

Article

A Novel Adaptive Correlation Spatio-Temporal-Frequency Graph Convolutional Framework for Multi-Source Information Fusion-Based Fault Diagnosis of Wind Turbines

Rongrong Peng 

School of Mechanical and Vehicle Engineering, Nanchang Institute of Science and Technology,
Nanchang 330108, China; 15294476178@163.com

Abstract

Wind-turbine fault diagnosis is limited by incomplete fault information in single-sensor signals and the poor adaptability of fixed graph structures to non-stationary operating responses. This study proposes an adaptive correlation spatio-temporal-frequency graph convolutional network (AC-STFGCN) to address these limitations in multi-source wind-turbine fault diagnosis. One branch uses one-dimensional convolutions to extract local time-domain features, while the other learns time–frequency representations based on the short-time Fourier transform (STFT). Vibration and acoustic features from these branches are fused and encoded using a bidirectional long short-term memory (Bi-LSTM) network. Local temporal segments are then represented as individual graph nodes. Feature correlations between nodes determine an adaptive, sample-specific adjacency matrix that captures non-Euclidean dependencies across temporal segments and multi-domain features. PoolGAT combines hierarchical graph pooling and graph attention to emphasize fault-relevant nodes and suppress redundant information. Experiments used multi-source signals collected from a wind-turbine drivetrain fault-simulation test rig. AC-STFGCN achieved 99.08% accuracy, 99.12% precision, 99.03% recall, and an F1-score of 99.07%. Confusion matrices, receiver operating characteristic (ROC) curves, and t-distributed stochastic neighbor embedding (t-SNE) visualizations also showed improved discrimination between similar fault patterns. Further experiments at load levels of 0%, 25%, 50%, 75%, and 100% assessed the sensitivity of AC-STFGCN to load-dependent signal variations. All four diagnostic metrics remained above 98% across the evaluated load range.

Keywords: wind turbine; fault diagnosis; multi-source information fusion; graph convolutional network; adaptive correlation graph; spatio-temporal-frequency features



Academic Editors: Xuefang Xu, Zijian Qiao, Mengdi Li and Peng Mei

Received: 10 August 2026

Revised: 14 September 2026

Accepted: 17 September 2026

Published: 24 September 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.

1. Introduction

The working condition of wind turbines, central components of modern wind-power systems, directly affects power-generation efficiency, maintenance costs, and operational reliability [1–5].

During service, drivetrain components experience varying mechanical loads and complex vibration excitations [6]. Bearings, gearboxes, and related transmission components can develop wear, localized damage, and misalignment, which may degrade performance and cause unplanned downtime [7,8]. Reliable condition monitoring and fault diagnosis are therefore needed to support wind-turbine operation and maintenance [9–12].

Traditional diagnostic methods typically extract fault-sensitive indicators from measured signals and use them as inputs to machine learning (ML)-based classifiers. Studies

have examined acoustic-emission features for early bearing damage identification and integrated spectral kurtosis for gearbox fault diagnosis [13–16]. These methods rely on signal characteristics associated with specific damage mechanisms, and their effectiveness depends on feature selection and acquisition conditions. Changes in load, rotational speed, and transmission paths can alter the measured responses. This variability motivates models that learn complementary representations directly from monitoring data [17–20].

Deep learning (DL)-based architectures address different aspects of feature learning for fault diagnosis [21,22]. Convolutional neural networks (CNNs) extract local patterns from signal sequences or time–frequency representations. Multi-scale feature extraction and attention can emphasize diagnostic information at different resolutions. Liang et al. combined multi-scale feature extraction with content-guided attention for bearing fault diagnosis [1]. For wind-turbine misalignment diagnosis, Zheng et al. proposed the lightweight MS-DSCNN-AttMPLSTM framework, which combines convolutional feature extraction with attention-enhanced recurrent modeling [4]. These studies show how convolutional and recurrent structures complement each other by extracting local response patterns and incorporating temporal context, respectively [23,24].

Transformer-based methods model dependencies between distant feature positions and interactions among information sources. Li et al. proposed MSIF-Convformer for multi-source fault diagnosis under strong noise [25]. Its parallel CNN feature extractors encode vibration and acoustic time–frequency representations. A CNN–Transformer architecture captures local and global dependencies, while cross-attention supports information fusion. This work establishes multi-source time–frequency learning and attention-based fusion as effective architectural foundations. When relationships among fused local representations are explicitly expressed as a graph, their organization and aggregation require further consideration [25,26].

Graph neural networks use connectivity-dependent feature propagation for fault diagnosis [27]. Yu et al. combined CNN and graph convolutional network features through weighted fusion for mechanical fault diagnosis under time-varying speeds [28]. Their subsequent I²SGCN framework addressed cross-domain diagnosis using multi-source sensing, sensor-importance estimation, feature-based and importance-based graphs, and subgraph learning [29]. These studies show that the diagnostic role of graphs depends on their node representations, connectivity, and learning objective. Sensor reliability has also informed research on adaptive multi-source fusion. In MSCAF, Yu et al. used uncertainty quantification and sensor-confidence assessment to adapt fusion weights. These weights account for differences in signal quality and fault-information consistency across sensors [16].

This study proposes an adaptive correlation spatio-temporal-frequency graph convolutional network (AC-STFGCN) for wind-turbine drivetrain fault diagnosis. MSCAF uses sensor-confidence weighting, whereas I²SGCN employs cross-domain graph learning. AC-STFGCN constructs sample-specific temporal-segment graphs and processes them hierarchically. Its main contributions concern node representation, adaptive connectivity, and pooling–attention processing, as detailed below.

- (1) A joint temporal-segment representation integrates complementary vibration and acoustic features. Parallel time-domain and STFT-based branches extract features that are then fused and temporally encoded by a Bi-LSTM network. Local temporal segments form the graph nodes, each combining multi-source information with temporal context for subsequent graph construction.
- (2) An adaptive correlation graph assigns sample-specific edge weights based on positive feature interactions between nodes within a prescribed temporal lag range. Temporally encoded multi-source features determine the connectivity, while the lag constraint

defines the temporal neighborhood. Together, these components allow graph relationships to adapt to each input sample.

- (3) PoolGAT combines learned pooling with graph attention for hierarchical graph aggregation. Pooling forms coarser graph representations, and attention then reweights neighboring contributions within each pooled graph. This sequence supports selective information propagation between aggregated feature units and is designed to reduce redundancy while preserving fault-relevant relationships.

The evaluation covers seven drivetrain health conditions, with comparative experiments conducted at 50% of the rated load. Additional experiments assess the framework at 0%, 25%, 75%, and 100% of the rated load.

Section 2 introduces the theoretical foundations of GCNs and Bi-LSTMs. Section 3 presents the AC-STFGCN framework and its optimization objective. Section 4 describes the experimental setup, dataset preparation, and comparative results. In Section 5, a discussion of the proposed method is carried out. Section 6 concludes the paper and outlines directions for future work.

2. Related Work

2.1. Graph Neural Network

Graph convolutional networks (GCNs) provide a useful framework for modeling data whose relationships are not naturally arranged on regular Euclidean grids [30]. Unlike conventional convolutional neural networks, which aggregate information through fixed local kernels, GCNs propagate node features according to a graph topology. This property is useful for multi-source wind-turbine fault diagnosis because dependencies among sensors, temporal segments and frequency-domain components are often irregular and sample-dependent. Fault-induced impacts, transmission-path effects and multi-source coupling can create correlations that are difficult to represent using sequence- or image-based structures alone. Therefore, GCNs can be used to model non-Euclidean relationships among fault-sensitive features, as shown in Figure 1.

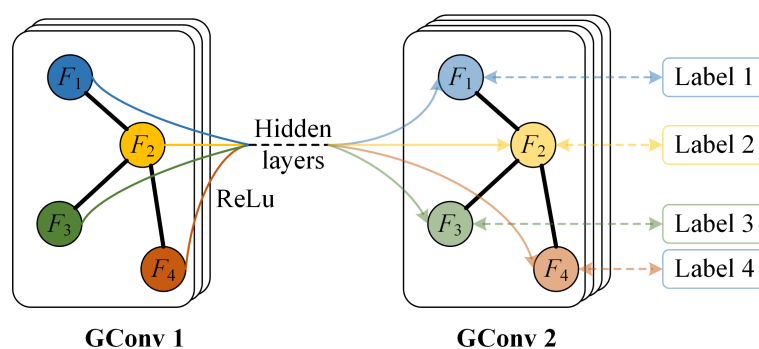


Figure 1. Standard structure of a GCN.

Let a graph be denoted by $\mathcal{G} = (\mathbf{V}, \mathbf{E}, \mathbf{A})$, where $\mathbf{V}$ is the node set, $\mathbf{E}$ is the edge set, $\mathbf{A} \in \mathbb{R}^{N \times N}$ is the adjacency matrix, and N is the number of nodes. The node feature matrix is defined as

$$\mathbf{X} \in \mathbb{R}^{N \times F}, \quad (1)$$

where F is the dimensionality of each node feature. The propagation rule of a standard GCN can be written as

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

with

$$\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}, \quad (3)$$

$$\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}, \quad (4)$$

where $\mathbf{H}^{(l)}$ denotes the node representation at the l -th layer, $\mathbf{H}^{(0)} = \mathbf{X}$, $\mathbf{W}^{(l)}$ is a learnable weight matrix, $\sigma(\cdot)$ is a nonlinear activation function, and $\mathbf{I}$ is the identity matrix. Through this propagation process, each node aggregates information from its neighbors and updates its representation according to both local features and graph-structured dependencies.

2.2. Bidirectional Long Short-Term Memory

Long short-term memory (LSTM) networks are recurrent neural networks designed to model temporal dependencies in sequential data. By introducing input, forget and output gates, LSTMs alleviate the vanishing-gradient problem of conventional recurrent neural networks and can capture long-range temporal dependencies. Wind-turbine monitoring signals often contain temporally evolving fault information, including impulsive responses, rotational-frequency modulation and gear-meshing components. Therefore, LSTMs are useful for extracting temporal-context features from non-stationary multi-source sensor signals. For an input sequence $\mathbf{z}_t$, the LSTM computation at time step t is given by

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \mathbf{z}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i), \quad (5)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \mathbf{z}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f), \quad (6)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o \mathbf{z}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o), \quad (7)$$

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c \mathbf{z}_t + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c), \quad (8)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t, \quad (9)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t), \quad (10)$$

where $\mathbf{i}_t$, $\mathbf{f}_t$ and $\mathbf{o}_t$ denote the input gate, forget gate and output gate, respectively. $\mathbf{c}_t$ is the cell state, $\mathbf{h}_t$ is the hidden state, $\sigma(\cdot)$ is the sigmoid activation function, and $\odot$ denotes the Hadamard product.

