

Adaptive adjustment graph representation learning method for rotating machinery fault diagnosis under noisy signals

Lei WANG^a, Peijie YOU^a, Xin ZHANG (✉)^b, Li JIANG^a, Yibing LI^a

^a School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430070, China

^b Mechanical and Aerospace Engineering, The Hong Kong University of Science and Technology, Hong Kong 999077, China

✉ Corresponding author. Email: mexzyl@ust.hk (Xin ZHANG)

© The Author(s) 2025. This article is published with open access at link.springer.com and journal.hep.com.cn

ABSTRACT Recently, intelligent fault diagnosis methods have been employed in the condition monitoring of rotating machinery. Among them, graph neural networks are emerging as a new feature extraction tool that can mine the relationship characteristics between samples. However, many existing graph construction methods suffer from structural redundancy or missing node relationships, thus limiting the diagnosis accuracy of the models in practice. In this paper, an adaptive adjustment k -nearest neighbor graph-driven dynamic-weighted graph attention network (AAKNN-DWGAT) is proposed to address this problem. First, time-domain signals are transformed into frequency-domain features by using fast Fourier transformation. Subsequently, a frequency similarity evaluation method based on dynamic frequency warping is proposed, which enables the conversion of distance measurements into a frequency similarity matrix (FSM). Then, an adaptive edge construction operation is conducted on the basis of FSM, whereby the effective domain is captured for each node using an adaptive edge adjustment method, generating an AAKNN graph (AAKNNG). Next, the constructed AAKNNG is fed into a dynamic-weighted graph attention network (DWGAT) to extract the fault features of nodes layer by layer. In particular, the proposed DWGAT employs a dynamic-weighted strategy that can update the edge weight periodically using high-level output features, thereby eliminating the adverse impacts caused by noisy signals. Finally, the model outputs fault diagnosis results through a softmax classifier. Two case studies verified the effectiveness and the superiority of the proposed method compared with other graph neural networks and graph construction methods.

KEYWORDS rotating machinery, fault diagnosis, graph neural network, adaptive adjustment

1 Introduction

Rotating machinery, such as engines, gearboxes, and axial flow pumps, are essential equipment in industrial production [1]. However, it is prone to various operational risks, including bearing wear and gear failure, which can result in economic losses and serious accidents [2,3]. Therefore, fault diagnosis technology is crucial for monitoring and analyzing mechanical operation states, promptly identifying potential problems, and improving equipment reliability.

Recently, intelligent fault diagnosis (IFD) has emerged as a prominent research area [4–7]. Traditional IFD methods usually utilize simple machine learning models such as artificial neural networks [8], extreme learning machines [9], and support vector machines [10,11].

Nevertheless, these methods suffer from a shallow architecture and exhibit suboptimal performance in fault diagnosis. With the advancement of deep learning, neural network models with complex structures are employed to fill this gap [12]. For instance, Li et al. [13] introduced a wavelet capsule network for compound fault, consisting of wavelet convolution layers and two capsule layers. This network achieved higher compound fault diagnosis accuracy under the scenario of incomplete fault data. Xing et al. [14] introduced deep belief networks into the fault diagnosis task, which can autonomously learn distribution-invariant features from raw vibration data, resulting in substantial recognition accuracy.

However, conventional deep learning methods assume independence and identical distribution of samples and neglect the organizational information within the data [15,16]. As a framework that can jointly represent data and their interconnections, graph theory provides an effective pathway for mining the internal features of

samples and their relationship network [17]. On this basis, graph neural networks (GNNs) have emerged, which can enable feature extraction by the transfer and aggregation of information between nodes [18]. It has been applied in many prediction or classification tasks, such as remaining useful life prediction [19], social network analysis [20], and knowledge graphs. As a widely adopted GNN method in different research areas, graph convolutional networks (GCNs) can extract graph structure features through convolutional operations [21,22]. For example, Chen et al. [23] employed the structural analysis method for pre-diagnosis and establishing the correlation graph, which is then inputted into a GCN to identify the fault types of a traction system. Zhao et al. [24] presented a temporal convolutional network model for traffic prediction in urban road networks. Chen et al. [25] proposed a multichannel domain adaptation GCN to extract domain-invariant features under varying working conditions. However, importance differences between neighbor nodes in the graph feature learning process of GCN were rarely considered in previous studies [26].

