

Article

Dynamics-Driven Dual-Stream Graph Neural Network with Adaptive Gated Fusion for Gearbox Fault Diagnosis

Jiashuo Yu ^{1,2}, Hanbin Xiao ^{1,2,*}, Min Liu ¹ and Dinglong Zhu ³

¹ School of Transportation and Logistics Engineering, Wuhan University of Technology, Wuhan 430063, China; 812581080@whut.edu.cn (J.Y.); liumin1989@whut.edu.cn (M.L.)

² Sanya Science Education Innovation Park, Wuhan University of Technology, Sanya 572000, China

³ Sinopec Petroleum Engineering Jiangnan Co., Ltd. (SPEJH), Wuhan 430223, China; zhudl.osec@sinopec.com

* Correspondence: xhbchina@126.com

Abstract

Conventional data-driven networks for gearbox fault diagnosis process multi-sensor streams as isolated sequences, failing to capture spatial-topological kinetic correlations and structural energy propagation pathways governed by multi-stage gearbox dynamics. To address these limitations, this study proposes a graph neural network-based fault diagnosis methodology integrating multi-dimensional attention and dynamic topological priors (KT-GNN-CBAM). Mesh stiffness characteristics are analytically evaluated to initialize physical topology edge weights, while a convolutional block attention module filters spatio-temporal features to suppress background noise. Node features are subsequently aggregated through a parallel dual-stream architecture comprising a physics-prior kinetic stream and a data-driven attention stream, which are dynamically fused via an adaptive gated mechanism. Experimental validation on the HP-GBS-2023 testbed under mixed operations and 6 dB noise shows that the proposed framework achieves an optimal diagnostic accuracy of 98.87%. Ablation evaluations confirm that omitting the mechanics-driven prior branch induces a 282.30% relative surge in the model's misclassification rate. Ultimately, embedding mechanical invariants as a physical inductive bias mitigates purely data-driven black-box constraints, offering an interpretable and robust solution for advanced intelligent fault diagnosis in complex gearbox systems.

Keywords: gearbox fault diagnosis; graph neural networks; physics-guided neural networks; adaptive gated fusion

1. Introduction

Heavy hoisting machinery, such as port cranes, constitutes critical infrastructure in modern logistics. Within these systems, the gearbox is a pivotal element of the power transmission chain operating under severe dynamic conditions. Continuous exposure to complex alternating stresses renders critical gears highly susceptible to structural degradation and fatigue failure, which can trigger catastrophic accidents. Consequently, early fault detection and prognostics and health management (PHM) are imperative to guarantee equipment reliability.

To address non-stationary vibration fluctuations under fluctuating operations, adaptive time-frequency techniques have been widely deployed. Methodologies including empirical mode decomposition (EMD) variants [1–4], intrinsic/local characteristic scale decompositions (ITD/LCD) [5–7], wavelet and empirical wavelet transforms (WT/EWT) [8–13], variational mode decomposition (VMD) [14–16], singular



Academic Editor: Ahmed Abu-Siada

Received: 16 July 2026

Revised: 1 August 2026

Accepted: 3 August 2026

Published: 5 August 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

value decomposition (SVD) [17,18], and synchrosqueezing/synchroextracting transforms (SST/SET) [19–26] have significantly advanced weak fault feature enhancement. Concurrently, classical supervised learning and statistical discriminant analysis frameworks have successfully mapped vibrational and kinematic signatures to mechanical health states, as demonstrated by Niola et al. [27] for micro-turbine vibrational behavior classification and Marticorena and García Peyrano [28] in bearing cage kinematics and acoustic emission monitoring. Nevertheless, these classical methodologies fundamentally process isolated signal sequences, relying on empirical parameters while failing to capture spatial-topological kinetic correlations across distributed sensors.

Driven by industrial big data, deep learning (DL) architectures have shifted fault diagnosis toward autonomous representation learning [29]. Generative networks (DBNs, autoencoders) established early paradigms for feature stacking [30–32]. To process complex dynamic sequences, convolutional neural networks (CNNs) and recurrent backbones (e.g., LSTMs) were introduced to capture spatiotemporal dependencies in multi-channel vibration streams [33–41]. Furthermore, attention mechanisms, time-series Transformers, and advanced statistical fusion models have enhanced representation capacity under noisy or small-sample constraints [42–45]. Despite these successes, existing deep networks encounter severe bottlenecks in industrial scenarios: (1) Absence of dynamic structural priors: prevailing architectures treat multi-sensor streams as isolated sequences, ignoring spatial sensor placement and deterministic energy propagation pathways. Although recent physics-guided GNNs attempt to embed domain knowledge, most rely on static topologies or soft penalty loss terms, lacking dynamic adaptability under cross-domain operational shifts. (2) Insufficient noise robustness: standard uniform weight-sharing fails to adaptively isolate transient impacts from heavy background noise.

To address these limitations, this study proposes a graph neural network-based fault diagnosis methodology integrating multi-dimensional attention and dynamic topological priors (KT-GNN-CBAM). The core contributions are as follows:

(1) Physical Topology Graph Construction: The equivalent mesh stiffness characteristics of the gearbox components are analytically evaluated to initialize the edge weight matrix of the GNN, constraining the network along physical energy propagation pathways.

(2) Feature Recalibration Mechanism: A Convolutional Block Attention Module (CBAM) is embedded within the feature extraction layers to filter channel and temporal features, suppressing background noise.

(3) Dual-Stream Information Fusion: A dual-stream graph neural network architecture is developed to aggregate node features through parallel physical constraints and data-driven attention, using a gated mechanism for feature fusion.

Consequently, the framework utilizes topological constraints as a physical inductive bias and attention mechanisms for adaptive selectivity, improving diagnostic accuracy and robustness under variable operating conditions.

The remainder of this paper is organized as follows: Section 2 presents the mathematical formulations and architectural design of the proposed KT-GNN-CBAM framework. Section 3 details the physical experimental setup, fault-injection scheme, sensor placements, and dataset partitioning strategy. Section 4 evaluates the comparative diagnostic performance, multi-metric analyses, feature manifold visualizations, training convergence profiles, and structural ablation studies. Finally, Section 5 concludes the paper, highlights core findings, and outlines potential future research directions.

2. Methods

This section describes the mathematical foundations of the component modules within the proposed KT-GNN-CBAM framework.

2.1. Foundations of Relevant Theories and Models

2.1.1. Graph Topological Formulations

To parameterize the spatial interactions within the multi-sensor gearbox network, the localized measurement configuration is modeled as a weighted undirected graph $G = (V, E, A)$. The vertex set $V = \{v_1, v_2, \dots, v_n\}$ contains the n sensor nodes, and the edge set $E \subseteq V \times V$ captures the structural transmission pathways. The graph signal is represented as an attribute matrix $X \in \mathbb{R}^{n \times F}$ with feature dimension F . The symmetric adjacency matrix $A \in \mathbb{R}^{n \times n}$ quantifies the connectivity strength via coupling weights w_{ij} ($A_{ij} = w_{ij}$ if $((v_i, v_j) \in E)$, else 0). To ensure numerical stability across multi-layer propagation, the symmetric normalized Laplacian matrix L_{sym} is utilized:

$$L_{sym} = I_n - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \quad (1)$$

where I_n denotes the $n \times n$ identity matrix, and D is the diagonal degree matrix defined by $D_{ii} = \sum_{j=1}^n A_{ij}$.

2.1.2. Spatial Graph Convolution and Attention Operations

The physical stream branch utilizes a Graph Convolutional Network (GCN). Based on first-order spectral graph approximation, the layer-wise feature propagation rule for the l -th GCN layer is formulated as:

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}) \quad (2)$$

where $H^{(0)} = X$, $\tilde{A} = A + I_n$, $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$, and $W^{(l)}$ represents the trainable layer weight matrix. The non-linear activation function $\sigma(\cdot)$ denotes the Rectified Linear Unit (ReLU). The baseline adjacency matrix A is structurally initialized by the dynamic stiffness matrix $A_{kinetic}$.

Concurrently, the data-driven Graph Attention Network (GAT) stream calculates dynamic correlation coefficients e_{ij} between connected nodes via a parameterized self-attention vector $\mathbf{a}$ and a shared weight matrix W :

$$e_{ij} = \text{LeakyReLU}(\mathbf{a}^T [\mathbf{W} \mathbf{h}_i \parallel \mathbf{W} \mathbf{h}_j]) \quad (3)$$

where $\mathbf{h}_i$ and $\mathbf{h}_j$ are the respective node feature vectors, and $\parallel$ denotes vector concatenation. The attention coefficient is normalized via a Softmax function over the first-order neighborhood N_i to yield the attention weight α_{ij} . Extending to a multi-head attention mechanism across K independent sub-spaces, the output layer computes the averaged node representation:

$$\mathbf{h}'_i = \sigma\left(\frac{1}{K} \sum_{k=1}^K \sum_{j \in N_i} \alpha_{ij}^{(k)} W^{(k)} \mathbf{h}_j\right) \quad (4)$$

where $\alpha_{ij}^{(k)}$ and $W^{(k)}$ correspond to the components of the k -th attention head.

2.1.3. Multidimensional Attentional Feature Recalibration and CNN Backbone

To extract local patterns from raw time-series sequences, a standard multi-layer 1D-CNN structured with cascading 1D convolutional layers, batch normalization (BN), ReLU activation, and max pooling operations is deployed as the front-end backbone to generate the initial node feature matrix X .

To isolate fault signatures from background interference, the Convolutional Block Attention Module (CBAM) processes the intermediate convolutional feature map $\mathbf{F} \in \mathbb{R}^{C \times L}$. The Channel Attention Module (CAM) generates a 1D channel attention map $\mathbf{M}_C(\mathbf{F})$ by pooling temporal context via average pooling $\mathbf{F}_{avg}^C$ and max pooling $\mathbf{F}_{max}^C$ through a shared Multi-Layer Perceptron (MLP):

$$\mathbf{M}_C(\mathbf{F}) = \sigma(\mathbf{W}_1(\mathbf{W}_0(\mathbf{F}_{avg}^C)) + \mathbf{W}_1(\mathbf{W}_0(\mathbf{F}_{max}^C))) \quad (5)$$

where $\mathbf{W}_0$ and $\mathbf{W}_1$ are the MLP weights, and $r = 16$ is the reduction ratio. The channel-refined feature map $\mathbf{F}'$ is derived via channel-wise multiplication $\mathbf{F}' = \mathbf{M}_C(\mathbf{F}) \otimes \mathbf{F}$. Subsequently, the Spatial Attention Module (SAM) tracks transient impacts along the temporal axis to output the spatial attention map $\mathbf{M}_S(\mathbf{F}') = \sigma(f^{7 \times 1}([\mathbf{F}_{avg}^S; \mathbf{F}_{max}^S]))$, where $f^{7 \times 1}$ signifies a 1D convolution with a kernel size of 7, and $[\cdot; \cdot]$ denotes channel-wise concatenation. The final attention-enhanced feature representation is derived via element-wise multiplication: $\mathbf{F}'' = \mathbf{M}_S(\mathbf{F}') \otimes \mathbf{F}'$, establishing the high-SNR feature inputs for downstream graph learning.

2.2. Construction of Kinetic Topological Graphs and Physical Parameter Mapping

To integrate mechanical invariants into the graph architecture, the physical gearbox is mapped into a topological graph structure compatible with non-Euclidean operators.