Conventional LSTMs process sequential information only in the forward temporal direction, which limits their ability to use both preceding and subsequent context at each time step. A bidirectional long short-term memory (Bi-LSTM) network addresses this limitation by using two independent LSTM layers to model the sequence in opposite temporal directions, as shown in Figure 2. The forward LSTM captures historical dependencies, whereas the backward LSTM captures subsequent contextual information. The final representation is obtained by concatenating the hidden states from both directions,

$$\mathbf{h}_t^{\text{bi}} = [\vec{\mathbf{h}}_t \parallel \overleftarrow{\mathbf{h}}_t], \quad (11)$$

where $\vec{\mathbf{h}}_t$ and $\overleftarrow{\mathbf{h}}_t$ denote the forward and backward hidden states, respectively, and $\parallel$ represents vector concatenation. Compared with unidirectional LSTMs, Bi-LSTMs provide a more complete temporal-context representation of non-stationary wind-turbine signals.

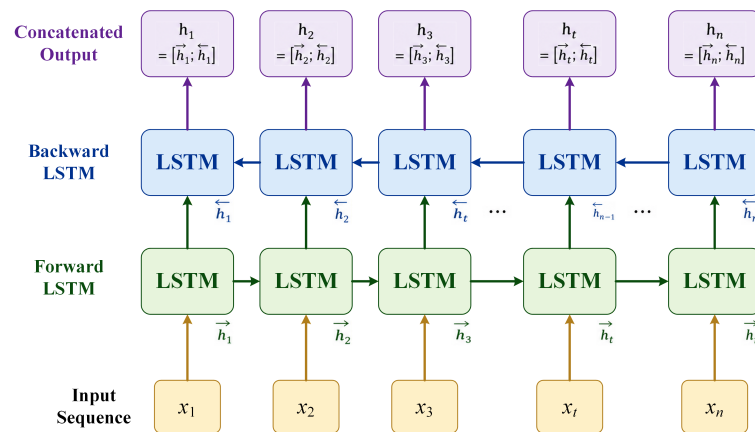


Figure 2. Standard structure of a Bi-LSTM.

3. Proposed Method

The overall architecture of the proposed AC-STFGCN is shown in Figure 3.

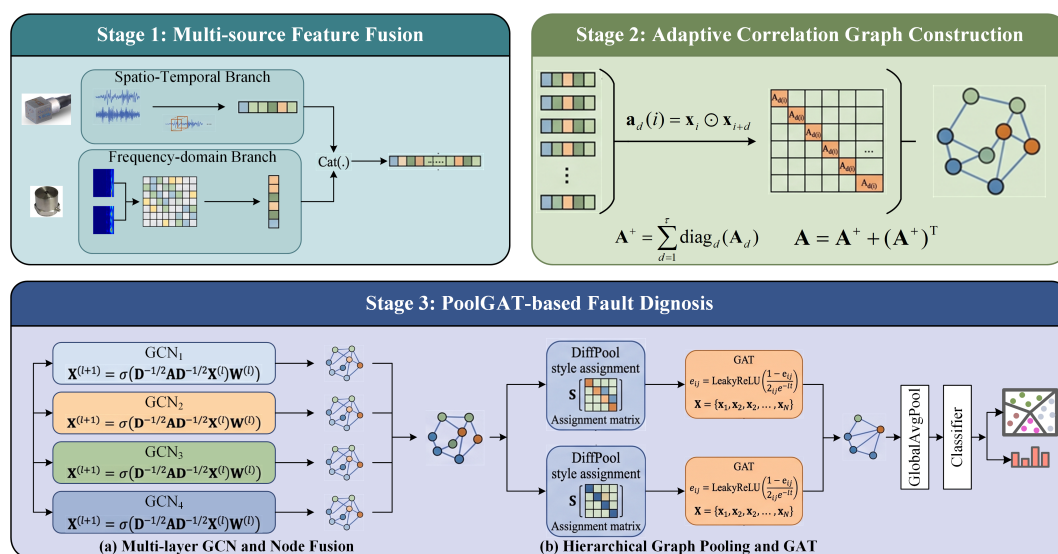


Figure 3. Overall architecture of the proposed AC-STFGCN.

3.1. Spatio-Temporal-Frequency Multi-Domain Feature Fusion

Multi-source wind-turbine signals contain local time-domain impulses, frequency-domain modulation patterns and cross-sensor coupling information. To extract these complementary features, AC-STFGCN uses a time-domain branch, a frequency-domain branch and a temporal-context encoding module. The time-domain branch processes vibration and acoustic signals as separate input channels and extracts local response features through batch normalization and one-dimensional convolution:

$$\mathbf{X}_{\text{TD}} = \text{Conv1D}[\text{BN}(\hat{\mathbf{x}})]. \quad (12)$$

Here, where $\hat{\mathbf{x}}$ denotes the normalized multi-source input signal, and $\text{BN}(\cdot)$ denotes batch normalization. Given an input feature u , batch normalization is defined as

$$\text{BN}(u) = \gamma_{\text{BN}} \frac{u - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} + \beta_{\text{BN}}, \quad (13)$$

where μ_B and σ_B^2 are the mean and variance of the corresponding channel features in the current batch, γ_{BN} and β_{BN} are learnable scaling and shifting parameters, respectively, and ε is a numerical stability term. The one-dimensional convolution is formulated as

$$y_o(t) = b_o + \sum_{c=1}^{C_{in}} \sum_{r=0}^{K_1-1} w_{o,c,r} u_c(t+r), \quad (14)$$

where $u_c(t)$ is the feature of the c -th input channel at position t , $y_o(t)$ is the feature of the o -th output channel, $w_{o,c,r}$ is the convolution kernel weight, and b_o is the bias term. C_{in} is the number of input channels, and K_1 is the one-dimensional convolution kernel size. In this study, the kernel size, stride and padding of the one-dimensional convolution are set to 7, 32 and 3, respectively.

The frequency-domain branch applies the short-time Fourier transform (STFT) using a finite-length Blackman window to obtain time–frequency representations. For channel $c \in \{\text{Vib}, \text{Aco}\}$, the STFT is defined as

$$X^c(m, k) = \sum_{n=0}^{N_{fft}-1} \hat{x}^c(n + mH) w(n) \exp\left(-j \frac{2\pi kn}{N_{fft}}\right), \quad (15)$$

where m denotes the frame index, n is the local sample index within each frame, k denotes the frequency-bin index, and H is the hop size. The Blackman window $w(n)$ is defined for $0 \leq n \leq N_{fft} - 1$ and is zero elsewhere. The STFT provides local spectral information that complements the time-domain convolutional features. Its uniform time–frequency representation supports convolutional processing of both vibration and acoustic signals. The analysis uses a 128-sample Blackman window, a 128-point FFT, and a hop size of 32 samples. At a sampling frequency of 12,800 Hz, these settings give a window duration of 10 ms, a frame interval of 2.5 ms, and 75% overlap. The nominal frequency-bin spacing is 100 Hz, but the effective frequency resolution also depends on the spectral characteristics of the Blackman window. Specifically, this STFT configuration is therefore intended for learning local spectral patterns rather than achieving optimal time–frequency resolution. The Blackman window is expressed as

$$w(n) = 0.42 - 0.5 \cos\left(\frac{2\pi n}{N_{fft} - 1}\right) + 0.08 \cos\left(\frac{4\pi n}{N_{fft} - 1}\right). \quad (16)$$

In this study, N_{fft} and H are set to 128 and 32, respectively. To reduce the dynamic range of the spectral magnitude, the logarithmic magnitude spectrum is used:

$$S^c(m, k) = \log(1 + |X^c(m, k)|). \quad (17)$$

The spectrograms of the vibration and acoustic signals are represented as

$$\mathbf{S} \in \mathbb{R}^{2 \times F \times T_f}, \quad (18)$$

where $F = N_{fft}/2 + 1$ is the number of frequency bins, and T_f is the number of time frames after STFT. The frequency-domain features are then compressed using two-dimensional convolution:

$$\mathbf{X}_{FD} = \text{Squeeze}\{\text{Conv2D}[\text{BN}(\mathbf{S})]\}. \quad (19)$$

The two-dimensional convolution is defined as

$$y_o(p, q) = b_o + \sum_{c=1}^{C_{in}} \sum_{r=0}^{K_h-1} \sum_{s=0}^{K_w-1} w_{o,c,r,s} u_c(p+r, q+s), \quad (20)$$

where p and q are two-dimensional indices, and K_h and K_w are the height and width of the convolution kernel. In this study, the two-dimensional convolution kernel size is set

to $(K_h, K_w) = (F, 1)$. This operation compresses frequency-domain information along the frequency dimension and preserves the temporal-frame dimension. $\text{Squeeze}(\cdot)$ denotes the operation that removes the frequency dimension with length 1.

The time-domain feature $\mathbf{X}_{\text{TD}}$ and frequency-domain feature $\mathbf{X}_{\text{FD}}$ are concatenated along the channel dimension:

$$\mathbf{X}_{\text{Cat}} = \text{Cat}(\mathbf{X}_{\text{TD}}, \mathbf{X}_{\text{FD}}). \quad (21)$$

where $\text{Cat}(\cdot)$ denotes channel-wise concatenation. The fusion module uses two 1×1 one-dimensional convolutional layers:

$$\mathbf{X}_{\text{SF}} = \text{Conv1D}_2\{\text{GELU}[\text{BN}(\text{Conv1D}_1(\mathbf{X}_{\text{Cat}}))]\}. \quad (22)$$

Here, $\text{GELU}(\cdot)$ denotes the Gaussian error linear unit activation function. To model temporal contextual dependencies, the fused feature sequence is fed into a bidirectional long short-term memory network:

$$\mathbf{H} = \text{BiLSTM}(\mathbf{X}_{\text{SF}}^T). \quad (23)$$

The Bi-LSTM output at time step t is obtained by concatenating the forward and backward hidden states:

$$\mathbf{h}_t^{\text{bi}} = [\vec{\mathbf{h}}_t \parallel \overleftarrow{\mathbf{h}}_t]. \quad (24)$$

Graph node features are then obtained through max pooling:

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \in \mathbb{R}^{N \times F_g}. \quad (25)$$

where F_g represents the dimensionality of each node feature, and N denotes the total number of graph nodes. Each node corresponds to a local temporal segment that integrates vibration, acoustic, frequency-domain and temporal-context information.