Graph attention network (GAT) overcomes the above limitations of the original GCN. It introduces an attention mechanism that fuses the attention coefficients of neighbor nodes into the information aggregation of the source node [27]. For instance, Jiang et al. [28] applied a multi-head GAT network (MHGAT) to bearing fault diagnosis, which obtains feature relationships between nodes through the parallel computation of multiple graph attention heads and achieves reliable diagnostic results. Ding et al. [29] introduced a multi-modal spatial-temporal GAT for multi-modal time series anomaly detection. Zhang et al. [30] proposed a multi-scale channel attention-driven graph dynamic fusion network for mechanical fault diagnosis, which explores the importance differences between channels at multiple scales. In general, the effect of these GNN-related methods is highly influenced by the graph construction process. Recently, some studies have used k -nearest neighbor graphs (KNNGs) to construct the graph data. For instance, Tao et al. [31] proposed a GAT model based on pooling KNNG, which achieves a high diagnosis accuracy under a small number of labeled samples. Yu et al. [32] extracted time and frequency-domain features for calculating Euclidean distance to build node-level KNNG. Xie et al. [33] proposed a graph construction method combining KNNG and kernel functions. However, original KNNGs usually require setting a fixed value of k , which represents the same number of neighbor nodes, thus inevitably causing redundancy in the graph structure. The redundant edge connections not only increase the model's training time but also introduce uncertain neighbor information. Particularly in

workplaces with high levels of noise from rotating machinery, the neighbor information from redundant edges may introduce more noisy signals, which negatively affect the judgment of the central node [34,35].

According to the above analysis, GNN-related fault diagnosis studies for rotating machineries face three main challenges: 1) Finding a reliable way to construct edge connections is a complex task that requires comprehensive consideration of the node attributes, and task requirements. 2) The required neighborhood information varies for different nodes because of differences in node attributes. This situation makes traditional KNNG with a fixed value of k susceptible to introducing incorrect information. 3) Considering the noise interference in actual operating conditions of rotating machinery, ensuring the reliability of the GNN model training process is challenging.

To address the above problems, this paper proposes an adaptive adjustment k -nearest neighbor graph-driven dynamic-weighted GAT (AAKNN-DWGAT). To address the first challenge, this paper presents a dynamic frequency warping (DFW) method to calculate the frequency similarities between the nodes, which can capture the feature relationships between nodes more accurately than the Euclidean distance metric. For the second challenge, the proposed method employs the 3σ criterion and the second-order difference method to optimize the original graph structure. This approach ensures that each node is assigned an appropriate number of edge connections. For the third challenge, a DWGAT is designed, which can dynamically adjust the edge weights according to the high-level output features. The model training process progresses toward loss reduction. Thus, high-level features obtained through multiple iterations contain more fault information, which can be used to update the edge weights and reduce the negative influence of noisy signals. The primary contributions of this paper are as follows:

(1) A DFW method is designed to achieve the time-frequency domain transformation and frequency similarity assessment, thereby enhancing the reliability of the graph structure.

(2) An adaptive edge adjustment method that identifies the inflection points of similarity distribution and eliminates outliers is proposed. It constructs a weighted topological graph structure with adaptive edge adjustment, thereby enhancing the graph's representation ability.

(3) A dynamic-weighted GAT (DWGAT) is developed to extract discriminative fault information from multiple scales and classify the final fault labels, where the proposed dynamic weighting strategy can reduce the impact of noise interference on the input graph to some extent.

(4) The effectiveness and noise robustness of the proposed AAKNN-DWGAT method are validated by two case studies for rotating machinery.

2 Preliminaries

2.1 Graph theory

Graph is a data structure composed of a set of nodes (vertices) and edges that connect these nodes. It frequently offers sophisticated and abstract depictions of real-world issues, such as social network graphs or citation network graphs.

Various graph construction methods exist, including the ε -ball neighborhood and KNNG. ε -graph, also known as epsilon graph, is a graph structure in which the connectivity between nodes is determined by their similarity. In an ε -graph, a similarity measure (such as Euclidean distance) is calculated for each pair of nodes. The obtained similarity value is then compared with a predefined threshold ε . If the similarity between two nodes exceeds ε , then they are considered neighbor nodes and connected by an edge in the graph.