2.2.1. Graph Definition of Sensor Nodes

Under the framework $G = (V, E, \mathbf{A})$, the node set V parameterizes the physical placement of multi-sensor measurement points. For the target two-stage gearbox transmission system, the node set is defined as:

$$V = \{v_{in}, v_{mid}, v_{out}\} \quad (6)$$

where v_{in} , v_{mid} and v_{out} correspond to the accelerometers mounted on the housing cover at the input shaft, intermediate shaft, and output shaft bearing locations, respectively. For each node $v_i \in V$, the associated feature vector $\mathbf{h}_i$ represents the localized time-domain sequence output by the front-end 1D-CNN-CBAM architecture.

2.2.2. Edge Definition of Physical Connection Relationships

The edge set E represents the energy transmission pathways within the mechanical powertrain, partitioned into a strong coupling edge set E_{strong} and a weak coupling edge set E_{weak} , $E = E_{strong} \cup E_{weak}$.

The strong coupling edge set E_{strong} covers sensor nodes with direct gear meshing interfaces (the input-intermediate and intermediate-output gear pairs). Conversely, the weak coupling edge set E_{weak} links nodes that lack direct gear-meshing contact but interact indirectly through vibrational cross-talk across the monolithic housing cover.

To provide a quantitative physical basis for the weak coupling edge weights rather than a qualitative assumption, the casing structural attenuation coefficient η_{shell} was evaluated through transient dynamics simulation of the gearbox housing assembly. By comparing the dynamic acceleration responses at the non-meshing housing measurement points with those at the direct meshing interfaces under simulated operating conditions, the energy

transmission loss across the housing structural path was quantified. The edge weights A_{ij} in the kinetic adjacency matrix are defined as:

$$A_{ij} = \begin{cases} w_{mesh}^{(k)}, & e_{ij} \in E_{strong} \\ w_{struct}, & e_{ij} \in E_{weak} \\ 0, & e_{ij} \notin E \end{cases} \quad (7)$$

where $w_{mesh}^{(k)}$ represents the dynamic gear-mesh transmission weight of the k -th stage, and w_{struct} denotes the casing structural attenuation weight. To explicitly parameterize w_{struct} based on the simulation-derived attenuation factor η_{shell} , w_{struct} is formulated as:

$$w_{struct} = \eta_{shell} \cdot w_{mesh}^{(k)} \quad (8)$$

Based on the transient dynamics simulation results, the energy transmission coefficient across the structural housing cover was quantified at $\eta_{shell} \approx 0.20$ (yielding $w_{struct} = 0.20 \cdot w_{mesh}^{(k)}$). This formulation ensures that the weak-coupling edge weights w_{struct} strictly reflect the dynamic energy dissipation along non-meshing transmission pathways.

2.2.3. Model Parameter Mapping and Weight Initialization

To enforce physical interpretability during spatial feature aggregation, a mechanism-driven initialization strategy maps analytically calculated mesh stiffness and structural support stiffness parameters onto the kinetic adjacency matrix $\mathbf{A}_{kinetic}$. To prevent numerical anomalies or gradient explosion induced by massive absolute stiffness values, a continuous mapping function $\psi(\cdot)$ projects the physical parameters into a bounded graph weight space:

$$A_{ij} = \psi(K_{ij}) = (1 - \varepsilon) \cdot \frac{K_{ij} - K_{min}}{K_{max} - K_{min}} + \varepsilon \quad (9)$$

where K_{ij} represents the physical equivalent stiffness between component i and component j , while $K_{max} = 5.0 \times 10^8$ N/m (gear mesh stiffness K_{mesh}) and $K_{min} = 1.0 \times 10^8$ N/m (bearing support stiffness K_{bear}) denote the stiffness bounds across the transmission domain. The connectivity preservation factor ε is set to 0.1 to retain structural information flow along weak lines and prevent isolated islands during early optimization stages.

The instantiated kinetic adjacency matrix $\mathbf{A}_{kinetic}$ exhibits a diagonally dominant hierarchy based on the testbed configuration:

$$\mathbf{A}_{kinetic} = \begin{bmatrix} 1.0 & w_{mesh}^{(1)} & w_{struct} \\ w_{mesh}^{(1)} & 1.0 & w_{mesh}^{(2)} \\ w_{struct} & w_{mesh}^{(2)} & 1.0 \end{bmatrix} \approx \begin{bmatrix} 1.0 & 0.95 & 0.15 \\ 0.95 & 1.0 & 0.92 \\ 0.15 & 0.92 & 1.0 \end{bmatrix} \quad (10)$$

The diagonal elements define the structural self-loops for local feature retention. The sub-diagonal components quantify the direct gear-pairing transmission efficiency, and the remote off-diagonal elements govern the cross-talk boundaries. This matrix acts as a mechanism-driven prior, regularizing the gradient update vectors during error backpropagation to constrain the topology within kinematic boundaries.

2.3. Architecture of the KT-GNN-CBAM Framework

The proposed KT-GNN-CBAM framework integrates a sequential temporal feature extractor, a parallel dual-stream topological learning network, and an adaptive gated fusion classifier (Figure 1). The architecture utilizes structural mechanics boundaries alongside data-adaptive correlations to execute fault diagnosis on multi-stage gearboxes.

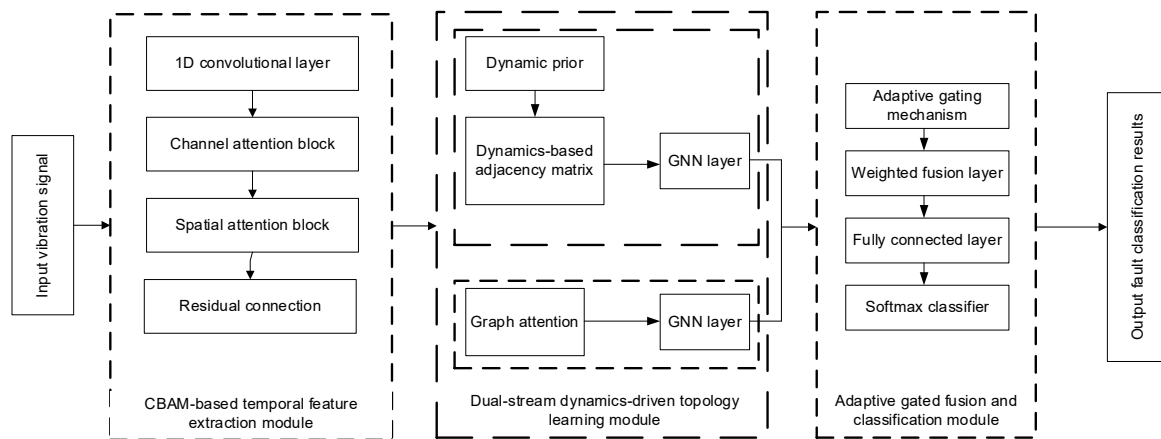


Figure 1. Schematic architecture of the proposed KT-GNN-CBAM.

2.3.1. Temporal Feature Extraction Module

The raw vibration sequence $\mathbf{X}$ is processed by a 1D-CNN backbone to generate high-dimensional feature tensors across C channels. To eliminate feature redundancy, the initial encoded tensor $\mathbf{F}_{enc}$ is scaled by the channel attention vector $\mathbf{M}_C$ via element-wise multiplication: $\mathbf{F}' = \mathbf{F}_{enc} \otimes \mathbf{M}_C$. The channel-modulated tensor $\mathbf{F}'$ is subsequently processed by the spatial attention map $\mathbf{M}_S$ to isolate transient shock impulses along the temporal axis: $\mathbf{F}'' = \mathbf{F}' \otimes \mathbf{M}_S$.

To stabilize gradient propagation and safeguard raw informational integrity, a residual connection superimposes the enhanced spatiotemporal tensor with the linearly projected raw inputs:

$$\mathbf{Y}_{feature} = \mathbf{F}'' + \text{Proj}(\mathbf{X}) \quad (11)$$

where $\text{Proj}(\cdot)$ is a linear projection operator deployed to align channel and temporal dimensional constraints. The output matrix $\mathbf{Y}_{feature}$ serves as the input node feature substrate for the downstream topological learning module.

2.3.2. Dual-Stream Kinetic Topological Learning Module

This module maps multi-sensor representations using parallel graph learning pathways: a physics-prior kinetic stream and a data-driven attention stream.

The physical stream branch inputs the parameterized kinetic adjacency matrix $\mathbf{A}_{kinetic}$ from Equation (9) as the deterministic topology for spatial graph convolutions. The feature propagation at the l -th layer follows the spectral graph aggregation rule:

$$\mathbf{H}_{phy}^{(l+1)} = \sigma(\mathbf{D}^{-\frac{1}{2}} \mathbf{A}_{kinetic} \mathbf{D}^{-\frac{1}{2}} \mathbf{H}^{(l)} \mathbf{W}_{phy}^{(l)}) \quad (12)$$

where $\mathbf{D}$ is the diagonal degree matrix computed as $D_{ii} = \sum_j A_{kinetic}(i,j)$, $\mathbf{W}_{phy}^{(l)}$ is the learnable weight transformation matrix, and $\sigma(\cdot)$ denotes the non-linear activation function.

Concurrently, the data-driven graph attention stream assumes a fully connected initial topology and calculates time-varying correlation weights. The dynamic attention coefficient e_{ij} between node i and node j is calculated as:

$$e_{ij} = \text{LeakyReLU}(\mathbf{a}^T [\mathbf{W}_{att} \mathbf{h}_i \parallel \mathbf{W}_{att} \mathbf{h}_j]) \quad (13)$$

where $\mathbf{W}_{att}$ is the learnable linear transformation matrix mapping the input features to a hidden space of dimension $\mathbf{F}'$, $\mathbf{a}$ is the parameterized interaction vector, and $\parallel$ denotes

the vector concatenation operation. The correlation coefficients are normalized across the neighborhood N_i via the Softmax function to yield the dynamic weights α_{ij} :

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})} \quad (14)$$

To capture heterogeneous feature dynamics across distinct representation subspaces, a K -head attention mechanism independently evaluates node correlations. The updated node representation $\mathbf{h}'_i$ is aggregated by concatenating the head outputs along the channel dimension:

$$\mathbf{h}'_i = \left\|_{k=1}^K \sigma \left(\sum_{j \in N_i} \alpha_{ij}^k \mathbf{W}^k \mathbf{h}_j \right) \right. \quad (15)$$

where $\mathbf{W}^k$ corresponds to the shared weight matrix of the k -th attention head.

2.3.3. Adaptive Gated Fusion and Fault Classification Module

To combine the structural representation $\mathbf{H}_{phy}$ from the physical stream and the dynamic coupling representation $\mathbf{H}_{att}$ from the attention stream, an adaptive gating unit dynamically generates a balance coefficient $\lambda \in (0, 1)$:

$$\lambda = \sigma(\mathbf{W}_g [\mathbf{H}_{phy} \parallel \mathbf{H}_{att}] + \mathbf{b}_g) \quad (16)$$

where $\mathbf{W}_g$ and $\mathbf{b}_g$ represent the weight matrix and bias vector of the gating network, respectively, and $\sigma(\cdot)$ denotes the Sigmoid function. The final decision feature vector $\mathbf{H}_{final}$ is structured via weighted fusion:

$$\mathbf{H}_{final} = \lambda \mathbf{H}_{phy} + (1 - \lambda) \mathbf{H}_{att} \quad (17)$$

An MLP classifier maps $\mathbf{H}_{final}$ into the specific fault category space covering c health classes ($Y = \{y_1, y_2, \dots, y_c\}$). The Softmax function in the output layer calculates the conditional probability distribution:

$$P(y = c|x) = \text{Softmax}(z_c) = \frac{\exp(z_c)}{\sum_{k=1}^c \exp(z_k)} \quad (18)$$

where $\mathbf{z} = \mathbf{W}_{cls} \mathbf{H}_{final} + \mathbf{b}_{cls}$ represents the logit vector, with $\mathbf{W}_{cls}$ and $\mathbf{b}_{cls}$ denoting the classification layer weights and biases.

