [CrossRef]
5. Ali, M.H.; Ali, F.; Muhammad, Z. FocusNet: Feature Oriented Contextual Understanding via Bidirectional Supervision for Surface Defect Detection. *ICCK Trans. Sens. Commun. Control* **2026**, *3*, 160–175. [CrossRef]
6. Zhang, T.; Jiang, L.; Liu, J.; Zhang, X.; Zhang, Q. A multi-scale deep feature memory and recovery network for multi-sensor fault diagnosis in the channel missing scenario. *Eng. Appl. Artif. Intell.* **2025**, *145*, 110228. [CrossRef]
7. Mohammadi, A.; Krysander, M.; Jung, D. Consistency-based diagnosis using data-driven residuals and limited training data. *Control Eng. Pract.* **2025**, *159*, 106283. [CrossRef]
8. Zheng, H.; Liu, H.; Liu, Z.; Tan, J. Federated temporal-context contrastive learning for fault diagnosis using multiple datasets with insufficient labels. *Adv. Eng. Inform.* **2024**, *60*, 102432. [CrossRef]
9. Xu, H.; Hu, Z.; Xu, Z.; Qian, Q. Zero-shot intelligent fault diagnosis via semantic fusion embedding. *Cogn. Robot.* **2025**, *5*, 37–47. [CrossRef]
10. Tie, J.; Xiao, P.; Zheng, L.; Ma, H.; Peng, D. Remote sensing scene classification based on attention network scale feature fusion. *J. South-Cent. Minzu Univ. (Nat. Sci. Ed.)* **2024**, *43*, 474–484.
11. Tian, L.; Chen, H.; Zheng, B. A spectral-spatial attention network combining multi-scale strategies for hyperspectral image classification. *J. South-Cent. Minzu Univ. (Nat. Sci. Ed.)* **2024**, *43*, 532–539.
12. Yang, C.; Liu, J.; Xu, Q.; Zhou, K. A generalized graph contrastive learning framework for few-shot machine fault diagnosis. *IEEE Trans. Ind. Inform.* **2023**, *20*, 2692–2701. [CrossRef]
13. Jiang, D.; He, C.; Chen, Z.; Zhao, J. Are novel deep learning methods effective for fault diagnosis? *IEEE Trans. Reliab.* **2024**, *74*, 4170–4184. [CrossRef]
14. Zhang, Z.; Zhou, F.; Wang, C.; Wen, C.; Hu, X.; Wang, T. A multiscale recursive attention gate federation method for multiple working conditions fault diagnosis. *Entropy* **2023**, *25*, 1165. [CrossRef] [PubMed]
15. Rameshkumar, K.; Sriram, R.; Saimurugan, M.; Krishnakumar, P. Establishing statistical correlation between sensor signature features and lubricant solid particle contamination in a spur gearbox. *IEEE Access* **2022**, *10*, 106230–106247. [CrossRef]
16. Sifat, M.S.I.; Kabir, M.A.; Islam, M.M.M.; Rehman, A.U.; Bermak, A. GAN-based data augmentation for fault diagnosis and prognosis of rolling bearings: A literature review. *IEEE Access* **2025**, *13*, 148083–148103. [CrossRef]
17. Lu, G.; Tsang, C.W.; Yim, H.N.; Lei, C.; Bu, S.; Yung, W.K.C.; Pecht, M. Interpretable fault diagnosis for overhead lines with covered conductors: A physics-informed deep learning approach. *Prot. Control Mod. Power Syst.* **2025**, *10*, 25–39. [CrossRef]
18. Li, S.; Xu, Y.; Feng, K.; Wang, Y.; Sun, B.; Yan, X.; Sheng, X.; Zhang, K.; Zheng, J.; Ni, Q. Joint threshold learning convolutional networks for intelligent fault diagnosis under nonstationary conditions. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–11. [CrossRef]
19. Lei, Z.; Deng, S.; Su, Y.; Li, Z.S.; Feng, K.; Wen, G.; Li, Z.; Chen, X. Adaptive weighted cost-sensitive learning-driven improved dense convolutional neural network for imbalanced fault diagnosis under limited fault samples. *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng.* **2025**, *11*, 04025013. [CrossRef]