3.2. Adaptive Correlation Graph Construction

A fixed adjacency matrix may not adapt well to non-stationary wind-turbine fault signals. Therefore, AC-STFGCN constructs a sample-specific graph structure from adaptive feature correlations among graph nodes. Given the node feature matrix $\mathbf{X} \in \mathbb{R}^{N \times F_g}$, the i -th node feature vector is defined as $\mathbf{x}_i = [x_i(1), x_i(2), \dots, x_i(F_g)]$. For two nodes separated by interval d , the dimension-wise correlation feature is computed as

$$\mathbf{a}_d(i) = \mathbf{x}_i \odot \mathbf{x}_{i+d}, \quad i = 1, 2, \dots, N - d. \quad (26)$$

The r -th dimension of this feature is

$$a_d(i, r) = x_i(r)x_{i+d}(r). \quad (27)$$

The maximum response along the feature dimension is selected, and ReLU activation is applied to retain positive correlations:

$$A_d(i) = \text{ReLU}\left[\max_{1 \leq r \leq F_g} a_d(i, r)\right]. \quad (28)$$

Next, A_d is embedded into the d -th upper diagonal of the adjacency matrix:

$$\mathbf{A}^+ = \sum_{d=1}^{\tau} \text{diag}_d(\mathbf{A}_d), \quad (29)$$

where $\text{diag}_d(\mathbf{A}_d)$ places vector $\mathbf{A}_d$ on the d -th upper diagonal. τ is the maximum correlation lag order, which is set to 11 in this study. The symmetric adjacency matrix is then constructed as

$$\mathbf{A} = \mathbf{A}^+ + (\mathbf{A}^+)^T. \quad (30)$$

Finally, amplitude normalization is applied:

$$\mathbf{A} \leftarrow \frac{\mathbf{A}}{\max(\mathbf{A}) + \varepsilon}. \quad (31)$$

Herein, the adaptive correlation strategy refers to a sample-dependent feature-interaction score computed directly from learned features. For nodes i and j , it is defined as follows:

$$s_{ij} = \max\left\{0, \max_r [x_i(r)x_j(r)]\right\}. \quad (32)$$

The score captures the strongest same-sign interaction across feature coordinates. Taking the maximum avoids cancellation across dimensions and encodes the assumption that a dominant shared feature response can support an edge. Connectivity is restricted to the prescribed temporal neighborhood and recomputed for each input sample. The positive scalars a and b can be defined as follows:

$$s(ax_i, bx_j) = ab s(x_i, x_j). \quad (33)$$

The cosine similarity technique normalizes vector magnitudes, and Pearson correlation also centers the feature vectors. The proposed AC-STFGCN instead uses the dominant component-wise product to define affinity, without computing a statistical correlation coefficient. Graph-level normalization does not remove sensitivity to node-specific or dimension-specific amplitude changes, and unusually large feature components can dominate the score. Empirical comparisons with alternative affinity measures are therefore needed to establish its relative effectiveness. Therefore, the adaptive correlation graph is obtained as follows:

$$\mathcal{G} = (\mathbf{X}, \mathbf{A}), \quad (34)$$

where $\mathbf{A}$ means the adaptive adjacency matrix generated from the node correlations in samples, and $\mathbf{X}$ denotes the graph node feature matrix.

3.3. PoolGAT-Based Graph Feature Extraction for Fault Diagnosis

In the graph convolutional layer, self-loops are first added to the adjacency matrix:

$$\hat{\mathbf{A}} = \mathbf{A} + \mathbf{I}, \quad (35)$$

where $\mathbf{I}$ denotes the identity matrix. Meanwhile, the degree matrix $\hat{\mathbf{D}}$ can be expressed as follows:

$$\hat{\mathbf{D}}_{ii} = \sum_j \hat{A}_{ij}. \quad (36)$$

The graph convolution propagation process is formulated as

$$\mathbf{X}^{(l+1)} = \sigma\left(\hat{\mathbf{D}}^{-1/2} \hat{\mathbf{A}} \hat{\mathbf{D}}^{-1/2} \mathbf{X}^{(l)} \mathbf{W}^{(l)}\right), \quad (37)$$

where $\mathbf{X}^{(l)}$ denotes the input node features at the l -th layer, $\mathbf{W}^{(l)}$ is the learnable weight matrix, and $\sigma(\cdot)$ is the ReLU activation function. Four graph convolutional layers are stacked, and their outputs are concatenated:

$$\mathbf{Z} = \text{Cat}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \mathbf{X}^{(3)}, \mathbf{X}^{(4)}). \quad (38)$$

To reduce graph redundancy and highlight informative nodes, a DiffPool-style hierarchical graph pooling strategy is adopted. The node assignment matrix is learned through a graph convolutional branch:

$$\mathbf{S}_\psi = \text{GCN}_{\text{assign}}(\mathbf{X}_\psi, \mathbf{A}_\psi), \quad (39)$$

where $\mathbf{S}_\psi \in \mathbb{R}^{N_\psi \times N_{\psi+1}}$ is the node assignment matrix, N_ψ denotes the number of nodes before pooling, and $N_{\psi+1}$ denotes the number of nodes after pooling. The softmax function is used to normalize the assignment probability of each node:

$$S_{ij} = \frac{\exp(q_{ij})}{\sum_{r=1}^{N_{\psi+1}} \exp(q_{ir})}, \quad (40)$$

where q_{ij} is the unnormalized score for assigning node i to the j -th cluster. Node embeddings are learned through another graph convolutional branch:

$$\mathbf{Z}_\psi = \text{GCN}_{\text{embed}}(\mathbf{X}_\psi, \mathbf{A}_\psi). \quad (41)$$

The pooling operation updates the node features and adjacency matrix as follows:

$$\mathbf{X}_{\psi+1} = \mathbf{S}_\psi^T \mathbf{Z}_\psi, \quad (42)$$

$$\mathbf{A}_{\psi+1} = \mathbf{S}_\psi^T \mathbf{A}_\psi \mathbf{S}_\psi. \quad (43)$$

A graph attention mechanism is introduced after pooling to assign different weights to neighboring nodes. For nodes i and j , linear mapping is first performed as

$$\mathbf{h}_i = \mathbf{W}\mathbf{x}_i, \quad \mathbf{h}_j = \mathbf{W}\mathbf{x}_j, \quad (44)$$

where $\mathbf{W}$ is a learnable mapping matrix. The attention score is computed as

$$e_{ij} = \text{LeakyReLU}(\mathbf{a}_s^T \mathbf{h}_i + \mathbf{a}_t^T \mathbf{h}_j), \quad (45)$$

where $\mathbf{a}_s$ and $\mathbf{a}_t$ are learnable attention parameters. The normalized attention coefficient is computed over the connected neighborhood set $\mathcal{N}(i)$:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{r \in \mathcal{N}(i)} \exp(e_{ir})}. \quad (46)$$

The node representation is then updated as

$$\tilde{\mathbf{x}}_i = \sum_{j \in \mathcal{N}(i)} \alpha_{ij} \mathbf{h}_j. \quad (47)$$

After two PoolGAT stages, the global average pooling (GAP) technique is then applied to obtain the final graph representation:

$$\mathbf{g} = \frac{1}{N'} \sum_{i=1}^{N'} \mathbf{x}'_i, \quad (48)$$

where N' is the final number of graph nodes, and $\mathbf{x}'_i$ is the final feature of the i -th node. The output of the fault diagnosis classifier is computed as

$$\mathbf{o} = \mathbf{W}_c \mathbf{g} + \mathbf{b}_c, \quad (49)$$

where $\mathbf{W}_c$ and $\mathbf{b}_c$ are the weight matrix and bias vector of the classifier. The probability of the k -th category is computed as

$$p_k = \frac{\exp(o_k)}{\sum_{q=1}^K \exp(o_q)}, \quad k = 1, 2, \dots, K. \quad (50)$$

The final predicted category is given by

$$\hat{y} = \arg \max_k p_k. \quad (51)$$

3.4. Overall Loss Function

The proposed method uses cross-entropy loss with label smoothing. The classification loss is expressed as

$$\mathcal{L}_{\text{CE}} = - \sum_{k=1}^K y_k^s \log p_k, \quad (52)$$

where y_k^s denotes the smoothed label of the k -th class, and p_k denotes the predicted probability. To regularize hierarchical graph pooling, a link preservation loss is introduced:

$$\mathcal{L}_{\text{link}}^\psi = \text{mean} \left[\left(\mathbf{A}_\psi - \mathbf{S}_\psi \mathbf{S}_\psi^\top \right)^2 \right], \quad (53)$$

where $\mathbf{S}_\psi \mathbf{S}_\psi^\top$ denotes the soft connectivity reconstructed from the node assignment matrix. This loss constrains the consistency of graph connectivity before and after pooling. An assignment entropy loss is also introduced:

$$\mathcal{L}_{\text{ent}}^\psi = - \text{mean} \left[\sum_j S_{ij}^\psi \log (S_{ij}^\psi + \epsilon) \right]. \quad (54)$$

This term regularizes the node assignment probabilities and prevents the assignment matrix from becoming overly diffuse. The auxiliary loss of the ψ -th pooling layer is defined as

$$\mathcal{L}_{\text{aux}}^\psi = 0.1 \mathcal{L}_{\text{link}}^\psi + 0.01 \mathcal{L}_{\text{ent}}^\psi. \quad (55)$$

The final optimization objective is

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda \sum_{\psi} \mathcal{L}_{\text{aux}}^\psi, \quad (56)$$

where λ controls the contribution of the auxiliary pooling losses.

The parameters of AC-STFGCN are updated using the AdamW optimizer:

$$\theta_{t+1} = \theta_t - \alpha_t \frac{\hat{\mathbf{m}}_t}{\sqrt{\hat{\mathbf{v}}_t + \epsilon}} - \alpha_t \omega \theta_t, \quad (57)$$

where θ_t denotes the model parameters at the t -th iteration, α_t denotes the learning rate, $\hat{\mathbf{m}}_t$ and $\hat{\mathbf{v}}_t$ are the bias-corrected estimates of the first-order and second-order moments, and ω denotes the weight decay coefficient. The learning rate is updated using a cosine annealing strategy:

$$\alpha_t = \alpha_{\min} + \frac{1}{2}(\alpha_0 - \alpha_{\min}) \left[1 + \cos\left(\frac{t}{T_{\max}}\pi\right) \right], \quad (58)$$

where α_0 is the initial learning rate, $\alpha_{\min}$ is the minimum learning rate, and $T_{\max}$ is the maximum number of training epochs. The algorithm of the proposed AC-STFGCN is described in Algorithm 1.