The total objective function L to train the model utilizes the cross-entropy loss augmented with an L_2 regularization term to prevent overfitting under small-sample constraints:

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(p_{i,c}) + \eta \|\Theta\|_2^2 \quad (19)$$

where N denotes the training batch size, $y_{i,c}$ is the true label of sample i for class c , $p_{i,c}$ is the corresponding predicted probability, Θ represents the set of all trainable network parameters, and η is the regularization coefficient. Optimization is conducted iteratively via the Adam algorithm based on error backpropagation.

3. Experimental Setup and Data Acquisition

To rigorously evaluate the diagnostic performance and practical applicability of the proposed dynamics-driven dual-stream graph neural network, systematic fault experiments were conducted on a custom multi-stage gearbox test rig. This section details the

structural composition of the experimental bench, the simulated gearbox fault categories, the operating profiles, and the vibration data acquisition methodology.

3.1. Physical Experimental Platform and Data Acquisition Scheme

Experimental verification is conducted on the HP-GBS-2023 gear fault prognosis testbed to capture authentic failure-transient signatures and establish clear dynamic propagation pathways under simulated industrial operational profiles.

3.1.1. Gearbox Fault Diagnosis Experimental Platform

The experimental platform is structured as a back-to-back closed-loop mechanical configuration comprising three functional blocks: a driving unit, a transmission testing unit, and an electrical load simulation unit (Figure 2).

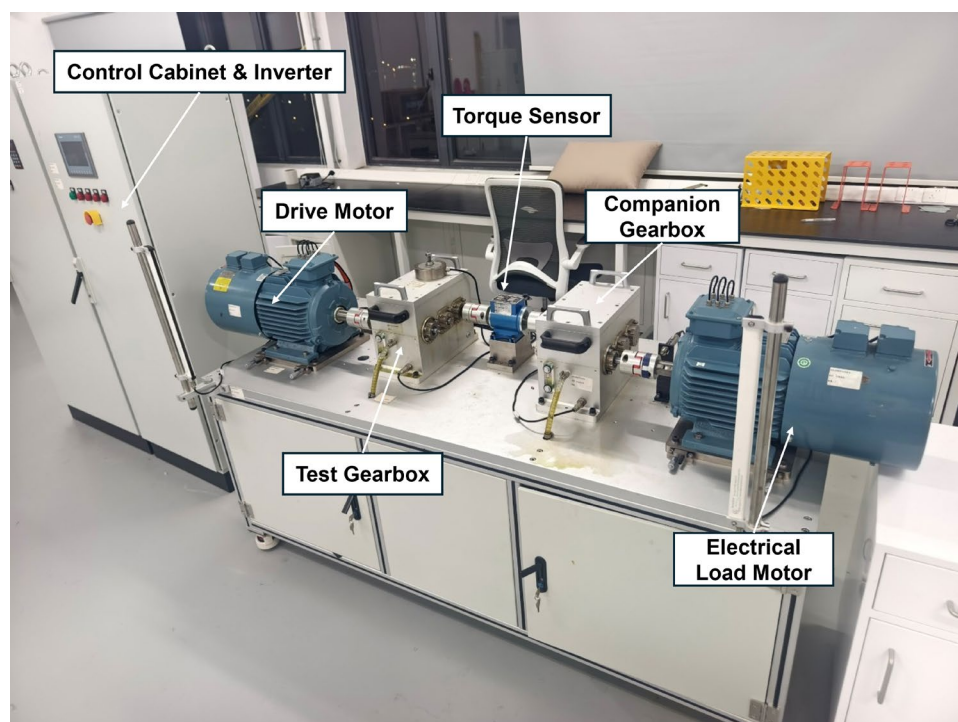


Figure 2. Schematic diagram of the HP-GBS-2023 gear fault testbed.

The driving unit utilizes an ABB QABP132S2P-B3 variable-frequency three-phase asynchronous motor maintaining a rated power of 7.5 kW and a rated speed of 2910 rpm. An inverter regulates continuous speed profiles within a 0–3000 rpm range to simulate the non-stationary speed fluctuations characteristic of port hoisting equipment. The transmission testing core consists of a two-stage unfolded cylindrical reduction gearbox with three parallel shaft systems: input, intermediate, and output. The first-stage and second-stage gear ratios are structured at 48:120 and 50:120, respectively. Output mechanical power is decelerated by the test gearbox and subsequently transmitted to a companion speed-increasing gearbox to maintain an overall system transmission ratio of 1:1. A modular quick-disassembly casing configuration allows for the substitution of specific faulty gear components without base displacement. The load simulation unit implements an identical ABB asynchronous motor to apply configurable load boundaries via regenerative braking, delivering smooth constant or dynamic torques with a control precision of 0.1 N·m.

3.1.2. Data Acquisition System Configuration and Sensor Placement

To ensure high-fidelity capture of high-frequency gear-meshing transient impacts, three accelerometers are rigidly screwed directly into threaded mounting holes reserved on the upper casing cover above the input, intermediate, and output shaft bearing housings. This rigid connection eliminates geometric wave scattering and signal distortion. Concurrently, the testbed incorporates a P4501-50KN tension-compression load cell to monitor radial support loading forces, alongside a photoelectric tachometer for real-time shaft rotational speed tracking. This sensor matrix achieves synchronous acquisition of mechanical load profiles and vibration responses. The systemic sampling frequency for all collaborative channels is fixed at 12,800 Hz.

3.1.3. Experimental Condition Design and Fault Injection Scheme

The experimental matrix comprises 11 discrete structural fault modes and 9 operating conditions, generating 108 operational combinations when including the baseline healthy state. Fault replication is executed via component replacement targeting two critical transmission interfaces: the high-speed intermediate-stage pinion and the final-stage output large gear.

The injected damage categories span diverse wear and structural fatigue profiles:

Final-stage output large gear (6 modes): Broken tooth, half-broken tooth, arcuate recession, distributed large-hole pitting, dense small-hole pitting, and uniform wear.

Intermediate-stage pinion (5 modes): Broken tooth, half-broken tooth, arcuate recession, dense small-hole pitting, and uniform wear.

The operational matrix combines three target rotational frequencies—low speed (300 rpm), medium speed (500 rpm), and high speed (800 rpm)—with three step-wise load torques regulated via the host control system: 0 N·m (no load), 5 N·m (half load), and 10 N·m (rated load). The data acquisition duration for each discrete fault-speed-load variant is fixed at 1 min. Under a uniform sampling frequency of 12,800 Hz, this sampling window establishes the necessary data density for downstream statistical and spectral feature analysis.

3.2. Experimental Dataset Partitioning and Model Parameter Configuration

3.2.1. Dataset Construction and Partitioning Strategy

To evaluate diagnostic robustness under operational domain shifts, a mixed-condition dataset is constructed by pooling vibration data across all predefined rotational speeds (300 rpm, 500 rpm, and 800 rpm) and load torques (0 N·m, 5 N·m, and 10 N·m). The continuous raw sequences are segmented via an overlapping sliding window technique with a window length of $L = 2048$ and a 50% overlap rate, generating 80,075 total samples. To simulate actual industrial deployment constraints, dataset partitioning is executed chronologically along the data acquisition axis using a strict 7:2:1 ratio (Table 1).

Table 1. Dataset partitioning details.

Dataset Subset	Number of Samples	Proportion	Purpose
Training set	56,341	70%	Iterative weight optimization and feature representation learning
Validation set	15,917	20%	Hyperparameter tuning and overparameterization monitoring
Testing set	7817	10%	Objective evaluation of finalized diagnostic performance

3.2.2. Physical and Dynamic Parameter Mapping

The mechanical invariants required for initializing the physical graph prior are configured according to the structural properties of the testbed powertrain. The radial support stiffness governing the bearing housing and transmission shaft assembly k_{supprt} is parameterized at 1.0×10^8 N/m. The equivalent gear-pair meshing stiffness is 5.0×10^8 N/m. These values parameterize the baseline kinetic adjacency matrix $\mathbf{A}_{\text{kinetic}}$ to constrain neighborhood feature propagation along physical energy transmission pathways.

3.2.3. Network Hyperparameters and Training Environment Configuration

Model training is executed within the PyTorch 2.8.0 framework and accelerated via CUDA 12.8 on a hardware stack comprising an AMD R9-8940HX CPU and an NVIDIA RTX 5060 Laptop GPU. Parameter optimization is driven by the Adam algorithm over a maximum limit of 20 epochs. The complete hyperparameter configuration and environmental properties are detailed in Table 2.

Table 2. Model hyperparameters and experimental environment configuration.

Category	Parameter	Value/Description
Input Configuration	Sample size	3×2048 (Three-channel synchronous input)
	Batch size	64
Network Structure	CNN kernel size	64-3-3 (Multi-scale receptive fields)
	CBAM reduction ratio	$R = 16$
	GAT attention heads	$K = 4$
	Hidden layer feature dimension	64
Training Strategy	Optimizer	Adam
	Initial learning rate	1×10^{-3}
	Learning rate decay strategy	Cosine annealing
	Maximum epochs	20
	Regularization coefficient	1×10^{-4}
Hardware Environment	CPU	AMD R9-8940HX
	GPU	NVIDIA RTX 5060 Laptop
	Acceleration library	CUDA 12.8

To eliminate empirical selection bias and ensure architectural reproducibility, the rationales and grid search spaces for the key hyperparameters listed in Table 2 are detailed as follows:

Convolutional Kernels (Selected: 64-3-3): Evaluated across {16,32,64,128}. Smaller initial kernels (16,32) failed to cover a full gear-meshing period under 12,800 Hz sampling, whereas 128 added unnecessary computational FLOPs without accuracy gains. The 64-3-3 hierarchy achieved the optimal trade-off by capturing low-frequency modulation envelopes in the first layer while isolating high-frequency shocks in deeper layers.

CBAM Reduction Ratio (Selected: $R = 16$): Evaluated within $R \in \{4, 8, 16, 32\}$. Smaller reduction ratios ($R = 4, 8$) introduced parameter redundancy, whereas $R = 32$ caused severe channel information bottlenecking. Setting $R = 16$ effectively suppressed noise-contaminated channels while preserving discriminative fault features with minimal parameter overhead.

GAT Attention Heads (Selected: $K = 4$): Tested across $K \in \{1, 2, 4, 8\}$. $K = 1, 2$ lacked capacity to disentangle overlapping spectral modes across shafts, while $K = 8$ led to attention weight dispersion and increased inference latency without accuracy improvement. $K = 4$ provided stable node weight convergence across independent representation subspaces.

Learning Rate (Selected: 1×10^{-3}): Searched within $\{10^{-2}, 10^{-3}, 10^{-4}\}$. A higher rate (10^{-2}) induced gradient oscillations around local minima, whereas 10^{-4} suffered from slow convergence. An initial rate of 1×10^{-3} paired with cosine annealing ensured rapid early convergence and smooth late-stage stabilization.

Epoch Horizon and Regularization: Capped at 20 epochs. Owing to the strong inductive bias from the physical stiffness matrix, validation loss plateaus near Epoch 15. A weight decay with a L_2 penalty coefficient of 1×10^{-4} coupled with a post-dense dropout operator prevents overparameterization and overfitting under non-stationary interference.