In the KNN algorithm, the k -nearest neighbors for each node are determined by calculating the distances or similarities relative to those neighbors. The obtained KNNG at the node level is an undirected graph that can be mathematically represented as $G = \{V, E, A, F\}$, where V denotes the set of nodes, E represents the collection of edges among samples that establishes the connections between nodes, and A denotes the adjacency matrix, which represents the connectivity relationship between nodes. $F \in \mathbb{R}^{m \times n}$ represents the feature matrix of the nodes, where m represents the number of nodes, and n represents the dimensionality of the node features. The first step in constructing KNNG is to calculate the distance matrix $H \in \mathbb{R}^{m \times m}$, where the elements $W_{ij} = \|F_1 - F_2\|_2$ represent the Euclidean distance between the feature vectors of nodes. This distance matrix is used to identify the k -nearest neighbors for each node, generating the collection of edges E among samples and constructing the adjacency matrix A .

$$A_{ij} = \begin{cases} 1 & \text{if } V_{ij} \in E, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

2.2 Multi-head graph attention network

The attention mechanism is designed to emulate the cognitive process of human information processing. Such a mechanism enables models to focus selectively on specific input segments and allocate more attention and resources to them, thereby improving model performance. Recent studies have successfully incorporated the attention mechanism into GNN, resulting in the

emergence of GATs. GATs can efficiently learning features and representations from graph data.

The MHGAT captures the feature relationships among nodes by performing parallel computations with multiple graph attention heads. Each attention head calculates attention coefficients between neighbor nodes and aggregates node information through a weighted summation approach. Ultimately, an improved final node representation can be obtained by concatenating or averaging the attention results obtained from multiple heads.

2.2.1 Attention coefficient calculation

In the feature learning process of GNN, attention coefficients serve as a crucial metric for determining the allocation of weights between nodes. The mathematical definition of the similarity coefficient e_{ij} , which represents the degree of similarity between any given node v_i , and its neighbor node v_j can be expressed as follows:

$$e_{ij} = a(\|Wh_i\|Wh_j), j \in N, W \in \mathbb{R}^{d(l+1) \times d(l)}, \quad (2)$$

where h_i denotes the feature vector of node v_i , while h_j signifies the feature vector of node v_j . $W \in \mathbb{R}^{d(l+1) \times d(l)}$ represents the weight parameters used for transforming node features, with $d(l)$ representing the length of the feature vector at the l th layer. N denotes the set of all neighbor nodes of v_i . As shown in Fig. 1, initially, the nodes employ linear transformation using the weight matrix W to enhance their features. Then, the enhanced node vectors v_i and v_j are concatenated. Subsequently, the concatenated high-dimensional features are mapped to the real domain using the function $a(\cdot)$, with LeakyReLU serving as the nonlinear activation function. Finally, the attention coefficients α_{ij} are obtained through softmax normalization. The above calculation process is expressed as Eq. (3).

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(a(Wh_i\|Wh_j)))}{\sum_{k \in N} \exp(\text{LeakyReLU}(a(Wh_i\|Wh_k)))}. \quad (3)$$

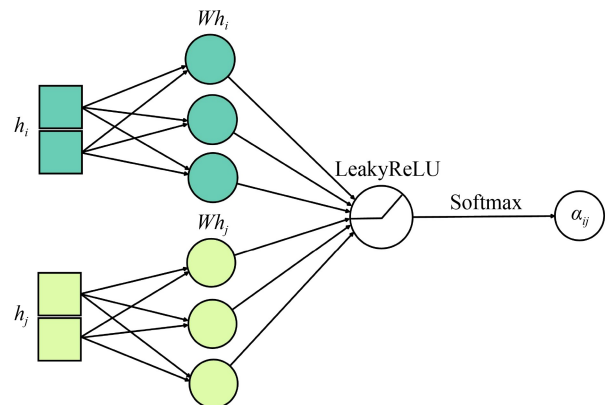


Fig. 1 Calculation of the attention coefficient.

2.2.2 Feature aggregation

Enhanced through feature combination and nonlinear activation function, the aggregation feature of the central node is obtained as follows:

$$h'_i = \text{Sigmoid} \left(\sum_{j \in N} \alpha_{ij} W h_j \right), \quad (4)$$

where Sigmoid means the activation function of this graph attention layer (GAL), which is used to achieve nonlinear transformation.