Algorithm 1 AC-STFGCN-based fault diagnosis

Require: Vibration signal $\mathbf{x}^{\text{Vib}}$, acoustic signal $\mathbf{x}^{\text{Aco}}$, trained model parameters, maximum correlation lag $\tau = 11$.

Ensure: Predicted health condition $\hat{y}$.

- 1: Normalize and assemble the signals into $\hat{\mathbf{x}}$.
 - 2: $\mathbf{X}_{\text{TD}} \leftarrow \text{Conv1D}(\text{BN}(\hat{\mathbf{x}}))$.
 - 3: $\mathbf{S} \leftarrow \log(1 + |\text{STFT}(\hat{\mathbf{x}})|)$.
 - 4: $\mathbf{X}_{\text{FD}} \leftarrow \text{Squeeze}(\text{Conv2D}(\text{BN}(\mathbf{S})))$.
 - 5: $\mathbf{X}_{\text{Cat}} \leftarrow \text{Cat}(\mathbf{X}_{\text{TD}}, \mathbf{X}_{\text{FD}})$.
 - 6: Fuse $\mathbf{X}_{\text{Cat}}$ using two 1×1 convolutional layers to obtain $\mathbf{X}_{\text{SF}}$.
 - 7: $\mathbf{H} \leftarrow \text{BiLSTM}(\mathbf{X}_{\text{SF}}^{\text{T}})$.
 - 8: Apply temporal max pooling to $\mathbf{H}$ to obtain node features $\mathbf{X}$.
 - 9: Construct $\mathbf{A}$ from node correlations with lags $1, \dots, \tau$.
 - 10: Symmetrize and normalize $\mathbf{A}$ to form $\mathcal{G} = (\mathbf{X}, \mathbf{A})$.
 - 11: Apply four stacked GCN layers and concatenate their outputs into $\mathbf{Z}$.
 - 12: **for** $s = 1, 2$ **do**
 - 13: Update $(\mathbf{Z}, \mathbf{A})$ through hierarchical graph pooling.
 - 14: Refine $\mathbf{Z}$ using graph attention.
 - 15: **end for**
 - 16: $\mathbf{g} \leftarrow \text{GAP}(\mathbf{Z})$.
 - 17: $\mathbf{p} \leftarrow \text{Softmax}(\mathbf{W}_c \mathbf{g} + \mathbf{b}_c)$.
 - 18: **return** $\hat{y} = \arg \max_k p_k$.
-

4. Case Study

4.1. Data Description

The proposed method was evaluated on a wind-turbine drivetrain fault-simulation test platform, as shown in Figure 4. The platform consisted of blades, a PPSC129 gearbox, a coupling, a YSSF450 generator, and the associated transmission components. The generator operated at a constant rotational speed of 1500 r/min under a load corresponding to 50% of its rated power. The rotational speed and load were regulated by a driving motor and a loading device, respectively.

Accelerometers were installed at the front main-shaft bearing, the low- and high-speed shaft bearings of the gearbox, and the front and rear bearings of the generator. An acoustic sensor was positioned near the gearbox to capture the radiated acoustic response. All vibration and acoustic signals were synchronously acquired at a sampling frequency of 12,800 Hz using a multi-channel data-acquisition system. One vibration channel and one acoustic channel were selected as the two model inputs.

The generator was equipped with SKF 6326 bearings at its front and rear ends. Seven health conditions were considered: normal condition (NC), front-bearing inner-race fault (FBIRF), front-bearing outer-race fault (FBORF), rear-bearing inner-race fault (RBIRF), rear-bearing outer-race fault (RBORF), simultaneous outer-race faults of the front

and rear bearings (FRBORF), and generator-shaft misalignment fault (GSMF). Localized bearing defects with a diameter of 0.5 mm were introduced by electrical discharge machining. Shaft misalignment was introduced by imposing a 0.5 mm horizontal radial offset between the coupled shaft centerlines at the coupling, and the prescribed offset was verified using a dial indicator. For each health condition, ten 20 s recordings were collected. The resulting dataset composition is summarized in Table 1.

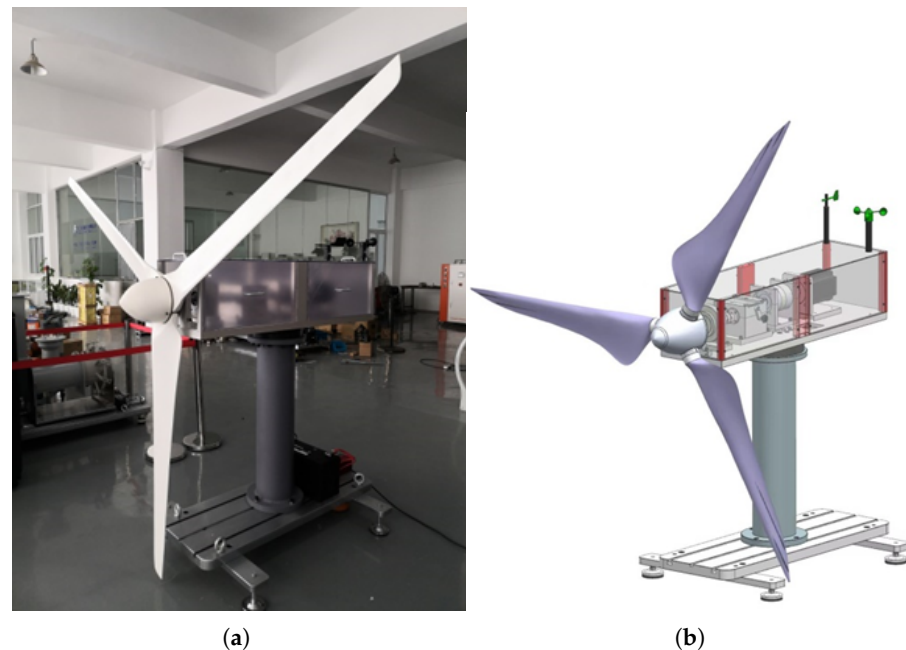


Figure 4. Wind-turbine drivetrain fault-simulation test platform. (a) Typical photograph; (b) Schematic diagram.

Table 1. Description of the wind-turbine drivetrain fault dataset.

Class	Fault Type	Recordings	Duration (s)	Segmented Samples	Operating Condition
NC	Normal condition	10	20	9970	1500 r/min, 50% load
FBIRF	Front-bearing inner-race fault	10	20	9970	1500 r/min, 50% load
FBORF	Front-bearing outer-race fault	10	20	9970	1500 r/min, 50% load
RBIRF	Rear-bearing inner-race fault	10	20	9970	1500 r/min, 50% load
RBORF	Rear-bearing outer-race fault	10	20	9970	1500 r/min, 50% load
FRBORF	Compound outer-race fault	10	20	9970	1500 r/min, 50% load
GSMF	Generator shaft misalignment fault	10	20	9970	1500 r/min, 50% load

4.2. Diagnosis Results

The comparative methods were selected to cover three categories of fault diagnosis approaches related to the proposed framework: multi-source sensor fusion, deep feature extraction and graph-based feature aggregation. This selection was used to evaluate whether AC-STFGCN provides additional benefit beyond existing fusion-based, sequence-based and graph-based diagnostic models. MSCAF is a multi-source sensor correlation adaptive fusion method with uncertainty quantification. It models inter-sensor dependencies and estimates fusion uncertainty to improve the reliability of multi-source fault diagnosis [16]. MSIF-Convformer is an end-to-end multi-source sensor fault diagnosis framework for noisy operating conditions [25]. It integrates convolutional operations and Transformer-based attention to extract local fault-sensitive features and capture global dependencies. MGGF is a multi-level graph-guided fault diagnosis model utilizing a multi-channel data fusion strategy [31]. It constructs hierarchical graph structures to characterize inter-channel dependencies in rotating machinery diagnosis under highly imbalanced class distributions. I2SGCN is a two-stage importance-aware subgraph convolutional

network for multi-source sensor data [29]. It learns domain-invariant representations through subgraph learning and importance-aware feature modeling. MSIDBNs integrate multi-sensor information fusion with an improved deep belief network for limited-sample health-condition diagnosis [32]. MSFDM is a DL-based model from multi-source sensing data for fault detection under imbalanced and limited data scenarios [7].

Vibration signals were recorded using PCB 356A01 triaxial ICP accelerometers with a nominal sensitivity of 5 mV/g ($\pm 20\%$) and a peak measurement range of ± 1000 g. AC-STFGCN used the X-axis signal from an accelerometer mounted on top of the monitored bearing housing, together with a single acoustic channel. Acoustic signals were recorded using a BEHRINGER ECM8000 (Behringer, Willich, Germany) omnidirectional electret condenser microphone positioned near the same bearing housing. The microphone had a nominal frequency response of 15 Hz to 20 kHz and an impedance of 600 Ω . A calibrated microphone sensitivity was unavailable. An NI USB-4431 data-acquisition device synchronously recorded the selected vibration and acoustic channels with 24-bit resolution and a ± 10 V input range. Both channels were sampled at 12,800 Hz within the same hardware-timed acquisition task, sharing a sampling clock and acquisition start time. The acquisition system applied anti-aliasing filtering before storing the signals. Subsequent processing used identical temporal boundaries for both channels to preserve their alignment. Each 20 s recording contained 256,000 samples per channel. Before segmentation, the original recordings were randomly assigned to training, validation, and test sets in a 6:2:2 ratio, stratified by class. Splitting at the recording level prevented overlapping windows from the same recording from entering different sets. The Z-score normalization technique was applied separately to the vibration and acoustic channels using the mean and standard deviation computed exclusively from the training set. The resulting normalization parameters were then applied unchanged to the validation and test sets. The validation set was used for the selection of the model, and the checkpoint achieving the highest macro-F1 validation score was retained for the final evaluation of the test set. All compared methods used the same preprocessing procedures and data partitions. The models were implemented in Python 3.8.5 using PyTorch 2.4.1 and trained on an NVIDIA RTX 4090 GPU. Training was performed for a maximum of 80 epochs with a batch size of 64. The AdamW optimizer was employed with an initial learning rate of 1×10^{-3} , a weight decay of 1×10^{-2} , and a cosine-annealing learning-rate schedule. Each experiment was repeated using ten different random seeds. Accuracy, precision, recall, and F1-score are reported as the mean and standard deviation over the ten runs. Precision, recall, and F1-score were macro-averaged across the seven health conditions, thereby assigning equal weight to each category. The diagnostic results are listed in Table 2.