4. Experimental Results and Performance Analysis

In this section, extensive comparative experiments are carried out using the acquired gearbox vibration datasets. The proposed model is benchmarked against several baseline deep learning architectures using multi-dimensional metric evaluations. Furthermore, feature manifold visualizations, training convergence profiles, and structural ablation studies are presented to thoroughly validate the superiority and robustness of the proposed framework.

4.1. Fault Diagnosis Accuracy Comparison

To evaluate the KT-GNN-CBAM framework under non-stationary configurations, five baseline architectures spanning conventional deep learning models and standard graph neural networks are constructed:

Baseline DNN: A fully connected architecture that sequentially compresses high-dimensional inputs through a 4096–2048–512–128 dense hierarchy, serving as a baseline lacking temporal locality awareness or spatial topology constraints.

Baseline 1D-CNN: A standard convolution-pooling backbone comprising three convolutional layers with kernel sizes of 64, 32, and 16, with dropout, evaluating purely data-driven localized feature extraction.

Baseline ResNet-1D: An advanced data-driven network implementing residual skip connections across three distinct blocks to bypass deep gradient degradation. It evaluates whether structural depth alone can substitute for mechanism-driven physical priors.

Standard GCN: A standard graph convolutional network utilizing a uniform adjacency matrix without dynamic stiffness priors, evaluating unconstrained spatial message passing.

Standard GAT: A standard graph attention network utilizing multi-head dynamic attention over a fully connected node topology, assessing whether data-driven attention alone can resolve cross-channel noise contamination.

To eliminate data leakage and evaluate cross-temporal generalization bounds, data partitioning strictly follows the chronological acquisition sequence rather than random shuffling. The first 70% of continuous data forms the training set, the middle 20% constitutes the validation set, and the final 10% serves as the independent testing set, enforcing strict physical isolation across temporal blocks.

To replicate the harsh noise profiles of operational environments, Gaussian white noise enforcing a strict signal-to-noise ratio (SNR) of 6 dB is injected during data preprocessing. All baseline architectures are executed on matching computing hardware using identical classification heads and hyperparameter configurations. The statistical classification results computed over five independent trials are summarized in Table 3.

As quantified in Table 3, the proposed KT-GNN-CBAM framework achieves the highest overall diagnostic performance, yielding a mean classification accuracy of $97.71\% \pm 1.13\%$, Precision of 97.98%, Recall of 97.70%, and F1-Score of 97.66%. Among the evaluated baselines, Standard GCN exhibits the lowest accuracy (83.61%) because unconstrained graph message passing accelerates random noise propagation across spatial channels. Standard GAT (91.47%), 1D-CNN (91.93%), Baseline DNN (92.23%), and ResNet-

1D (94.42%) achieve improved accuracy by processing temporal features independently; however, their purely data-driven architectures lack physical structural constraints. Paired *t*-tests confirm that the accuracy improvements of KT-GNN-CBAM over all baseline models are statistically significant ($p < 0.01$). By embedding the dynamic stiffness matrix as an analytical graph prior, the proposed framework constrains feature aggregation strictly along physical energy-transmission pathways, suppressing noise propagation across multi-stage gearbox shafts.

Table 3. Gearbox fault diagnosis results (5-run mean $\pm$ std).

Model Architecture	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	<i>p</i> -Value (vs. Ours)
DNN	92.23 $\pm$ 0.47	92.43 $\pm$ 0.41	92.23 $\pm$ 0.47	92.27 $\pm$ 0.46	0.0002
1D-CNN	91.93 $\pm$ 1.96	92.49 $\pm$ 1.80	91.93 $\pm$ 1.97	91.89 $\pm$ 2.06	0.0011
ResNet-1D	94.42 $\pm$ 1.40	94.75 $\pm$ 1.31	94.41 $\pm$ 1.40	94.41 $\pm$ 1.38	0.0039
Standard GCN	83.61 $\pm$ 3.17	85.88 $\pm$ 2.17	83.61 $\pm$ 3.17	83.57 $\pm$ 3.11	<0.0001
Standard GAT	91.47 $\pm$ 1.91	92.23 $\pm$ 1.80	91.47 $\pm$ 1.91	91.43 $\pm$ 1.96	0.0006
KT-GNN-CBAM	97.71 $\pm$ 1.13	97.98 $\pm$ 1.05	97.70 $\pm$ 1.13	97.66 $\pm$ 1.15	N/A

To further evaluate model diagnostic boundaries under extreme noise interference, a severe noise stress test (0 dB SNR) was conducted across all architectures. The statistical classification results over five independent runs are summarized in Table 4.

Table 4. Diagnostic accuracy comparison under severe noise stress test (0 dB, 5-run mean $\pm$ std).

Model	Accuracy (%)	<i>p</i> -Value (vs. Ours)
DNN	86.11 $\pm$ 0.44	0.0684
1D-CNN	66.40 $\pm$ 6.67	0.0100
ResNet-1D	64.09 $\pm$ 7.61	0.0047
Standard GCN	45.44 $\pm$ 3.96	<0.0001
Standard GAT	46.73 $\pm$ 7.90	<0.0001
KT-GNN-CBAM	91.26 $\pm$ 4.22	N/A

As presented in Table 4, unconstrained baseline architectures (including 1D-CNN, ResNet-1D, Standard GCN, and Standard GAT) experience severe diagnostic degradation under heavy 0 dB background noise. Notably, standard graph networks (Standard GCN and Standard GAT) collapse to 45.44% and 46.73% accuracy, respectively, because unconstrained spatial message passing accelerates noise propagation across node channels. In sharp contrast, the proposed KT-GNN-CBAM preserves superior diagnostic efficacy at 91.26%, outperforming standard graph architectures by over 44 percentage points. Although Baseline DNN exhibits numerical resistance (86.11%) due to dense linear spatial smoothing across 6144 flattened dimensions, it relies on an excessively large parameter footprint (6.58 M parameters, as detailed later in Section 4.4). The proposed framework achieves higher classification accuracy (91.26%) with a lightweight footprint of only 52.01 K parameters (less than 1% of the DNN footprint). The moderate performance variance of the proposed model reflects the active, data-adaptive feature recalibration executed by the CBAM and adaptive gating modules across stochastically perturbed noise profiles. This trade-off confirms that embedding mechanical stiffness priors effectively safeguards feature-representation manifolds under extreme environmental interference.

To further evaluate class-wise discrimination and performance under varying decision thresholds, multi-class Receiver Operating Characteristic (ROC) and Precision-Recall (P-R) curves were generated for comparative analysis (Figures 3 and 4).

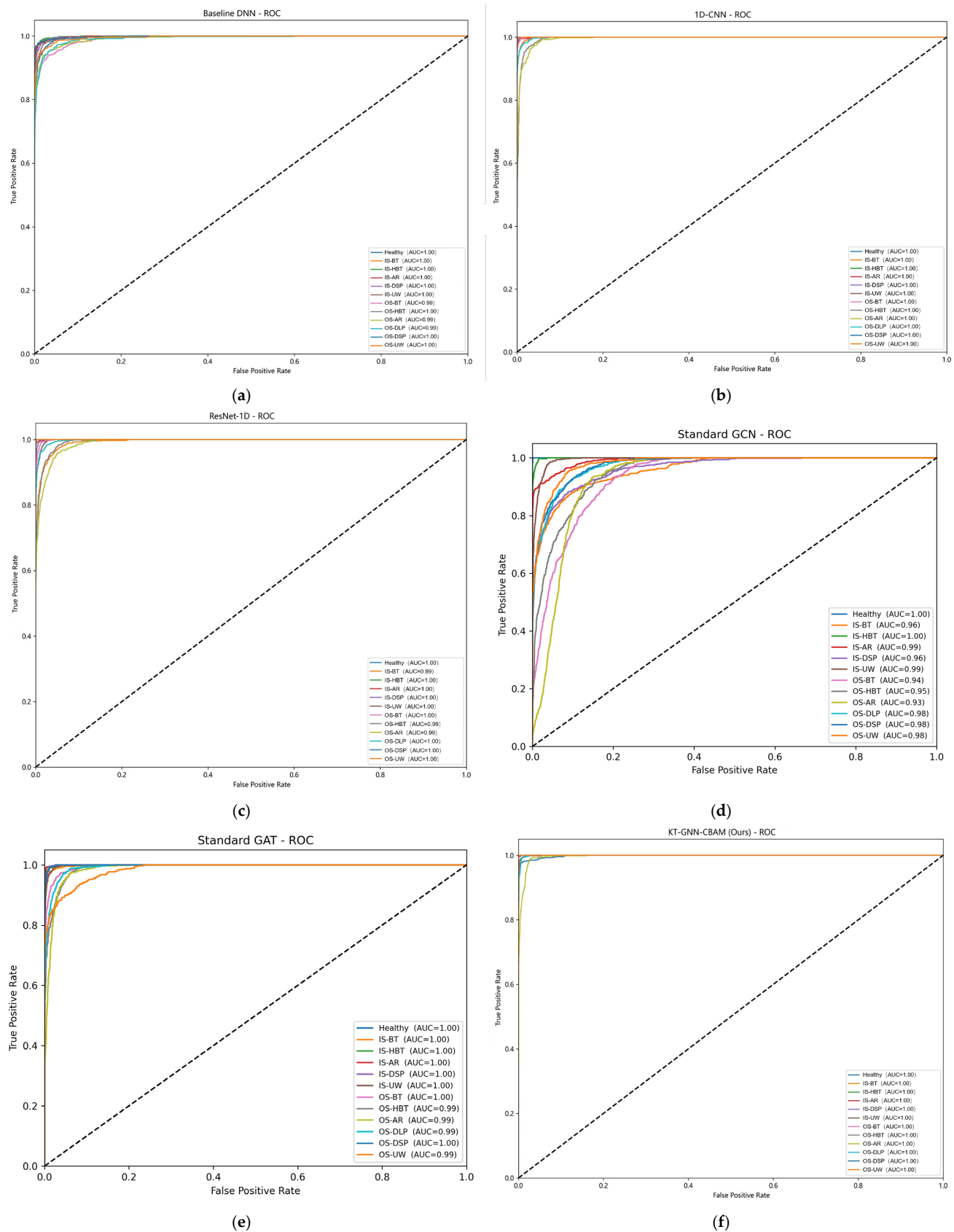


Figure 3. ROC curves of different methods for gearbox fault diagnosis: (a) ROC curve of DNN; (b) ROC curve of 1D-CNN; (c) ROC curve of ResNet-1D; (d) ROC curve of GCN; (e) ROC curve of GAT; (f) ROC curve of KT-GNN-CBAM. (IS: Intermediate shaft; OS: Output shaft; BT: Broken tooth; HBT: Half-broken tooth; AR: Arcuate recession; DSP: Dense small-hole pitting; DLP: Distributed large-hole pitting; UW: Uniform wear).