2.2.3 Multi-head attention mechanism

The MHGAT integrates several attention heads to concurrently calculate the aggregated feature h''_i from different attention scales. A more comprehensive and abundant node representation can then be obtained, as illustrated in Fig. 2.

$$h''_i = \parallel_{L=1}^L \text{Sigmoid} \left(\sum_{j \in N} \alpha_{ij}^L W^L h_j \right), \quad (5)$$

where $\parallel_{L=1}^L$ denotes the feature concatenation operation, with L attention scales. α_{ij}^L signifies the attention coefficients of the L th attention head, and W^L denotes the weight matrix of the corresponding attention head.

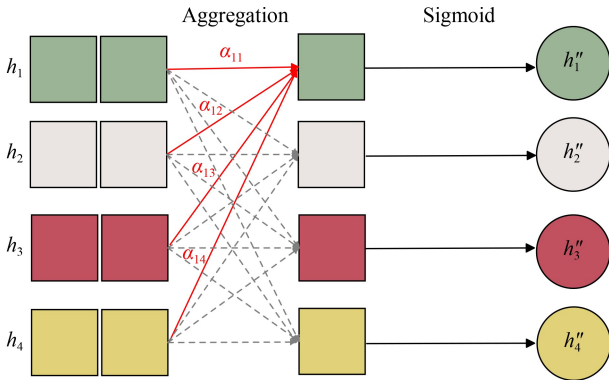


Fig. 2 Information aggregation of MHGAT.

3 Proposed fault diagnosis method

The main procedure consists of four steps: node feature extraction, distance matrix calculation, edge connection establishment, and fault diagnosis based on DWGAT. The details are shown in Fig. 3. The proposed method can be summarized into two parts: the proposed AAKNN graph (AAKNNG) construction and fault diagnosis based on DWGAT.

3.1 Proposed AAKNNG construction

3.1.1 Node feature embedding

The nodes within the AAKNNG represent the collected samples, and the node features V are embedded from the raw time-domain signals.

3.1.2 Adaptive edge connection construction

After node embedding, the next step is to establish reliable edge connections between nodes. A novel adaptive edge adjustment method is proposed to find the proper number of neighbors for different nodes. This method can be divided into two stages.

(1) Preliminary construction of edge connections based on DFW

In contrast to the conventional Euclidean distance-based metric approach, we propose a DFW method to measure the similarity between nodes and assign edge connections to highly similar nodes. DFW not only considers the distance between nodes but also takes into account the interrelationships between frequency nodes in the sample, thereby being more appropriate for capturing local similarity with fault features in the samples. Furthermore, DFW exhibits a certain degree of robustness to noise and deformation in the samples, enabling it to adapt to minor deformations and distortions in nodes.

Fault features are usually more pronounced in the frequency domain. Thus, fast Fourier transform (FFT) needs to be used for subsequent edge connection adjustments.

$$F_i = |\text{FFT}(X_i)|^2, \quad i = 1, \dots, m, \quad (6)$$

where $X_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,2n}\}$ denotes the i th sample. After FFT, the result is a centrosymmetric spectrum and the half spectrum $F_i = \{f_{i,1}, f_{i,2}, \dots, f_{i,n}\}$ is intercepted. Then, the obtained two groups of node features, F_i and F_j , are computed by using the Euclidean distance method to obtain the value of $D \in \mathbb{R}^{m \times m}$. The calculation equation is provided below.

$$D_{ij} = \sqrt{(F_i - F_j)^2}, \quad i, j = 1, \dots, m. \quad (7)$$

In the next step, we accumulate the distance matrix D and compute the minimum distance from the starting point to each point, which is then represented by the cumulative distance matrix C . The equation for calculating $C(a,b)$ can be expressed as

$$C(a,b) = D(a,b) + \min\{C(a-1,b), C(a,b-1), C(a-1,b-1)\}. \quad (8)$$

As shown in Fig. 4, in the cumulative distance matrix C , the value of $C(n,n)$ in the upper right corner denotes the minimum cumulative distance from the starting point to $f_{i,n}$ and $f_{j,n}$. To minimize the cumulative distance, the