Table 2. Performance comparison of different methods.

Metric	MSCAF	MSIF-Convformer	MGGF	I2SGCN	MSIDBNs	MSFDM	Ours
Accuracy	96.42 $\pm$ 0.38	97.15 $\pm$ 0.31	95.86 $\pm$ 0.44	97.63 $\pm$ 0.29	93.28 $\pm$ 0.61	94.72 $\pm$ 0.53	99.08 $\pm$ 0.18
Precision	96.35 $\pm$ 0.41	97.08 $\pm$ 0.34	95.94 $\pm$ 0.39	97.51 $\pm$ 0.32	93.41 $\pm$ 0.58	94.89 $\pm$ 0.49	99.12 $\pm$ 0.16
Recall	96.21 $\pm$ 0.45	97.23 $\pm$ 0.28	95.67 $\pm$ 0.47	97.58 $\pm$ 0.30	93.06 $\pm$ 0.66	94.55 $\pm$ 0.57	99.03 $\pm$ 0.20
F1-score	96.28 $\pm$ 0.40	97.15 $\pm$ 0.30	95.80 $\pm$ 0.43	97.54 $\pm$ 0.28	93.23 $\pm$ 0.62	94.71 $\pm$ 0.51	99.07 $\pm$ 0.17

As shown in Table 2, the proposed AC-STFGCN achieved the best diagnostic performance across all evaluation metrics. Compared with the second-best method (i.e., I2SGCN, AC-STFGCN), it improved accuracy, precision, recall and F1-score by 1.45%, 1.61%, 1.45% and 1.5%, respectively. This improvement indicates that adaptive graph construction and PoolGAT-based aggregation provide additional discriminative information beyond subgraph convolution and importance-aware feature modeling. MSIF-Convformer and MSCAF also achieved strong performance, with accuracy values of 97.15% and 96.42%, respectively, in-

dicating the effectiveness of convolution-attention modeling and sensor correlation fusion. MGGF achieved an accuracy of 95.86%, whereas MSIDBNs and MSFDM showed lower performance under the same experimental settings. AC-STFGCN also produced the lowest accuracy standard deviation, at 0.18%, suggesting stable performance across repeated runs.

To further examine the feature representations and diagnostic behaviour of the compared methods, confusion matrices, receiver operating characteristic (ROC) curves and t-distributed stochastic neighbor embedding (t-SNE) visualizations were used. The confusion matrices were used to identify class-wise misclassification patterns. ROC curves and the corresponding area under the curve (AUC) values were used to evaluate threshold-dependent discriminative capability for each fault category [33]. t-SNE was used as a qualitative visualization tool to project learned features from acquired signals into a two-dimensional featurespace and compare the clustering patterns of different health conditions.

The confusion matrices in Figure 5 show that the proposed method achieved stable diagnostic performance across all fault categories, with an overall accuracy of 99.14%. Compared with MSCAF, MSIF-Convformer, MGGF and I2SGCN, the proposed method produced fewer misclassifications among similar fault categories. This indicates that spatio-temporal-frequency features constructed from vibration and acoustic signals improved the separability of different fault patterns. The adaptive correlation graph characterized latent relationships among fault segments, whereas PoolGAT aggregated critical nodes and suppressed redundant information. These results support the effectiveness of multi-source information fusion, adaptive graph construction and hierarchical graph feature extraction for wind-turbine fault diagnosis.

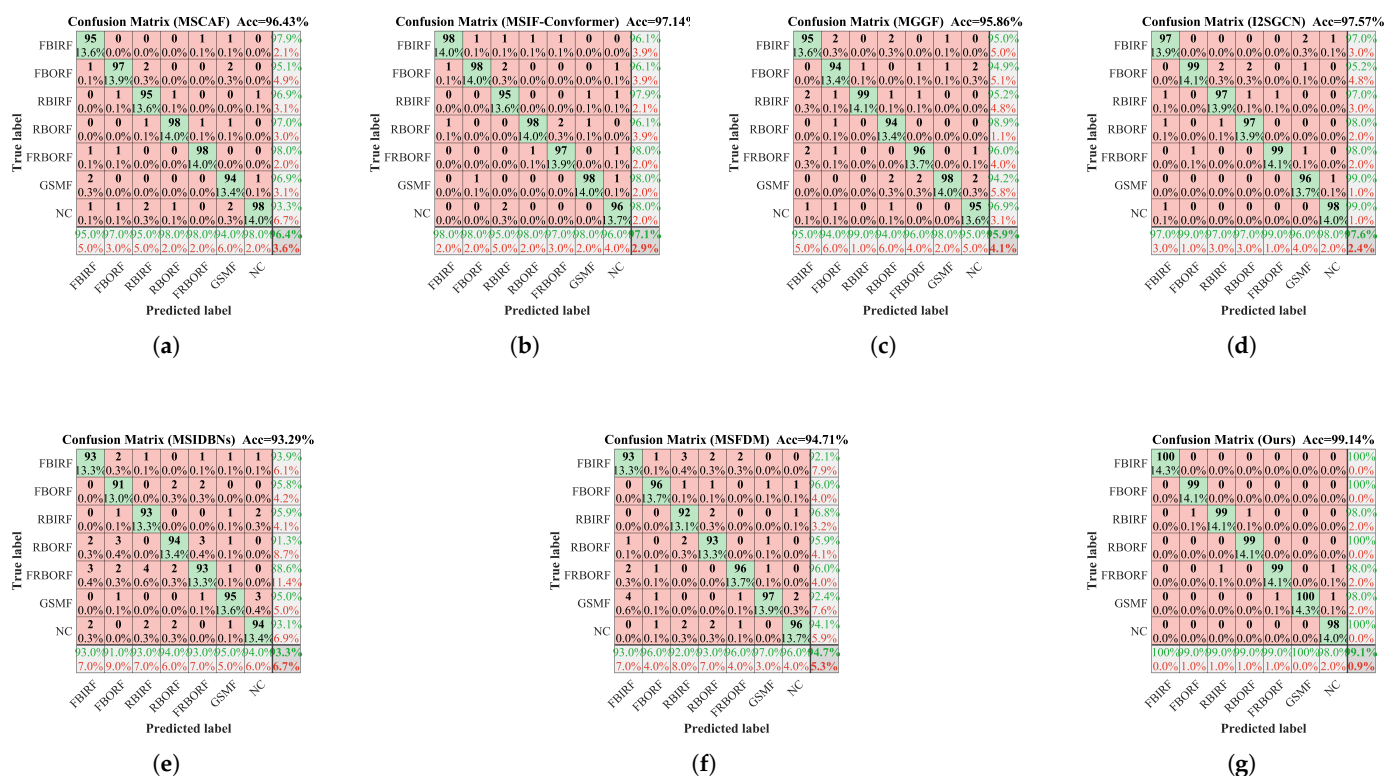


Figure 5. Comparison of confusion matrices for different methods. (a) MSCAF; (b) MSIF-Convformer; (c) MGGF; (d) I2SGCN; (e) MSIDBNs; (f) MSFDM; (g) Ours.

The ROC curves in Figure 6 show that AC-STFGCN achieved strong threshold-dependent discrimination across the seven health conditions. Its curves were closer to the upper-left corner than those of the compared methods for most fault categories, indicating higher true-positive rates and lower false-positive rates over a wide range of decision thresholds. The corresponding AUC values were also closer to 1 for most categories, show-

ing that AC-STFGCN maintained reliable probabilistic discrimination in addition to high classification accuracy. This improvement is consistent with the use of multi-source feature fusion, adaptive graph construction and PoolGAT-based feature aggregation.

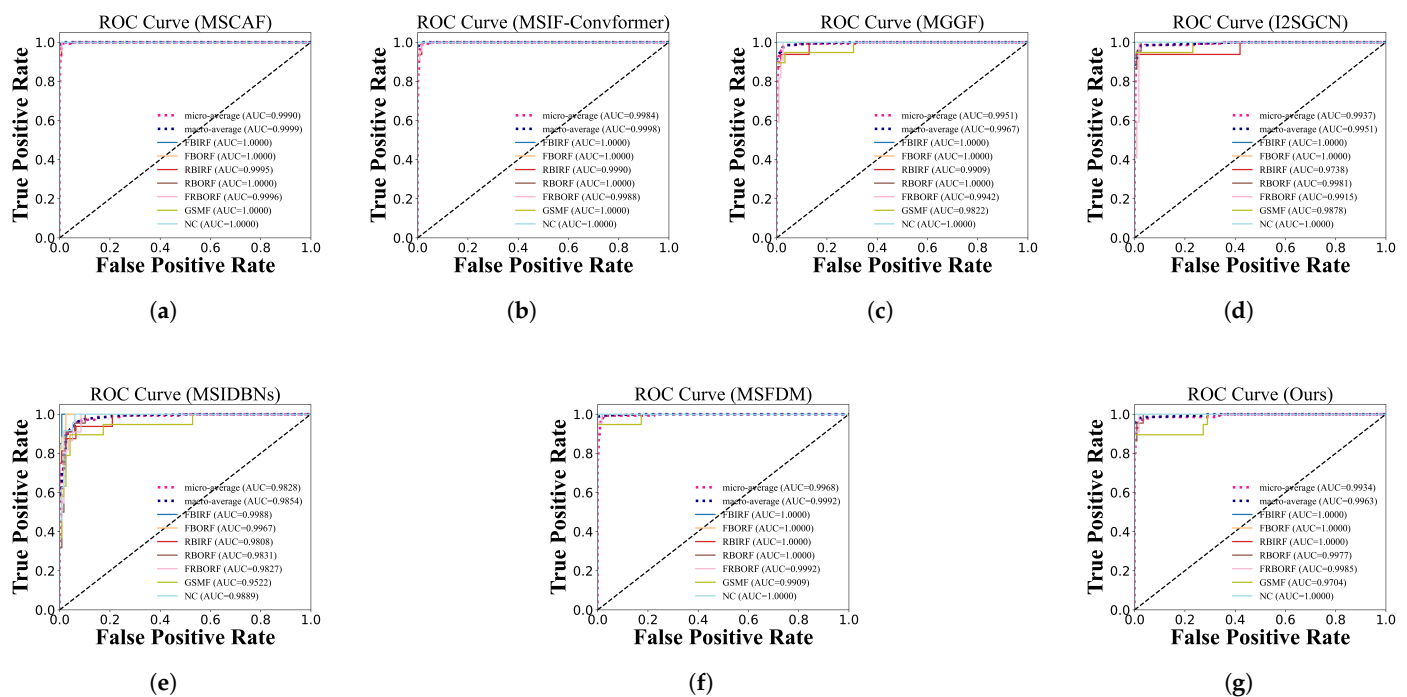


Figure 6. Comparison of ROC curves for different methods. (a) MSAF; (b) MSIF-Convformer; (c) MGGF; (d) I2SGCN; (e) MSIDBNs; (f) MSFDM; (g) Ours.