20. Ma, X.; He, X.; Wu, X.; Wen, C. Multi-level federated learning based on cloud-edge-client collaboration and outlier-tolerance for fault diagnosis. *Meas. Sci. Technol.* **2023**, *34*, 125148. [CrossRef]
21. Abbas, H.; Chinnam, R.B. Advancing intelligent fault diagnosis through enhanced mechanisms in transfer learning. *Machines* **2025**, *13*, 1120. [CrossRef]

22. Iqbal, M.; Lee, C.K.M.; Keung, K.L.; Zhao, Z. Intelligent fault diagnosis across varying working conditions using triplex transfer LSTM for enhanced generalization. *Mathematics* **2024**, *12*, 3698. [CrossRef]
23. Yi, F.; Zhong, R.; Zhu, W.; Zhou, R.; Wang, Y.; Guo, L. Leakage fault diagnosis of wind tunnel valves using wavelet packet analysis and vision transformer-based deep learning. *Mathematics* **2025**, *13*, 3195. [CrossRef]
24. Li, X.; Li, S.; Huang, J.; Zhang, F.; Kong, X.; Xie, Z.; Chu, F. MFSFormer: A novel end-to-end rotating machinery fault diagnosis framework for noise and small samples. *IEEE Trans. Reliab.* **2024**, *74*, 4158–4169. [CrossRef]
25. Xu, Y.; Yan, X.; Sun, B.; Zhai, J.; Liu, Z. Multireceptive field denoising residual convolutional networks for fault diagnosis. *IEEE Trans. Ind. Electron.* **2021**, *69*, 11686–11696. [CrossRef]
26. Wang, Z.; Ji, J.C.; Xu, Y.; Li, S.; Sun, B.; Yang, X. Multiview contrastive shapelet learning: A novel semisupervised approach for explainable machine fault diagnosis with insufficient annotated data. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 1–11. [CrossRef]
27. Kannan, N.; Saimurugan, M.; Sowmya, S. Enhanced quadratic discriminant analysis with sensor signal fusion for speed-independent fault detection in rotating machines. *Meas. Sci. Technol.* **2023**, *34*, 125158. [CrossRef]
28. Jain, N.; Singh, V.K.; Singh, C.; Tripathi, G.; Rathore, R.S.; Chaudhari, D. Noise-resilient deep learning architecture for bearing fault diagnosis in integrated IoT sensing–computing environments. *IEEE Trans. Consum. Electron.* **2025**, *72*, 1685–1698. [CrossRef]
29. Praveen Kumar, T.; Buvaanesh, R.; Saimurugan, M.; Naresh, G.; Muthiya, S.J.; Basavanakattimath, M. Performance evaluation of deep learning approaches for fault diagnosis of rotational mechanical systems using vibration, sound, and acoustic emission signals. *J. Low. Freq. Noise Vib. Act. Control* **2024**, *43*, 1363–1380. [CrossRef]
30. Wang, T.; Wang, M.; Wang, B.; Ma, L. Deep transfer learning based fault diagnosis for electromagnetic pulse valve faults under small sample. *Chin. J. Mech. Eng.* **2025**, *38*, 182. [CrossRef]
31. Feng, J.; Bao, S.; Xu, X.; Zhang, Z.; Hou, P.; Steyskal, F.; Dustdar, S. Rotating machinery fault diagnosis based on feature extraction via an unsupervised graph neural network. *Appl. Intell.* **2023**, *53*, 21211–21226. [CrossRef]
32. Yu, Z.; Wang, Y.; Zong, X.; Wu, J.; Zhou, Q. GRUDMU-DSCNN: An edge computing method for fault diagnosis with missing data. *Appl. Intell.* **2025**, *55*, 140. [CrossRef]
33. Kai, T.C.Y.; Saptoro, A.; Putra, Z.A.; Lim, K.H.; Yeo, W.S.; Sunarso, J. Supervised deep learning algorithms for process fault detection and diagnosis under different temporal subsequence length of process data. *Appl. Intell.* **2025**, *55*, 883. [CrossRef]