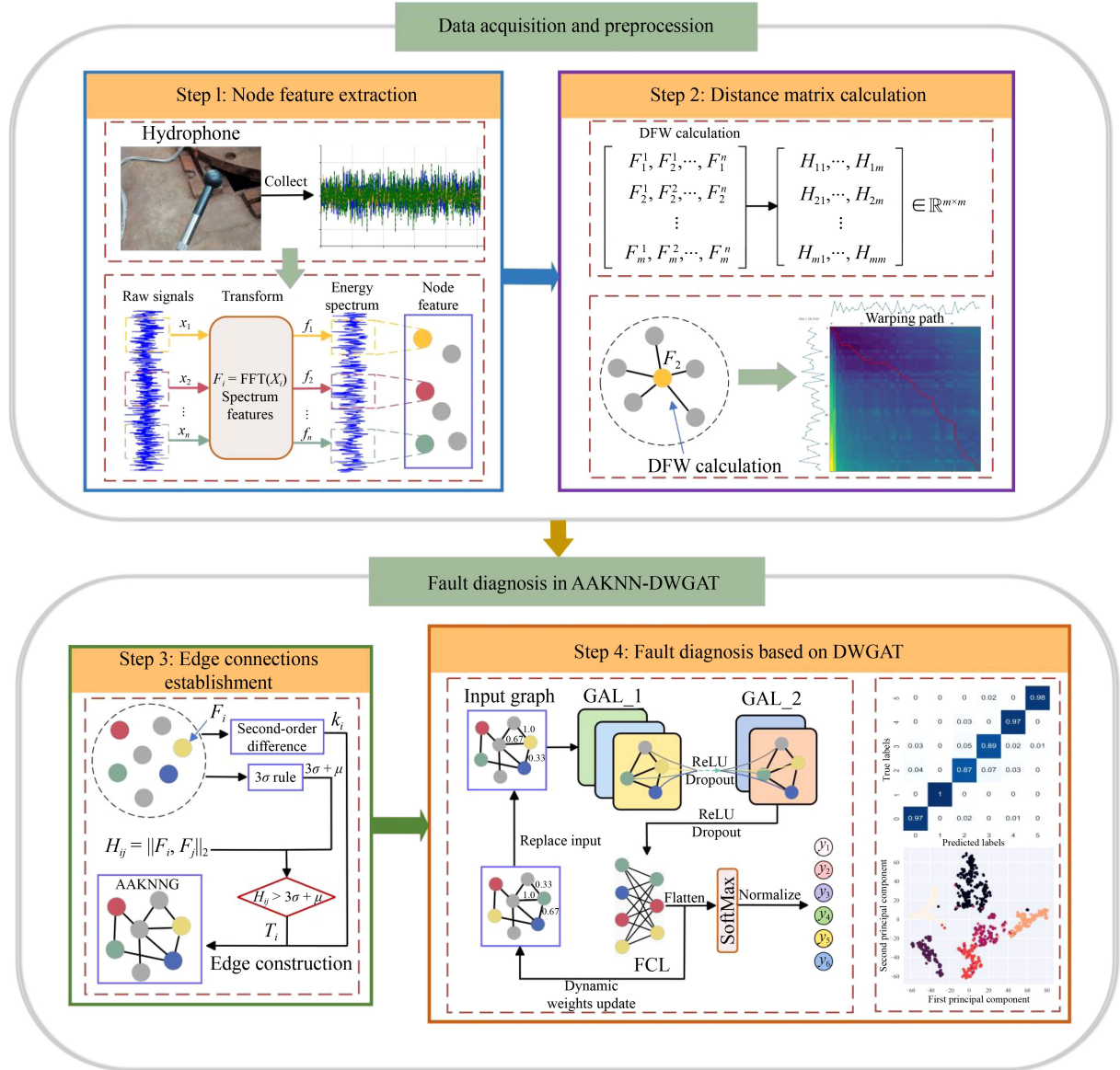


Fig. 3 Procedure of the proposed method.