Figure 7 shows the t-SNE visualizations of features extracted by different methods. MSAF, MSIF-Convformer, MGGF, I2SGCN, MSIDBNs and MSFDM separated the seven health conditions to some extent, but several categories still showed intra-class dispersion or inter-class overlap. In contrast, the features extracted by AC-STFGCN formed more compact clusters with clearer boundaries in the two-dimensional embedding space. This qualitative visualization suggests that vibration-acoustic fusion, joint time–frequency feature extraction, adaptive correlation graph construction and PoolGAT-based aggregation helped improve the separability of fault-sensitive representations.

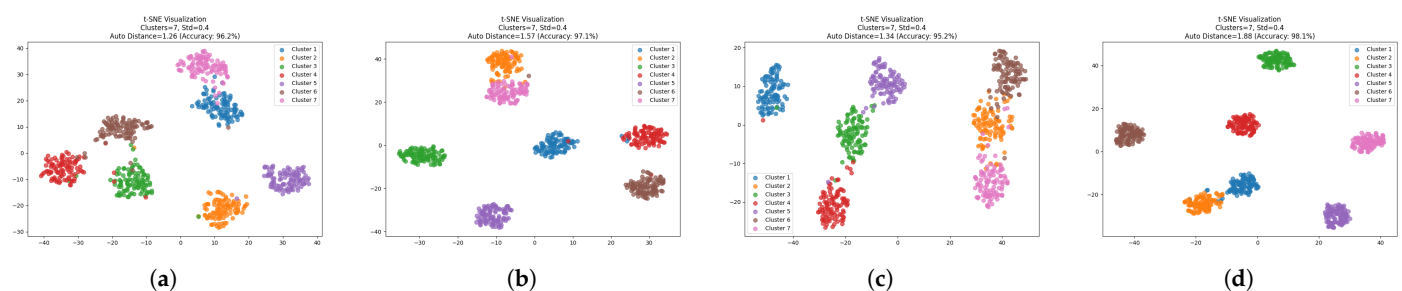


Figure 7. Cont.

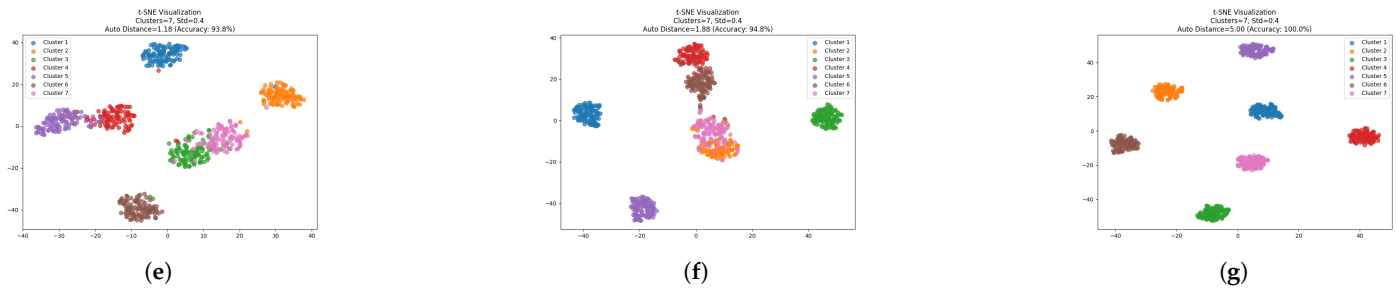


Figure 7. Comparison of t-SNE visualizations for different methods. (a) MSCAF; (b) MSIF-Convformer; (c) MGGF; (d) I2SGCN; (e) MSIDBNs; (f) MSFDM; (g) Ours.

5. Model Discussion

5.1. Ablation Study

To test the contribution of each innovation of the proposed AC-STFGCN, an ablation study was designed for multi-source information fusion, frequency-domain feature extraction, bidirectional temporal modeling, adaptive correlation graph construction and PoolGAT-based graph aggregation. The results are reported in Table 3. All variants used the same data partition, training parameters and hardware environment as the complete model. Each variant removed or replaced only one module, while the remaining components were kept unchanged.

Six ablation variants were designed. In w/o Aco, the acoustic signal was removed, and only vibration signals were used. This variant was used to evaluate the contribution of acoustic information. In w/o Vib, the vibration signal was removed, and only acoustic signals were retained. This variant examined the contribution of vibration information to fault-induced response representation. In w/o STFT, the STFT-based frequency-domain branch was removed, leaving only the one-dimensional convolutional branch for time-domain feature extraction. In w/o Bi-LSTM, the bidirectional temporal-context module was removed, and the fused features were directly mapped to graph nodes. In w/o ACG, the adaptive correlation-based adjacency matrix was replaced with a fixed adjacency matrix to evaluate the effect of sample-specific graph construction. In w/o PoolGAT, hierarchical graph pooling and graph attention were removed, and conventional graph convolution with global average pooling was used for graph-level feature aggregation.

Table 3. Ablation study of the proposed AC-STFGCN.

Model	Accuracy	Precision	Recall	F1-Score
w/o Aco	96.34 ± 0.43	96.41 ± 0.40	96.18 ± 0.47	96.29 ± 0.44
w/o Vib	95.72 ± 0.51	95.83 ± 0.48	95.56 ± 0.55	95.69 ± 0.52
w/o STFT	96.91 ± 0.39	96.98 ± 0.37	96.74 ± 0.42	96.86 ± 0.40
w/o Bi-LSTM	97.35 ± 0.34	97.42 ± 0.31	97.21 ± 0.37	97.31 ± 0.35
w/o ACG	97.86 ± 0.30	97.91 ± 0.28	97.73 ± 0.33	97.82 ± 0.31
w/o PoolGAT	98.31 ± 0.26	98.37 ± 0.24	98.18 ± 0.29	98.27 ± 0.27
AC-STFGCN	99.08 ± 0.18	99.12 ± 0.16	99.03 ± 0.20	99.07 ± 0.17

As shown in Table 3, the complete AC-STFGCN achieved the best fault diagnosis results across all evaluation metrics. Compared with the single-sensor variants, the complete model achieved higher performance by jointly using vibration and acoustic signals. This result indicates that the two signal sources provided complementary fault information. The lower performance of w/o Aco and w/o Vib also shows that relying on a single sensor can lead to incomplete health-condition representation. Removing the

STFT-based frequency-domain branch reduced model performance, indicating that time-domain convolution alone was insufficient to capture modulation frequencies, impact periodicities and time–frequency evolution in the fault signals. The frequency-domain branch therefore provided complementary spectral information within local time windows. Removing Bi-LSTM also decreased performance, suggesting that temporal-context modeling contributed to the representation of non-stationary monitoring signals. By combining forward and backward sequence information, Bi-LSTM improved the encoding of temporal dependencies before graph construction. Replacing the adaptive correlation graph with a fixed adjacency matrix reduced the diagnostic performance, indicating that sample-specific node relationships were useful for representing fault-dependent feature interactions. This result suggests that relationships among local temporal segments and multi-domain features vary with fault type, response intensity and multi-source coupling. Removing PoolGAT also decreased performance, showing that hierarchical graph pooling and graph attention helped aggregate fault-relevant nodes and reduce redundant graph information. Overall, the ablation results support the contribution of vibration-acoustic fusion, STFT-based frequency-domain extraction, Bi-LSTM temporal-context modeling, adaptive correlation graph construction and PoolGAT-based aggregation. Among these components, multi-source fusion and time–frequency representation produced the largest performance decreases when removed, whereas adaptive graph construction and PoolGAT further improved graph-level feature representation.

5.2. Training and Loss Curves

The training behaviour of the proposed AC-STFGCN was further compared with that of the other methods. Figure 8 shows the variations in loss and accuracy during training. The proposed AC-STFGCN exhibited a faster decrease in training loss and reached a more stable convergence state than the compared methods. Its accuracy curve also increased more rapidly during the early training stage and remained stable after convergence. These trends indicate that the proposed feature-fusion and graph-aggregation modules improved optimization stability under the same training settings.

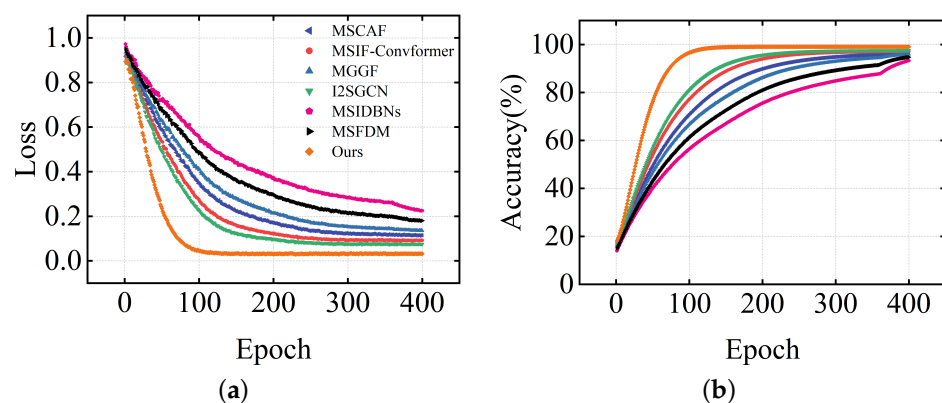


Figure 8. Comparison of loss and accuracy values during method training. (a) Training Loss. (b) Training accuracy.

5.3. Friedman Test and Holm Post Hoc Analysis

To further examine the overall differences among the compared fault diagnosis methods, the Friedman non-parametric test was used to analyse their average ranks across repeated experimental trials. The average ranking results are reported in Table 4. A smaller rank indicates better overall diagnostic performance.

Table 4. Average accuracy ranks of different fault diagnosis methods.