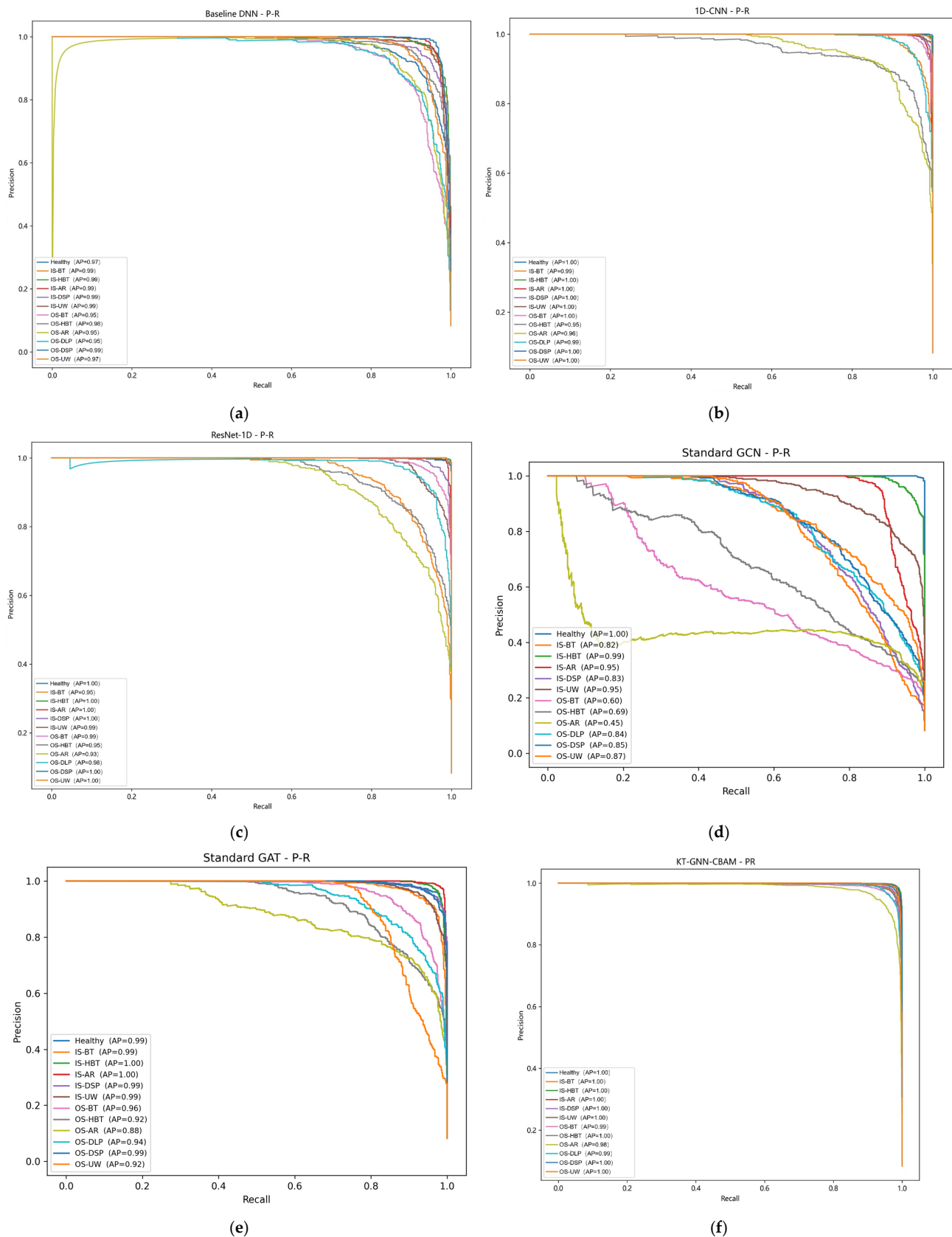


Figure 4. P-R curves of different methods for gearbox fault diagnosis: (a) P-R curve of DNN; (b) P-R curve of 1D-CNN; (c) P-R curve of ResNet-1D; (d) P-R curve of GCN; (e) P-R curve of GAT; (f) P-R curve of KT-GNN-CBAM. (IS: Intermediate shaft; OS: Output shaft; BT: Broken tooth; HBT: Half-broken tooth; AR: Arcuate recession; DSP: Dense small-hole pitting; DLP: Distributed large-hole pitting; UW: Uniform wear).

The multi-class ROC curves on the testing set display high baseline metrics across most architectures, leading to dense curve clustering toward the ideal (0,1) coordinate. However, structural disparities are evident among the baseline models: Standard GCN exhibits observable AUC contractions due to noise amplification during spatial message passing, whereas Standard GAT achieves moderate metric recovery by dynamically weighting node dependencies. While the Baseline DNN and ResNet-1D display minor AUC-tracking reductions, the proposed paradigm approaches an AUC of 1.00 across all 12 failure modes, maintaining a stable True Positive Rate (TPR) under highly restricted False Positive Rate (FPR) boundaries.

The P-R curves further clarify structural performance disparities under simulated operational limits and class-imbalance constraints:

Baseline DNN: Exhibits unstable tracking boundaries, undergoing an immediate vertical drop in the output shaft arcuate recession mode under high-confidence intervals due to high noise susceptibility.

Baseline 1D-CNN: Triggers earlier, gradual precision drops across diverse classifications, indicating decision ambiguity along adjacent decision thresholds.

Baseline ResNet-1D: Demonstrates enhanced deep feature isolation but undergoes a localized sharp vertical plunge in the output shaft distributed large-hole pitting mode, revealing feature aliasing under operational domain shifts.

Standard GCN: Demonstrates the most severe performance degradation, with precision-recall curves plunging drastically in output shaft failure modes. Uniform, unconstrained graph convolution allows heavy vibration noise to propagate globally across nodes, causing severe feature over-smoothing and inter-node aliasing.

Standard GAT: Outperforms Standard GCN by leveraging dynamic attention coefficients to suppress noise-contaminated connections, maintaining AP values and high overall precision metrics across most health states. However, it still exhibits moderate precision decay in subtle fault overlaps, demonstrating that purely data-driven attention mechanisms without physical stiffness topologies cannot fully isolate noise from genuine kinetic pathways.

KT-GNN-CBAM: Adheres tightly to the ideal (1,1) upper-right limit without abnormal plunges or threshold degradation, indicating that pairing a physical topology prior with CBAM channel recalibration guarantees stable threshold resilience under strict recall requirements.

To evaluate the microscopic discriminative behaviors across specific fault categories under heavy background noise, raw confusion matrices are introduced to conduct a fine-grained analysis of the predicted distributions (Figure 5). The matrices visualize the exact sample counts misclassified between distinct failure modes, revealing the localized diagnostic bottlenecks and structural failure trajectories for each evaluated architecture.

Baseline DNN: While establishing a high overall classification baseline, its prediction errors are sparsely scattered across various non-diagonal grid positions. This wide error distribution indicates a generalized vulnerability to stochastic noise and a lack of structural localization constraints.

Baseline 1D-CNN: Despite delivering an overall performance improvement over the DNN, it triggers a severe diagnostic bottleneck involving 213 misclassified sample cases between the output shaft half-broken tooth and intermediate shaft broken tooth categories. This high error aggregation demonstrates that purely data-driven convolutional backbones lack sufficient physical discriminative depth when processing cross-shaft failures possessing similar transient shock envelopes.

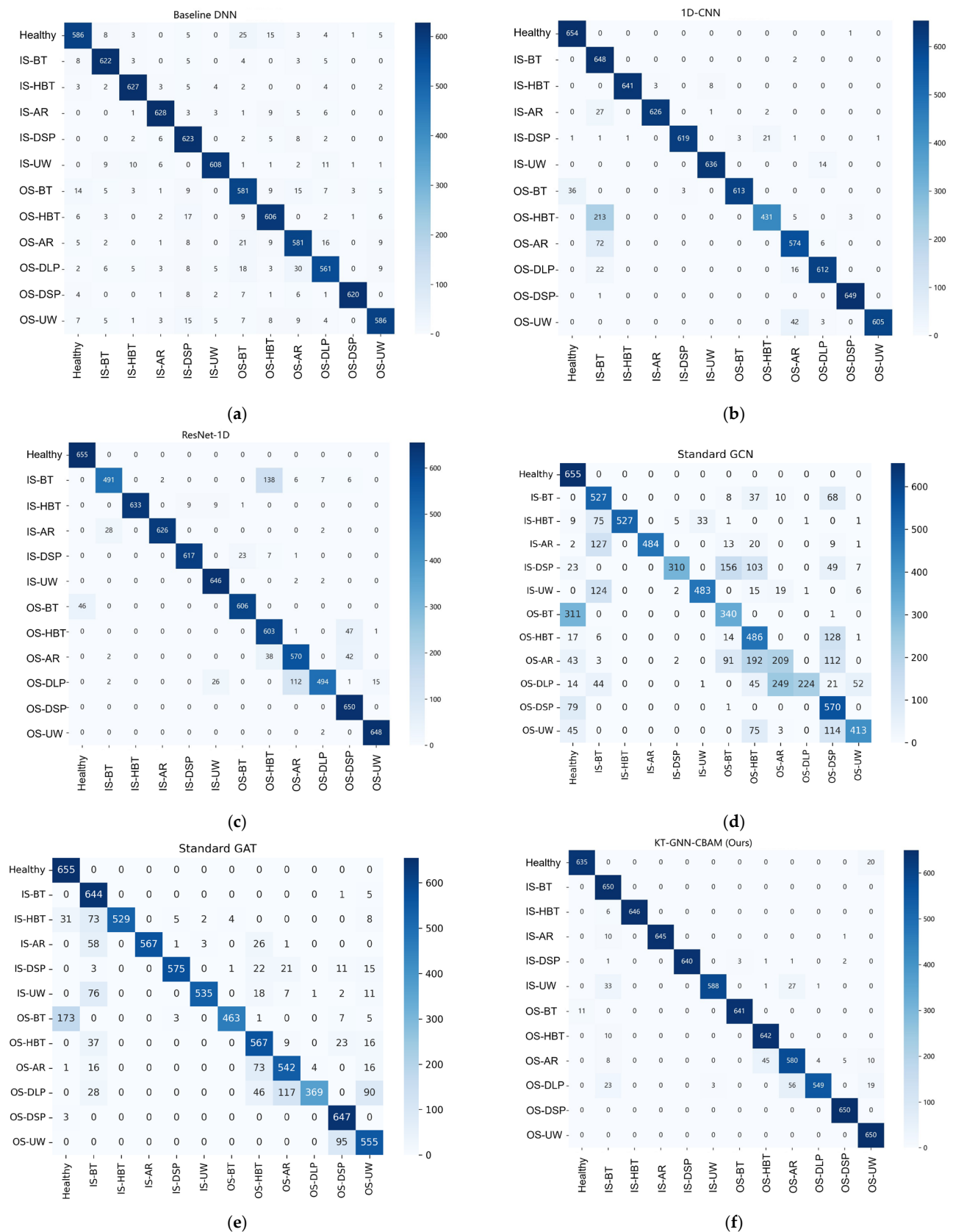


Figure 5. Confusion matrices of different methods for gearbox fault diagnosis: (a) Confusion matrix of DNN; (b) Confusion matrix of CNN; (c) Confusion matrix of ResNet-1D; (d) Confusion matrix of GCN; (e) Confusion matrix of GAT; (f) Confusion matrix of KT-GNN-CBAM. (IS: Intermediate shaft; OS: Output shaft; BT: Broken tooth; HBT: Half-broken tooth; AR: Arcuate recession; DSP: Dense small-hole pitting; DLP: Distributed large-hole pitting; UW: Uniform wear).

Baseline ResNet-1D: Similarly encounters localized classification bottlenecks across adjacent transmission domains, yielding 138 misclassified instances between the intermediate shaft broken tooth and output shaft half-broken tooth modes, alongside 112 misclassified instances between the output shaft distributed large-hole pitting and output shaft arcuate recession failure modes.

Standard GCN: Exhibits the most severe off-diagonal error dispersion and cluster aggregation across multiple transmission components. The unconstrained, uniform graph aggregation leads to extensive decision ambiguity—most prominently confusing severe output shaft tooth breakage with healthy operational states, as well as inducing widespread mutual aliasing among adjacent output shaft recession and pitting modes. This catastrophic error scattering confirms that unweighted spatial graph convolution over-smooths discriminative vibration features under heavy operational noise.