34. Li, X.; Zhang, J.; Zhao, C.; Ding, J.; Sun, Y. Nonlinear industrial process fault diagnosis with latent label consistency and sparse Gaussian feature learning. *J. Cent. South Univ.* **2022**, *29*, 3956–3973. [CrossRef]
35. Zhang, W.; Peng, G.; Li, C.; Chen, Y.; Zhang, Z. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors* **2017**, *17*, 425. [CrossRef] [PubMed]
36. Xu, Y.; Zhang, L.; Chen, L.; Tan, T.; Wang, X.; Xiao, H. Attention-guided residual spatiotemporal network with label regularization for fault diagnosis with small samples. *Sensors* **2025**, *25*, 4772. [CrossRef] [PubMed]
37. Lu, M.; Wang, G. Temporal heterogeneous relationships enhanced deep clustering network for fault diagnosis using unlabeled multisensor data. *IEEE Trans. Ind. Inform.* **2024**, *21*, 912–921. [CrossRef]
38. Cao, Y.; Tang, J.; Shui, S.; Cai, D.; Zhang, L.; Xiong, P. Fault diagnosis techniques for electrical distribution network based on artificial intelligence and signal processing: A review. *Processes* **2025**, *13*, 48. [CrossRef]
39. Godhade, S.; Kumar, J. Robust phasor measurement unit (PMU)-based fault diagnosis in transmission lines using a spatio-temporal hybrid transformer with graph contrastive learning. *Eng. Sci. Technol. Int. J.* **2026**, *80*, 102440. [CrossRef]
40. Tsiakalos, A.; Tsiakalos, A. Wind-Aware Angle-of-Arrival Consistency Validation for Spoofed Guidance Signal Rejection in Unmanned Aerial Vehicles. *ICCK Trans. Intell. Syst.* **2026**, *3*, 186–195. [CrossRef]
41. Godhade, S.; Kumar, J. Robust ensemble framework for fault detection in transmission lines using hybrid classifiers and deep Q-networks. *SSRG Int. J. Electron. Commun. Eng.* **2026**, *13*, 257–274. [CrossRef]
42. Sadeque, M.G.; Ahmmed, T.; Hossain, M.A.; Rahman, M.N.; Sarker, M.T.; Ramasamy, G. A novel CWT-CNN framework for driver fatigue detection from EEG signals. In Proceedings of the 2025 Multimedia University Engineering Conference (MECON), Cyberjaya, Malaysia, 21–23 July 2025.
43. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
44. Agarap, A.F. An architecture combining convolutional neural network (CNN) and support vector machine (SVM) for image classification. *arXiv* **2019**, arXiv:1712.03541. [CrossRef]
45. Zhang, M.; Kong, W.; Teng, J.; Wang, Z. A CNN-BiGRU text classification model merging positional encoding and attention mechanism. In Proceedings of the WCSE 2022 Spring Event: 2022 9th International Conference on Industrial Engineering and Applications, Sanya, China, 18 April 2022; pp. 1–10. [CrossRef]
46. Ali, M.; Ahmmed, T.; Emran, M.A.; Roy, D.; Refat, K.A.; Khan, R. Driver fatigue detection using CWT-extracted features and a deep learning approach (CNNLSTM). In Proceedings of the 2024 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON), Dhaka, Bangladesh, 28–29 November 2024.

47. Li, Y.; Sun, S.; Zhao, J. TiRGN: Time-guided recurrent graph network with local-global historical patterns for temporal knowledge graph reasoning. In Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22), Main Track, Vienna, Austria, 23–29 July 2022; pp. 2152–2158.
48. Liu, K.; Zhao, F.; Xu, G.; Wang, X.; Jin, H. RETIA: Relation-entity twin-interact aggregation for temporal knowledge graph extrapolation. In Proceedings of the 2023 IEEE 39th International Conference on Data Engineering (ICDE), Anaheim, CA, USA, 3–7 April 2023; pp. 1761–1774. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.