Method	MSCAF	MSIF-Convformer	MGGF	I2SGCN	MSIDBNs	MSFDM	AC-STFGCN
Average rank	4.05	2.95	4.95	2.05	7.00	5.90	1.10

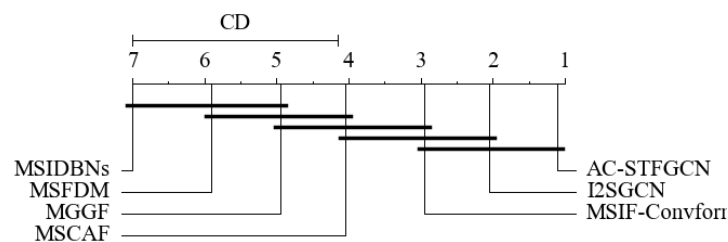
Based on the average ranks, the Friedman statistic is defined as

$$\tau_{\chi^2} = \frac{12N}{k(k+1)} \left(\sum_{i=1}^k r_i^2 - \frac{k(k+1)^2}{4} \right), \quad (59)$$

where N is the number of repeated experimental trials, k is the number of compared methods, and r_i denotes the average rank of the i -th method. The corresponding Iman-Davenport statistic is expressed as

$$\tau_F = \frac{(N-1)\tau_{\chi^2}}{N(k-1) - \tau_{\chi^2}}. \quad (60)$$

In this study, $N = 10$ and $k = 7$. The Friedman test produced $\tau_{\chi^2} = 57.49$ with $p < 0.001$, indicating statistically significant differences among the compared methods. Meanwhile, the critical difference (CD) diagram is shown in Figure 9.

**Figure 9.** CD diagram of different methods.

Because AC-STFGCN obtained the best average rank, Holm post hoc analysis was further conducted by comparing AC-STFGCN with each baseline method. The Holm post hoc comparison results are given in Table 5. The Holm-corrected comparisons show that the proposed AC-STFGCN achieved the highest mean diagnostic accuracy among the compared methods. The differences relative to MSIF-Convformer and I2SGCN did not reach statistical significance after Holm correction.

Table 5. Holm post hoc comparison between AC-STFGCN and baseline methods.

Comparison	Rank Difference	z Value	Holm-Adjusted p	Significant
AC-STFGCN vs. MSIDBNs	5.90	6.11	<0.001	Yes
AC-STFGCN vs. MSFDM	4.80	4.97	<0.001	Yes
AC-STFGCN vs. MGGF	3.85	3.99	<0.001	Yes
AC-STFGCN vs. MSCAF	2.95	3.05	0.0068	Yes
AC-STFGCN vs. MSIF-Convformer	1.85	1.91	0.111	No
AC-STFGCN vs. I2SGCN	0.95	0.98	0.326	No

5.4. Evaluation Under Different Load Conditions

To investigate the influence of various working loads on diagnostic performance, experiments were conducted at 0%, 25%, 50%, 75%, and 100% of the rotating load, as given in Table 6.

Table 6. Diagnostic performance of AC-STFGCN under different load conditions.

Rated Load (%)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
0	98.26 ± 0.37	98.39 ± 0.32	98.18 ± 0.41	98.28 ± 0.35
25	98.87 ± 0.24	98.96 ± 0.21	98.79 ± 0.28	98.86 ± 0.25
50	99.08 ± 0.18	99.12 ± 0.16	99.03 ± 0.20	99.07 ± 0.17
75	98.92 ± 0.23	99.01 ± 0.19	98.84 ± 0.26	98.91 ± 0.22
100	98.61 ± 0.31	98.73 ± 0.27	98.52 ± 0.34	98.60 ± 0.30

It can be seen that the proposed AC-STFGCN achieves all four diagnostic metrics above 98% across the different loads, with only limited diagnostic performance variation. The best performance is observed at 50% load, with comparable results at 25% and 75% load, whereas modest decreases and greater fluctuations are observed under no-load and rated-load conditions. These results indicate that AC-STFGCN exhibits some sensitivity to load-dependent signal variations while maintaining consistently high diagnostic performance across the evaluated load range.

5.5. Computational Efficiency and Practical Limitations

Computational efficiency affects the practical deployment of fault diagnosis methods. Table 7 compares the training and inference times of all methods on the experimental platform described in Section 4.

Table 7. Training and inference times of different methods.

Metric	MSCAF	MSIF-Convformer	MGGF	I2SGCN	MSIDBNs	MSFDM	Ours
Total training time (s)	848	912	784	816	624	704	960
Training time per epoch (s)	10.6	11.4	9.8	10.2	7.8	8.8	12.0
Inference time per sample (ms)	2.4	2.7	2.2	2.3	1.7	1.9	2.8

It can be seen that the proposed AC-STFGCN achieves improved diagnostic performance with moderately longer training and inference times than the benchmark methods. The additional computational cost arises primarily from multi-domain feature extraction, temporal encoding, and adaptive graph processing. These results reflect a trade-off between diagnostic performance and computational efficiency. More efficient implementations and faster hardware could reduce this overhead.

5.6. Diagnostic Performance Under Added Noise

Noise robustness is evaluated at 50% of the rated load and a rotational speed of 1500 r/min. Independent zero-mean Gaussian white noise is added to both vibration and acoustic test signals. The models are trained on the original data and tested on the noise-corrupted signals without retraining. Noise intensity is defined by the signal-to-noise ratio (SNR), which can be defined as follows:

$$\text{SNR} = 10 \log_{10} \left(\frac{P_s}{P_n} \right), \quad (61)$$

where P_s and P_n are the powers of the original signal and the added noise, respectively. Both channels use the same target SNR, with noise power calculated separately for each channel. The evaluated SNR levels are −10, −8, −6, −4, −2, 0, and 2 dB, as shown in Figure 10. The proposed AC-STFGCN is compared with its vibration-only and acoustic-only variants using identical data partitions and the corresponding noise-corrupted inputs.

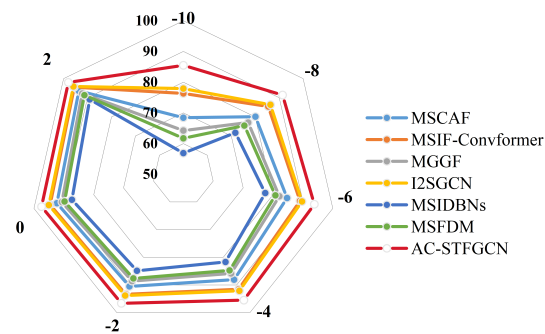


Figure 10. Diagnostic performance at 50% load under different (SNRs).

It can be seen that diagnostic accuracy of all methods decreases with SNR. The proposed AC-STFGCN is the most robust to noise across the evaluated SNR range and maintains satisfactory accuracy on all tasks.

5.7. Sensitivity to the Maximum Correlation Lag

The maximum correlation lag τ defines the temporal range within which graph nodes can connect. Its effect is evaluated at 50% of the load and 1500 r/min [34,35]. Only τ is varied, with data partitions, other network parameters, and training settings held fixed. Each configuration is trained independently over ten runs. For a graph with N nodes, the number of candidate undirected edges, excluding self-loops, is given by

$$E_{\tau} = \sum_{d=1}^{\tau} (N - d) = N\tau - \frac{\tau(\tau + 1)}{2}, \quad 1 \leq \tau < N. \quad (62)$$

The results is given in Table 8. The diagnostic accuracy increases from 96.34% at $\tau = 1$ to a maximum of 99.08% at $\tau = 11$, before decreasing slightly to 98.95% at $\tau = 13$ and 98.73% at $\tau = 15$. This trend represents that a narrow connectivity range excludes useful temporal relationships, whereas an overly broad range introduces redundant interactions. The small accuracy differences near $\tau = 11$ suggest a relatively broad range of settings with comparable performance.

Table 8. Sensitivity of diagnostic accuracy to the maximum correlation lag.

Maximum Correlation Lag τ	Simulated Accuracy (%)
1	96.34
3	97.12
5	98.05
7	98.61
9	98.86
11	99.08
13	98.95
15	98.73

6. Conclusions

To address the non-stationary of wind-turbine drivetrain fault signals, the coupling of multi-source information and the limited adaptability of conventional fixed graph structures, this study proposes an adaptive correlation spatio-temporal-frequency graph convolutional network (AC-STFGCN) for multi-source wind-turbine fault diagnosis. The proposed method integrates vibration and acoustic signals and jointly represents multi-source and multi-domain fault features through one-dimensional convolution, STFT technique, Bi-LSTMs, adaptive graph construction and PoolGAT-based graph feature aggregation. The main conclusions are summarized as follows.

- (1) AC-STFGCN effectively integrates complementary fault information from vibration and acoustic signals. Compared with single-sensor inputs, multi-source information fusion provides a more comprehensive representation of fault states and reduces the influence of sensor installation locations, signal transmission paths and noise interference on diagnostic performance.
- (2) The time-domain convolutional branch and the STFT-based frequency-domain branch extract local impact responses and time–frequency modulation characteristics, respectively. Bi-LSTM further enhances the temporal-context representation of non-stationary fault signals. The ablation results show that removing the acoustic signal, vibration signal, STFT branch or Bi-LSTM module leads to different degrees of performance degradation, indicating the importance of multi-source time–frequency features and temporal modeling for wind-turbine fault identification.
- (3) The adaptive correlation graph construction method dynamically generates adjacency matrices according to sample-specific features. Compared with fixed graph structures, this strategy is more suitable for describing non-Euclidean correlations among local temporal segments and multi-domain features under different fault conditions. Pool-GAT further aggregates fault-sensitive nodes through hierarchical graph pooling and graph attention mechanisms while suppressing redundant graph information, thereby improving the discriminative capability of graph-level feature representations.
- (4) On the dataset collected from a wind-turbine drivetrain fault-simulation test platform, AC-STFGCN achieves accuracy, precision, recall and F1-score values of 99.08%, 99.12%, 99.03% and 99.07%, respectively. These results outperform those of the compared multi-source fusion, deep feature extraction and graph-neural-network-based diagnostic methods. The confusion matrices, ROC curves and t-SNE visualizations further indicate that the proposed method provides strong discrimination between similar fault types and stable classification performance.
- (5) The proposed method has currently been validated mainly on a single wind-turbine drivetrain test platform. Its generalization capability across operating conditions, equipment types and fault-severity levels therefore requires further investigation. Future work will incorporate a broader range of operating conditions, different load levels and independent test datasets to further evaluate the engineering applicability and robustness of the proposed model.

The proposed AC-STFGCN uses a relatively complex processing pipeline and requires synchronized multi-source signals. Its computational efficiency on embedded hardware and diagnostic reliability in the field require further validation. Future work will focus on lightweight architectures and robustness to missing or degraded sensor signals. Industrial data will support validation across different machines and working conditions.