Standard GAT: Substantially mitigates broad off-diagonal scattering compared to Standard GCN, recovering strong diagonal dominance by leveraging dynamic attention weighting. Nevertheless, localized misclassification bottlenecks persist, particularly regarding severe output shaft tooth structural faults misidentified as healthy states, alongside persistent aliasing between adjacent surface pitting and uniform wear modes. This performance indicates that while dynamic node weighting suppresses random background noise, it lacks kinetic topology priors to fully disentangle overlapping multi-scale shock signatures.

KT-GNN-CBAM: Exhibits prominent diagonal dominance with near-zero off-diagonal counts, effectively resolving the cross-shaft and cross-component semantic confusion boundaries encountered by the baselines. This performance verifies that injecting mechanism-driven physical priors via kinetic topologies successfully establishes mechanics-compliant neighborhood constraints during the spatial feature propagation phase. Concurrently, the integrated CBAM module adaptively recalibrates critical channel and temporal vectors to isolate weak discriminative fault components from dominant environmental noise interferences, achieving superior decision robustness in complex gearbox diagnostics.

4.2. Deep Analysis of Feature Space Visualization

To evaluate high-dimensional feature extraction capabilities under noise contamination and operational domain shifts, the t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm is deployed. High-dimensional feature vectors extracted from the penultimate layer (the fully connected layer or global pooling layer) of each architecture are mapped onto a two-dimensional manifold to analyze fault-mode decoupling capacities and environmental robustness qualitatively (Figure 6).

Baseline DNN: The feature distribution displays extensive fracturing, where identical fault categories fragment into multiple discontinuous sub-clusters. This topological degradation indicates a high susceptibility to varying rotational speeds and load configurations, causing the network to misidentify operational fluctuations as dominant discriminative features and disrupting physical continuity within the feature manifold. Furthermore, distinct fault clusters entangle within the central region to form a disordered aggregate, verifying that fully connected layers encounter diagnostic blind spots when isolating weak fault excitations from background noise under strongly coupled conditions.

Baseline 1D-CNN, ResNet-1D, Standard GCN and Standard GAT: While localized convolutional and standard graph operators improve cluster grouping compared to the DNN, these baseline networks retain a visible tendency to fracture into separated sub-clusters across multiple fault modes due to operational domain shifts.

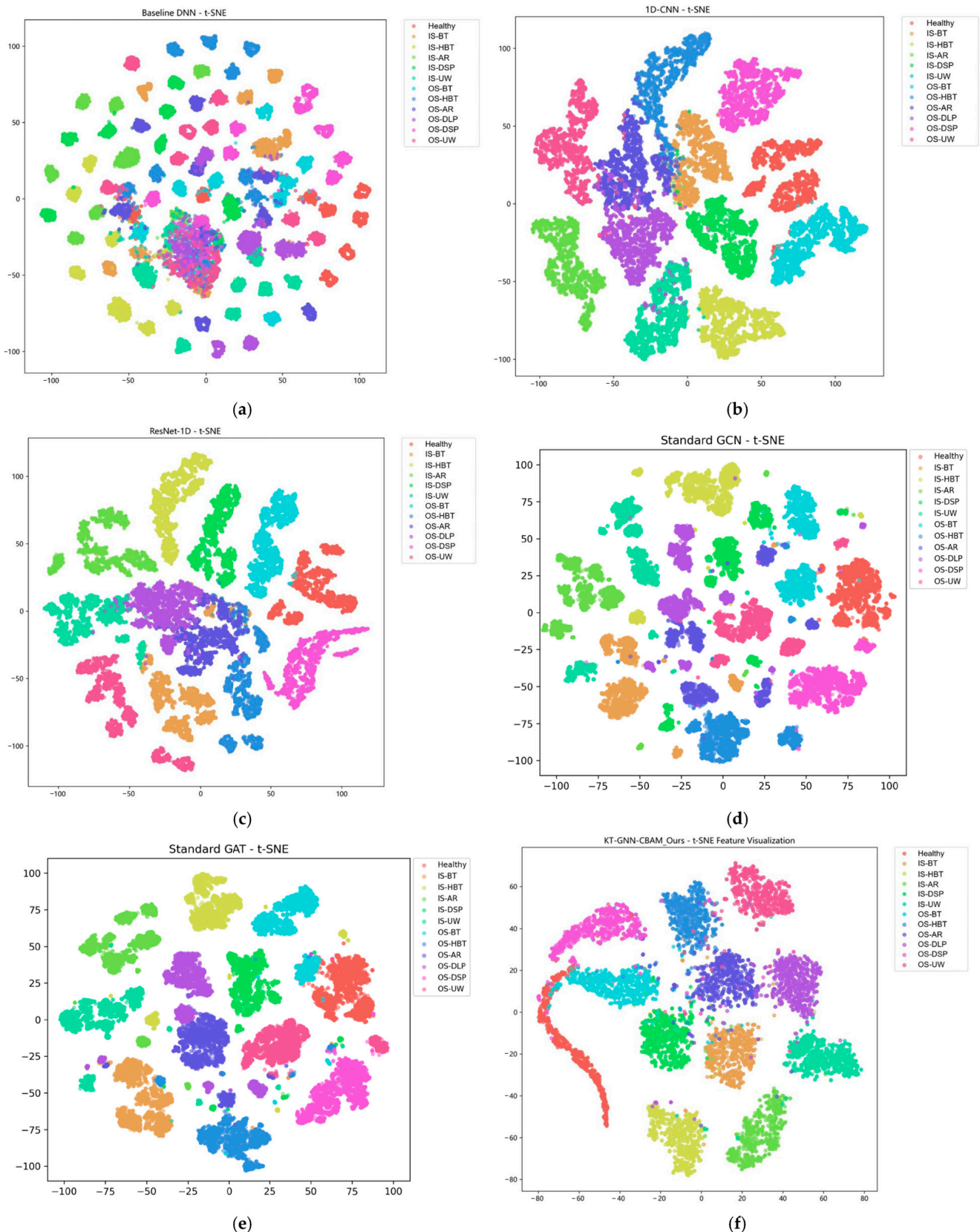


Figure 6. t-SNE visualizations of different methods for gearbox fault diagnosis: (a) t-SNE visualization of DNN; (b) t-SNE visualization of CNN; (c) t-SNE visualization of ResNet-1D; (d) t-SNE visualization of GCN; (e) t-SNE visualization of GAT; (f) t-SNE visualization of KT-GNN-CBAM. (IS: Intermediate shaft; OS: Output shaft; BT: Broken tooth; HBT: Half-broken tooth; AR: Arcuate recession; DSP: Dense small-hole pitting; DLP: Distributed large-hole pitting; UW: Uniform wear).

KT-GNN-CBAM: The 12 target health categories form highly compact intra-class clusters with distinct inter-class separation boundaries. This global clustering configuration verifies the efficacy of the parameterized kinetic topology as a physical inductive bias, which guides the network to decouple invariant mechanical failure signatures from non-stationary speed and load fluctuations. Concurrently, the integrated CBAM module increases the representation weights of fault-sensitive channels, transitioning from generic statistical feature mapping to dynamics-guided feature representations. This manifold separation directly corroborates the classification robustness demonstrated in the corresponding P-R curves and confusion matrices.

4.3. Convergence Analysis of the Training Process

The training and validation loss curves alongside the corresponding classification accuracy curves of the proposed model are illustrated in Figure 7.

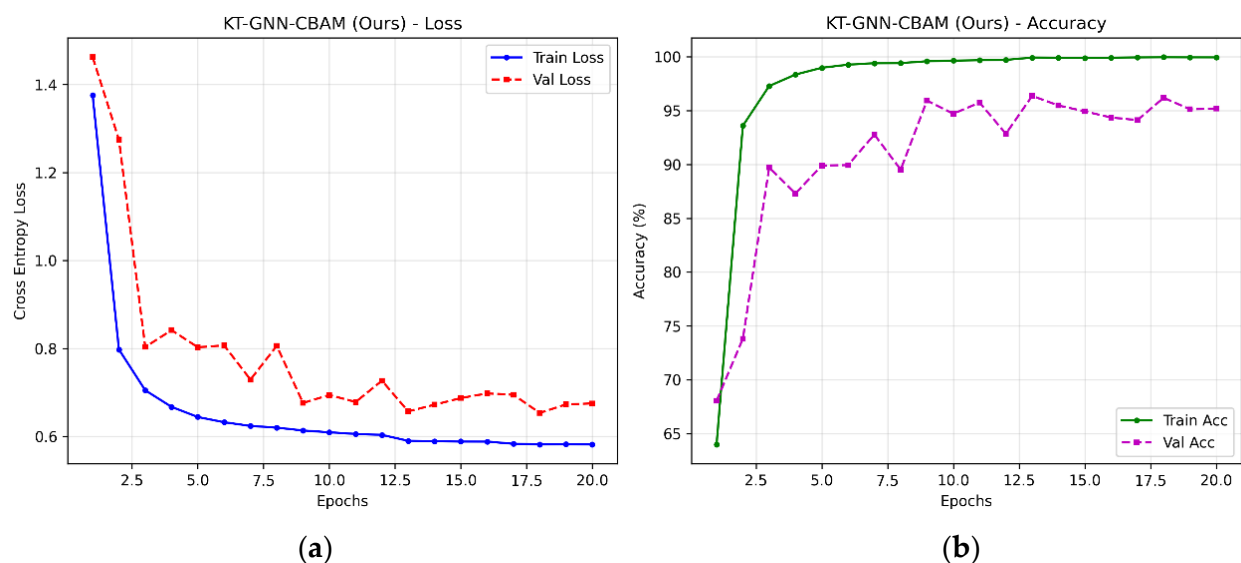


Figure 7. Loss and accuracy curves of the proposed KT-GNN-CBAM: (a) Loss curves of the proposed KT-GNN-CBAM; (b) Accuracy curves of the proposed KT-GNN-CBAM.

As illustrated, the training loss curve decreases monotonically within the first 5 epochs, matching a simultaneous asymptotic ascent in training accuracy. This rapid stabilization trajectory indicates that the front-end CBAM layers effectively isolate salient transient impact signatures along the channel and temporal axes, while the physical structural weights embedded in the KT-GNN module suppress environmental noise interference to accelerate parameter convergence toward the target solution domain.

The validation loss and accuracy curves track the macro trajectory of the training metrics but superimpose a serrated, oscillating morphology. This pattern reflects localized decision-boundary adjustments within the high-dimensional feature space, driven by the inherent non-stationarity of the multi-condition gearbox vibration inputs rather than by structural overfitting.

During the late optimization phase, the validation loss experiences minor tracking fluctuations and a marginal rebound, accompanied by a localized saturation in validation accuracy. To prevent overparameterization and preserve generalization limits near the global minimum, an early-stopping protocol triggers checkpoint preservation at the peak validation accuracy instance. This optimal weight configuration yields a final diagnostic accuracy of 98.84% on the independent testing set.

4.4. Computational Complexity and Real-Time Performance Analysis

To evaluate the operational feasibility for real-time deployment on embedded port crane controllers, the computational complexity and inference latency of the proposed KT-GNN-CBAM were quantified alongside baseline architectures. As summarized in Table 5, model parameters (Params), floating-point operations per sample (FLOPs), and average single-sample inference execution time across 500 benchmark trials on both GPU (NVIDIA RTX) and CPU hardware were recorded.

Table 5. Model complexity and inference latency analysis.

Model Architecture	Parameters (K)	FLOPs (M)	GPU Latency (ms)	CPU Latency (ms)
DNN	1615.95	3.23	0.39	0.16
1D-CNN	4.13	2.27	0.48	0.26
ResNet-1D	7.68	5.67	0.85	0.53
Standard GCN	35.29	21.78	0.71	0.73
Standard GAT	35.29	21.78	0.99	0.74
KT-GNN-CBAM	52.01	21.96	2.23	1.69

The proposed KT-GNN-CBAM contains 52.01 K trainable parameters with a computational footprint of 21.96 M FLOPs per 2048-point time-series sample. The average single-sample inference latency is 2.23 ms on GPU and 1.69 ms on CPU. Although lightweight 1D-CNNs exhibit lower latency (0.48 ms on GPU) due to minimal parameter counts (4.13 K), their classification accuracy degrades significantly under 6 dB noise conditions (91.93%). Considering that a 2048-point signal sequence sampled at 12,800 Hz corresponds to a physical acquisition window of 160 ms, the 1.69 ms execution time of KT-GNN-CBAM consumes approximately 1% of the sampling interval. This confirms that the dual-stream physics-guided model achieves an optimal balance between high-noise resilience, structural interpretability, and real-time execution capability for industrial edge monitoring.

4.5. Ablation Study and Model Component Effectiveness Analysis

To isolate the individual quantitative performance contributions of the core structural modules within the complete KT-GNN-CBAM framework, three model variants are evaluated under identical dataset partitioning and noise boundary conditions:

w/o CBAM: Removes the channel and spatial attentional layers from the front end of the feature extractor, retaining only the standard 1D-CNN backbone as the preprocessing block. It evaluates the influence of multidimensional attentional filtering when capturing transient impact signatures under noise-contaminated backgrounds.

w/o Physics: Eliminates the physical GCN branch initialized by the dynamic stiffness matrix while retaining exclusively the data-driven GAT pathway. The module collapses into a purely data-driven graph neural network to evaluate the regularizing effect of mechanical prior knowledge on the topology of the feature space.

w/o Gating: Preserves the dual-stream graph network structure but removes the learnable adaptive gating unit, forcing direct cross-stream feature fusion of the physical stream matrix $\mathbf{H}_{phy}$ and the data-driven stream matrix $\mathbf{H}_{att}$ via element-wise addition under a fixed 1:1 weight allocation. It validates the role of the dynamic weight distribution mechanism during heterogeneous information fusion.

The statistical results computed across the ablation variants are summarized in Table 6.

Table 6. Statistical results of the ablation study.

Model	Accuracy (%)	Performance Degradation (%)	Relative Increase in Misclassification Rate (%)
KT-GNN-CBAM	98.87	-	-
w/o Gating	96.50	2.37	209.73
w/o CBAM	96.42	2.45	216.81
w/o Physics	95.68	3.19	282.30

As quantified in Table 5, the exclusion of the physics-prior stream (w/o Physics) triggers the most significant performance degradation, reducing diagnostic accuracy from 98.87% to 95.68%. While the absolute precision drop is 3.19 accuracy points, evaluated from the perspective of reliability engineering, the relative misclassification rate surges abruptly by 282.30%. This performance collapse verifies that for highly nonlinear mechanical systems operating under complex variable operating conditions, purely data-driven models struggle to isolate the invariant physical features underlying distinct fault mechanisms. Conversely, incorporating physical topological constraints yields a rigid inductive skeleton for the network, effectively rectifying domain-shift-induced feature drift at the low-dimensional manifold level and extending the pattern recognition upper bound in high-dimensional feature spaces.

The performance degradation of the w/o CBAM variant, representing an absolute decrease of 2.45 percentage points, highlights the importance of front-end feature filtering. Raw vibration sequences are heavily contaminated by stochastic background noise. Without the channel attention mechanism to weight specific fault-salient resonance bands and the spatial attention mechanism to focus on temporal impact envelopes, the discriminative quality of features propagated to the downstream graph neural networks degrades. The ablation results demonstrate that the CBAM module acts as a crucial filtering baseline, preventing weak, incipient fault signatures from being submerged by environmental noise interferences and maintaining high diagnostic sensitivity in low-SNR environments.

The removal of the adaptive gating unit (w/o Gating) results in a 2.37% reduction in classification accuracy. This degradation establishes that a simplistic linear superposition of physical priors and data-driven features does not guarantee optimal joint representation learning. When the physics stream and data stream encounter semantic conflicts across distinct feature dimensions—such as during localized measurement anomalies or transient signal distortions—forced fusion lacking an adaptive gating mechanism introduces deleterious feature noise into the joint representation space. Conversely, deploying a learnable soft-gating mechanism dynamically updates cross-stream feature weights, successfully filtering out mutually interfering components to maximize the synergistic advantages of the dual-stream architecture.

To provide quantitative and visual evidence explaining the performance degradation observed in the w/o Gating variant, the degree of feature divergence between the physical stream (h_{phy}) and the data-driven attention stream (h_{att}) was quantified using the cosine similarity metric $S_{cos}(h_{phy}, h_{att})$. As illustrated in Figure 8a, the cosine similarity distribution spans a negative-dominant spectrum ranging from -0.55 to 0.25, yielding a mean similarity score of -0.176. This negative correlation confirms that the mechanics-driven kinetic prior stream and the data-driven attention stream extract orthogonal, highly divergent feature representations. Under heavy background noise, forcing a 1:1 static equal addition ($h_{phy} + h_{att}$, as in w/o Gating) induces severe vector cancellation and semantic conflicts, degrading joint representation capacity. Furthermore, Figure 8b demonstrates the dynamic regulation mechanism executed by the learnable adaptive gating unit across diverse health states. Rather than adhering to the static 0.5 equal-fusion baseline (rep-

resented by the dotted line), the dynamic gate weight λ spans flexibly from 0.0 to 0.95 across distinct fault categories. When semantic conflicts intensify ($S_{\cos} < 0$), the gating unit dynamically recalibrates stream contributions, prioritizing the physical stream for complex fault categories while favoring the attention stream for localized noise suppression. This adaptive weighting prevents mutual feature corruption, confirming the necessity of the gated fusion mechanism in the proposed KT-GNN-CBAM framework.

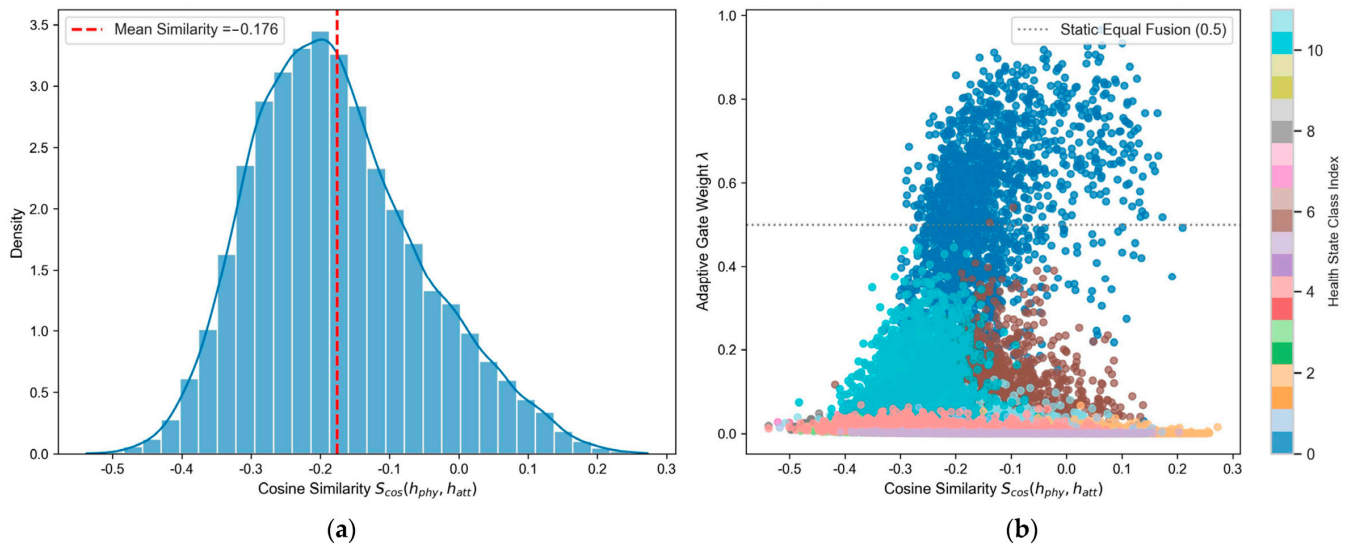


Figure 8. Quantitative evaluation of cross-stream feature interactions: (a) Cosine similarity distribution between streams; (b) Adaptive gate weight response vs. feature divergence.

5. Conclusions

This study presents a comprehensive performance evaluation of the physics-prior topology-guided graph neural network framework (KT-GNN-CBAM) for multi-stage gearbox fault diagnosis under variable operating conditions and structural noise interferences. The primary research insights and structural conclusions are summarized as follows:

(1) **Diagnostic Efficacy of the Dual-Stream Architecture:** Multi-model comparative evaluations demonstrate that the proposed KT-GNN-CBAM framework yields a classification accuracy of 98.87% under a strict 6 dB noise constraint, outperforming baseline DNN, 1D-CNN, and ResNet-1D architectures. Integrating a physical graph stream derived directly from the analytical dynamic stiffness matrix suppresses stochastic noise propagation and cross-domain feature overfitting, resolving the characteristic stability boundaries encountered by purely data-driven network backbones.

(2) **Quantitative Sensitivity of Component Streams:** Component-stripping ablation testing maps the exact classification dependencies of individual structural modules. The mechanics-driven graph pathway serves as the primary regularizer within the latent space; its removal triggers a 282.30% relative surge in the architectural misclassification rate. Furthermore, the adaptive gating unit functions as a soft-weight filter that dynamically resolves semantic representation conflicts, eliminating feature interference between deterministic physical priors and data-driven dynamic attention maps under non-stationary vibration profiles.

(3) **Manifold Evolution and Inductive Biases:** High-dimensional feature space visualization via the t-SNE algorithm confirms optimal intra-class cluster compactness and explicit inter-class separation boundaries generated by the complete dual-stream network. The structural boundary constraints governing gear meshing properties act as a robust physical inductive bias, aligning the hidden representation manifolds with the actual kinematic distribution of failures to maximize cross-temporal threshold resilience.

Despite its high diagnostic accuracy and noise robustness, the proposed framework has three main limitations. First, constructing the kinetic adjacency matrix requires prior analytical knowledge of gear mesh and bearing support stiffness, which can be challenging to obtain for uncalibrated mechanical systems. Second, the current experimental validation was conducted on a single back-to-back closed-loop test rig (HP-GBS-2023). When migrating to authentic industrial port crane scenarios, practical challenges may arise from non-ideal sensor installation/mounting constraints, severe transient load shocks during heavy hoisting operations, and complex multi-path noise interference across extended transmission chains. Third, while the dual-stream graph architecture adds minimal inference latency, it increases memory overhead during edge-node training compared to standard 1D-CNNs.

Future research will address these constraints by: (1) estimating stiffness parameters automatically from dynamic data via physics-informed neural networks (PINNs); (2) integrating domain adaptation and dynamic load-shock resilience mechanisms to bridge the gap between lab testbeds and field port crane deployments; and (3) compressing the model via quantization and distillation for lightweight deployment on embedded edge controllers in port cranes.