Funding: This work was supported by the Science and Technology Research Project of Jiangxi Provincial Department of Education, grant number GJJ2402809, and Private Special Project of Nanchang Institute of Science and Technology, grant number NGMBZX-26-04.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT-6 As-tra solely for English-language polishing and grammatical correction. The authors subsequently reviewed and revised the text and take full responsibility for the content of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liang, Y.; Chen, J.; Liu, R.; Zhou, H.; Kang, N.; Zhou, N. A Novel Bearing Fault Diagnosis Framework with a Multi-Scale Feature Extraction Module and Efficient Content-Guided Attention Mechanism. *Lubricants* **2026**, *14*, 121. [\[CrossRef\]](#)
2. Yu, Y.; Karimi, H.R.; Shi, P.; Peng, R.; Zhao, S. A New Multi-Source Information Domain Adaptation Network Based on Domain Attributes and Feature Transfer for Cross-Domain Fault Diagnosis. *Mech. Syst. Signal Process.* **2024**, *211*, 111194. [\[CrossRef\]](#)
3. Tian, J.; Yu, Y.; Karimi, H.R.; Gao, F.; Lin, J. A continual test-time domain adaptation method for online machinery fault diagnosis under dynamic operating conditions. *Neural Netw.* **2026**, *194*, 108192. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Zheng, X.; Xiao, Y.; Li, X. Lightweight MS-DSCNN-AttMPLSTM for High-Precision Misalignment Fault Diagnosis of Wind Turbines. *Machines* **2026**, *14*, 155. [\[CrossRef\]](#)
5. Yang, Z.; Wu, J.; Ma, Y. Towards Reliable Electrified Powertrain Fault Diagnosis under Imperfect Data via Gradient Correlative Alignment. *Reliab. Eng. Syst. Saf.* **2026**, *277*, 113202. [\[CrossRef\]](#)
6. Qin, G.; Jin, Y.; Qiao, L.; Wu, Z. Advances, Challenges, and Recommendations for Non-Destructive Testing Technologies for Wind Turbine Blade Damage: A Review of the Literature from the Past Decade. *Sensors* **2026**, *26*, 1773. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Shi, P.; Yu, Y.; Gao, H.; Hua, C. A novel multi-source sensing data fusion-driven method for detecting rolling mill health states under imbalanced and limited datasets. *Mech. Syst. Signal Process.* **2022**, *171*, 108903. [\[CrossRef\]](#)
8. Tian, J.; Yu, Y.; Karimi, H.R.; Lin, J. A test-time adaptation method using evidential deep learning for online machinery fault diagnosis. *Knowl.-Based Syst.* **2026**, *331*, 114831. [\[CrossRef\]](#)
9. Su, X.; Han, J.; Chen, C.; Lu, J.; Ma, W.; Dai, X. Intelligent Workshop Bearing Fault Diagnosis Method Based on Improved Convolutional Neural Network. *Lubricants* **2025**, *13*, 521. [\[CrossRef\]](#)
10. Yu, Y.; Karimi, H.R.; Gelman, L.; Cetin, A.E. MSIFT: A Novel End-to-End Mechanical Fault Diagnosis Framework under Limited and Imbalanced Data Using Multi-Source Information Fusion. *Expert Syst. Appl.* **2025**, *274*, 126947. [\[CrossRef\]](#)
11. Mohamed Razick, F.R.; Musilek, P. Deep Learning for Short-Circuit Fault Diagnostics in Power Distribution Grids: A Comprehensive Review. *Computers* **2026**, *15*, 76. [\[CrossRef\]](#)
12. Voltolini, E.; Toscani, A.; Armelloni, E.; Cocconcelli, M.; Fendillo, L.; Manconi, E. Non-Contact Bearing Fault Diagnostics: Experimental Investigation of Microphones Position and Distance. *Appl. Sci.* **2026**, *16*, 3670. [\[CrossRef\]](#)
13. Lehmann, B.; Trompeter, P.; Gutiérrez Guzmán, F.; Jacobs, G. Evaluation of Wear Models for the Wear Calculation of Journal Bearings for Planetary Gears in Wind Turbines. *Lubricants* **2023**, *11*, 364. [\[CrossRef\]](#)
14. Peng, H.; Zhang, H.; Fan, Y.; Shangguan, L.; Yang, Y. A Review of Research on Wind Turbine Bearings' Failure Analysis and Fault Diagnosis. *Lubricants* **2023**, *11*, 14. [\[CrossRef\]](#)
15. Nakai, T.; Lu, R.; Tani, H.; Koganezawa, S.; Wang, J. Early-Stage Damage Diagnosis of Rolling Bearings Based on Acoustic Emission Signals Interpreted by Friction Behavior and Machine Learning. *Lubricants* **2026**, *14*, 95. [\[CrossRef\]](#)
16. Yu, Y.; Karimi, H.R.; Gelman, L.; Tian, J.; Mei, P. A Novel Multi-Source Sensor Correlation Adaptive Fusion Framework with Uncertainty Quantification for Intelligent Fault Diagnosis. *Reliab. Eng. Syst. Saf.* **2025**, 111812. [\[CrossRef\]](#)
17. Gelman, L.; Kerrouche, R.; Abdullahi, A.O. Novel Fault Diagnosis Technology Based on Integrated Spectral Kurtosis for Gearboxes. *Sensors* **2026**, *26*, 2185. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Zhao, S.; Karimi, H.R.; Yu, Y. Multi-Source Domain Adaptation for Fault Diagnosis: A Unified Framework for Feature and Relation Transfer. *Signal Image Video Process.* **2025**, *19*, 787. [\[CrossRef\]](#)
19. Vlachou, V.I.; Kudelina, K.; Efstathiou, D.E.; Vologiannidis, S.D.; Baraškova, T.; Shirokova, V.; Karakatsanis, T.S. Correlation-Driven Multisensory Fusion for Intelligent Fault Analysis in Induction Motors. *Machines* **2026**, *14*, 606. [\[CrossRef\]](#)
20. Ambarita, E.E.; Karlsen, A.; Hasan, A. Joint Actuator and Sensor Fault Diagnosis in Wind Turbine Systems. *Wind Energy* **2025**, *28*, e70059. [\[CrossRef\]](#)
21. Paolini, D.; Dini, P.; Elhanashi, A.; Saponara, S. Advanced Fault Detection and Diagnosis Exploiting Machine Learning and Artificial Intelligence for Engineering Applications. *Electronics* **2026**, *15*, 476. [\[CrossRef\]](#)
22. Yu, Y.; Karimi, H.R.; Pawelczyk, M. MLFNet: A Novel Multi-Level Fusion Mechanical Fault Diagnosis Network under Limited and Imbalanced Datasets Using Multi-Source Information. *Mech. Syst. Signal Process.* **2026**, *243*, 113674. [\[CrossRef\]](#)
23. Xu, X.; Ou, X.; Ge, L.; Qiao, Z.; Shi, P. Simulated data-assisted fault diagnosis framework with dual-path feature fusion for rolling element bearings under incomplete data. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 3542617. [\[CrossRef\]](#)
24. Shi, P.; Liu, H.; Xu, X.; Zhao, D.; Hua, C. M4oE: A Multi-Task Multi-Input Multi-Scale Mixture-of-Experts Method for Multi-Sensor Fusion Diagnosis. *IEEE Internet Things J.* **2026**, *13*, 22103–22118.
25. Li, M.; Huang, J.; Zhang, F.; Yu, Y.; Gu, F.; Chu, F. MSIF-Convformer: A Novel End-to-End Fault Diagnosis Framework with Multi-Source Sensors under Strong Noise. *Inf. Fusion* **2025**, 104000. [\[CrossRef\]](#)
26. Udmale, S.S.; Nath, A.G.; Singh, D.; Singh, S.K. Data-Driven Approach under Varying Operating Conditions for Bearing Fault Diagnosis. *Ann. Oper. Res.* **2026**, *358*, 495–528. [\[CrossRef\]](#)
27. Nie, Q.; Geng, J.; Liu, C. A Review of Fault Diagnosis Methods: From Traditional Machine Learning to Large Language Model Fusion Paradigm. *Sensors* **2026**, *26*, 702. [\[CrossRef\]](#) [\[PubMed\]](#)

28. Yu, Y.; Karimi, H.R.; Liu, C. Weighted Feature Fusion of Convolutional Neural Network and Graph Convolutional Network for Mechanical Fault Diagnosis under Time-Varying Speeds. *IFAC-PapersOnLine* **2024**, *58*, 729–733. [[CrossRef](#)]
29. Yu, Y.; He, Y.; Karimi, H.R.; Gelman, L.; Cetin, A.E. A Two-Stage Importance-Aware Subgraph Convolutional Network Based on Multi-Source Sensors for Cross-Domain Fault Diagnosis. *Neural Netw.* **2024**, *179*, 106518. [[CrossRef](#)] [[PubMed](#)]
30. Wang, B.; Zhao, S. MTAGCN: Multi-Task Graph-Guided Convolutional Network with Attention Mechanism for Intelligent Fault Diagnosis of Rotating Machinery. *Machines* **2025**, *13*, 347. [[CrossRef](#)]
31. Yu, Y.; Karimi, H.R.; Kundu, P.; Zio, E.; Feng, K. A Multi-Channel Data Fusion-Enabled Multilevel Graph-Guided Framework for Fault Diagnosis in Rotating Machinery under Extreme Biased Data. *Inf. Fusion* **2026**, *134*, 104339. [[CrossRef](#)]
32. Yu, Y.; Shi, P.; Tian, J.; Xu, X.; Hua, C. Rolling Mill Health States Diagnosing Method Based on Multi-Sensor Information Fusion and Improved DBNs under Limited Datasets. *ISA Trans.* **2023**, *134*, 529–547. [[CrossRef](#)] [[PubMed](#)]
33. Yilmaz, Ü. Behavioral Fault Diagnosis in Inverter-Driven PMSM Systems Using a Hybrid CNN–BiLSTM–Attention Deep Learning Framework with SHAP-Based Interpretability. *Machines* **2026**, *14*, 638. [[CrossRef](#)]
34. Liu, Z.; Zhang, Z.; Shao, H.; Zhao, K. A Universal Domain Adaptation Fault Diagnosis Method Based on Feature Invariance and Collaborative Mutual Learning. *Mech. Syst. Signal Process.* **2026**, *258*, 114723. [[CrossRef](#)]
35. Garavagno, A.M.; Ragusa, E.; Gastaldo, P.; Frisoli, A.; Zunino, R. BearingNAS: Obtaining In-Sensor Intelligent Fault Diagnosis Systems for Bearings Using a Laptop. *arXiv* **2026**, arXiv:2607.18287.

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.