In summary, the deep integration of mechanical invariants alongside data-adaptive attention modules mitigates the interpretability bottlenecks and generalization constraints characteristic of conventional grid-based deep learning diagnostics. The developed methodology establishes a verified algorithmic foundation and scalable technical framework for advanced intelligent fault isolation and prognostics and health management (PHM) deployments in complex multi-stage industrial gearbox systems.

Author Contributions: Conceptualization, J.Y. and H.X.; methodology, J.Y.; software, J.Y.; validation, J.Y., M.L. and D.Z.; formal analysis, J.Y.; investigation, J.Y.; resources, H.X.; data curation, J.Y.; writing—original draft preparation, J.Y.; writing—review and editing, J.Y.; visualization, J.Y.; supervision, M.L.; project administration, H.X.; funding acquisition, H.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to privacy concerns.

Acknowledgments: During the preparation of this manuscript, the authors used Gemini 3.6 Flash for the purposes of language translation and polishing. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: Author Dinglong Zhu was employed by the company Sinopec Petroleum Engineering Jiangnan Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Yu, D.; Cheng, J.; Yang, Y. Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings. *Mech. Syst. Signal Process.* **2005**, *19*, 259–272. [[CrossRef](#)]
2. Lei, Y.; Lin, J.; He, Z.; Zuo, M.J. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* **2013**, *35*, 108–126. [[CrossRef](#)]
3. Imaouchen, Y.; Kedadouché, M.; Alkama, R.; Thomas, M. A frequency-weighted energy operator and complementary ensemble empirical mode decomposition for bearing fault detection. *Mech. Syst. Signal Process.* **2017**, *82*, 103–116. [[CrossRef](#)]
4. Chen, J.; Zhou, D.; Lyu, C.; Lu, C. An integrated method based on CEEMD-SampEn and the correlation analysis algorithm for the fault diagnosis of a gearbox under different working conditions. *Mech. Syst. Signal Process.* **2018**, *113*, 102–111. [[CrossRef](#)]

5. Frei, M.G.; Osorio, I. Intrinsic time-scale decomposition: Time–frequency–energy analysis and real-time filtering of non-stationary signals. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2007**, *463*, 321–342.
6. Feng, Z.; Lin, X.; Zuo, M.J. Joint amplitude and frequency demodulation analysis based on intrinsic time-scale decomposition for planetary gearbox fault diagnosis. *Mech. Syst. Signal Process.* **2016**, *72*, 223–240. [[CrossRef](#)]
7. Wang, L.; Liu, Z. An improved local characteristic-scale decomposition to restrict end effects, mode mixing and its application to extract incipient bearing fault signal. *Mech. Syst. Signal Process.* **2021**, *156*, 107657. [[CrossRef](#)]
8. Fan, X.; Zuo, M.J. Gearbox fault detection using Hilbert and wavelet packet transform. *Mech. Syst. Signal Process.* **2006**, *20*, 966–982. [[CrossRef](#)]
9. Yan, R.; Gao, R.X.; Chen, X. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Process.* **2014**, *96*, 1–15. [[CrossRef](#)]
10. Zheng, J.; Pan, H.; Yang, S.; Cheng, J. Adaptive parameterless empirical wavelet transform based time-frequency analysis method and its application to rotor rubbing fault diagnosis. *Signal Process.* **2017**, *130*, 305–314. [[CrossRef](#)]
11. Wang, D.; Zhao, Y.; Yi, C.; Tsui, K.-L.; Lin, J. Sparsity guided empirical wavelet transform for fault diagnosis of rolling element bearings. *Mech. Syst. Signal Process.* **2018**, *101*, 292–308. [[CrossRef](#)]
12. Yu, K.; Lin, T.R.; Tan, J.; Ma, H. An adaptive sensitive frequency band selection method for empirical wavelet transform and its application in bearing fault diagnosis. *Measurement* **2019**, *134*, 375–384. [[CrossRef](#)]
13. Teng, W.; Ding, X.; Cheng, H.; Han, C.; Liu, Y.; Mu, H. Compound faults diagnosis and analysis for a wind turbine gearbox via a novel vibration model and empirical wavelet transform. *Renew. Energy* **2019**, *136*, 393–402. [[CrossRef](#)]
14. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2013**, *62*, 531–544.
15. Wang, X.B.; Yang, Z.X.; Yan, X.A. Novel particle swarm optimization-based variational mode decomposition method for the fault diagnosis of complex rotating machinery. *IEEE/ASME Trans. Mechatron.* **2017**, *23*, 68–79.
16. Lin, Y.; Xiao, M.; Liu, H.; Li, Z.; Zhou, S.; Xu, X.; Wang, D. Gear fault diagnosis based on CS-improved variational mode decomposition and probabilistic neural network. *Measurement* **2022**, *192*, 110913. [[CrossRef](#)]
17. Zhao, M.; Jia, X. A novel strategy for signal denoising using reweighted SVD and its applications to weak fault feature enhancement of rotating machinery. *Mech. Syst. Signal Process.* **2017**, *94*, 129–147. [[CrossRef](#)]
18. Cong, F.; Zhong, W.; Tong, S.; Tang, N.; Chen, J. Research of singular value decomposition based on slip matrix for rolling bearing fault diagnosis. *J. Sound. Vib.* **2015**, *344*, 447–463. [[CrossRef](#)]
19. Daubechies, I.; Lu, J.; Wu, H.T. Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Appl. Comput. Harmon. Anal.* **2011**, *30*, 243–261. [[CrossRef](#)]
20. Hazra, B.; Sadhu, A.; Narasimhan, S. Fault detection of gearboxes using synchro-squeezing transform. *J. Vib. Control* **2017**, *23*, 3108–3127.
21. Yu, G.; Wang, Z.; Zhao, P. Multisynchrosqueezing transform. *IEEE Trans. Ind. Electron.* **2018**, *66*, 5441–5455.
22. Bao, W.; Li, F.; Tu, X.; Hu, Y.; He, Z. Second-order synchroextracting transform with application to fault diagnosis. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 1–9. [[CrossRef](#)]
23. Liu, Q.; Wang, Y.; Xu, Y. Synchrosqueezing extracting transform and its application in bearing fault diagnosis under non-stationary conditions. *Measurement* **2021**, *173*, 108569. [[CrossRef](#)]
24. Xu, Y.; Wang, L.; Yu, G.; Wang, Y. Generalized S-synchroextracting transform for fault diagnosis in rolling bearing. *IEEE Trans. Instrum. Meas.* **2021**, *71*, 1–14. [[CrossRef](#)]
25. Lv, Y.; Yuan, R.; Li, H. Longitudinal synchroextracting transform: A useful tool for characterizing signals with strong frequency modulation and application to machine fault diagnosis. *Measurement* **2021**, *182*, 109750. [[CrossRef](#)]
26. Yu, G. A multisynchrosqueezing-based high-resolution time-frequency analysis tool for the analysis of non-stationary signals. *J. Sound. Vib.* **2021**, *492*, 115813. [[CrossRef](#)]
27. Niola, V.; Savino, S.; Quaremba, G.; Cosenza, C.; Nicolella, A.; Spirto, M. Discriminant Analysis of the Vibrational Behavior of a Gas Micro-Turbine as a Function of Fuel. *Machines* **2022**, *10*, 925. [[CrossRef](#)]
28. Marticorena, M.; García Peyrano, O. Rolling Bearing Condition Monitoring Technique Based on Cage Rotation Analysis and Acoustic Emission. *J. Dyn. Monit. Diagn.* **2022**, *1*, 57–65. [[CrossRef](#)]
29. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
30. Hinton, G.E.; Osindero, S.; Teh, Y.W. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [[CrossRef](#)] [[PubMed](#)]
31. Shao, H.; Jiang, H.; Lin, Y.; Li, X. A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders. *Mech. Syst. Signal Process.* **2018**, *102*, 278–297. [[CrossRef](#)]
32. Wang, D.; Hong, R.; Lin, X. A method for predicting hobbing tool wear based on CNC real-time monitoring data and deep learning. *Precis. Eng.* **2021**, *72*, 847–857. [[CrossRef](#)]
33. Chen, T.; Wang, Z.; Yang, X.; Jiang, K. A deep capsule neural network with stochastic delta rule for bearing fault diagnosis on raw vibration signals. *Measurement* **2019**, *148*, 106857. [[CrossRef](#)]

34. Li, X.; Li, J.; Zhao, C.; Qu, Y.; He, D. Gear pitting fault diagnosis with mixed operating conditions based on adaptive 1D separable convolution with residual connection. *Mech. Syst. Signal Process.* **2020**, *142*, 106740. [\[CrossRef\]](#)
35. Choudhary, A.; Mishra, R.K.; Fatima, S.; Panigrahi, B. Multi-input CNN based vibro-acoustic fusion for accurate fault diagnosis of induction motor. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105872. [\[CrossRef\]](#)
36. Tong, J.; Liu, C.; Bao, J.; Pan, H.; Zheng, J. A novel ensemble learning-based multisensor information fusion method for rolling bearing fault diagnosis. *IEEE Trans. Instrum. Meas.* **2022**, *72*, 1–12. [\[CrossRef\]](#)
37. Yang, L.; Ma, Y.; Zeng, F.; Peng, X.; Liu, D. Improved deep learning based telemetry data anomaly detection to enhance spacecraft operation reliability. *Microelectron. Reliab.* **2021**, *126*, 114311. [\[CrossRef\]](#)
38. Ravikumar, K.N.; Yadav, A.; Kumar, H.; Gangadharan, K.; Narasimhadhan, A. Gearbox fault diagnosis based on Multi-Scale deep residual learning and stacked LSTM model. *Measurement* **2021**, *186*, 110099. [\[CrossRef\]](#)
39. Sun, H.; Zhao, S. Fault diagnosis for bearing based on 1DCNN and LSTM. *Shock. Vib.* **2021**, *2021*, 1221462. [\[CrossRef\]](#)
40. Gu, K.; Zhang, Y.; Liu, X.; Li, H.; Ren, M. DWT-LSTM-based fault diagnosis of rolling bearings with multi-sensors. *Electronics* **2021**, *10*, 2076. [\[CrossRef\]](#)
41. Shi, J.; Peng, D.; Peng, Z.; Zhang, Z.; Goebel, K.; Wu, D. Planetary gearbox fault diagnosis using bidirectional-convolutional LSTM networks. *Mech. Syst. Signal Process.* **2022**, *162*, 107996. [\[CrossRef\]](#)
42. Jin, Y.; Hou, L.; Chen, Y. A time series transformer based method for the rotating machinery fault diagnosis. *Neurocomputing* **2022**, *494*, 379–395. [\[CrossRef\]](#)
43. Zhao, Y.; Zhang, X.; Wang, J.; Wu, L.; Liu, Z.; Wang, L. A new data fusion driven-sparse representation learning method for bearing intelligent diagnosis in small and unbalanced samples. *Eng. Appl. Artif. Intell.* **2023**, *117*, 105513. [\[CrossRef\]](#)
44. Li, X.; Cheng, J.; Shao, H.; Liu, K.; Cai, B. A fusion CWSMM-based framework for rotating machinery fault diagnosis under strong interference and imbalanced case. *IEEE Trans. Ind. Inform.* **2021**, *18*, 5180–5189. [\[CrossRef\]](#)
45. Wang, Y.; Jin, Q.; Sun, G.; Sun, C.F. Planetary gearbox fault feature learning using conditional variational neural networks under noise environment. *Knowl. Based Syst.* **2019**, *163*, 438–449. [\[CrossRef\]](#)

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.