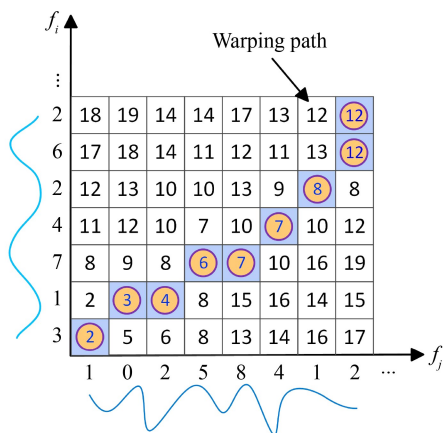


Fig. 4 Schematic of the DFW calculation.

warping path starts from $C(n,n)$, and then moves down, left, and bottom-left along the minimum value in C until $C(1,1)$ is reached, thereby obtaining the sum of the distances on the shortest backtracking path, which determines the similarity H_{ij} between the two samples. Subsequently, we extend the above calculation to m samples and obtain the frequency similarity matrix (FSM) H , where $H \in \mathbb{R}^{m \times m}$.

(2) Adaptive edge adjustment

With the differences between node attributes, they have different dependence degrees on neighborhood information. In this situation, a fixed number of neighbors would cause a redundant graph structure and introduce uncertain information. Therefore, an adaptive edge adjustment method is presented to remove the needless edge connections for each node, which consists of the

steps below.

First, to identify the outlier nodes whose deviations from any other node exceed the normal range, we employ the 3σ criteria to determine the outlier nodes. We initially sort the FSM and take the nonzero minimum similarity between each node and the other nodes to obtain the nonzero minimum similarity matrix $H^N = \{H_{11}^N, H_{21}^N, \dots, H_{m1}^N\}$. According to the 3σ criterion, the node can be identified as an outlier or abnormal situation if its deviation surpasses the range of the mean plus or minus 3 standard deviations, as described in Eq. (9).

$$T_i : \forall |H_{i1}^N - H_{j1}^N| > \mu \pm 3\sigma, j \in \{1, 2, \dots, m\}, \quad (9)$$

where T_i denotes the calculated abnormal node, H_{i1}^N represents the i th node's nonzero minimum similarity, H_{j1}^N represents any other node's nonzero minimum similarity, and σ is the standard deviation of the nonzero minimum similarity among all samples. By applying this criterion, we can obtain outlier nodes that demonstrate significant differences from any other node. These outlier nodes will be isolated and excluded from the edge construction process.

Second, a second-order difference method is employed to apply adaptive edge adjustment to the remaining nodes. The second-order difference can apply two discrete differences to data, which effectively converts a node's sequence of similarity with other nodes into a representation of its rate of change or gradient and identifies the inflection points of similarity distribution, thereby obtaining more abundant node information. After H is acquired, the similarity between each node and other nodes is initially ordered to procure the ordered similarity matrix denoted as H' . Subsequently, second-order difference processing is carried out on H' . Calculations are shown as follows:

$$S_i = \frac{\partial^2 H'}{\partial x^2} = H'_i(x+1) + H'_i(x-1) - 2H'_i(x), \quad x \in 2, 3, \dots, m-1, \quad (10)$$

where S_i denotes the second-order difference result of the i th node, and $H'_i(x)$ represents the ordered frequency similarity matrix between the i th node and x th node. The calculation of the second-order difference matrix displays the trend of similarity changes between the i th node and every other node. Subsequently, the inflection point, at which the similarity changes between nodes are more abrupt, can be identified, i.e., the minimum value point of the second-order difference matrix. In our method, the location (point index) of the inflection point is used as the setting number (k_i) of edges for this central node i . Simultaneously, k_{\max} is established to prevent excessively high k_i values for specific nodes. Finally, we construct the adaptive edges (E^k) of the nodes on the basis of k_i and k_{\max} . The proposed method can effectively and adaptively identify the optimal number of edge connections for each

central node.

Edge weights are incorporated in the construction of AAKNNG to highlight the priority of edge connections, thereby underscoring the significant degree of relationships between nodes. Preliminarily, the edge weights are calculated by Eq. (11), which will be further updated in the subsequent graph feature learning process.

$$w_{ij} = 1 - \frac{j}{k_i}, j = 0, 1, \dots, k_i - 1, \quad (11)$$

where j denotes any neighbor of node i . With increasing distance, the weight of the edge diminishes gradually in a linear trend. Varied weights on the edges signify distinct levels of feature extraction for neighbor nodes. Larger weights facilitate greater fault feature acquisition, while nodes with smaller weights exhibit the opposite effect, thereby reducing the impact of incorrect edge connections and enhancing the robustness of the proposed AAKNNG algorithm to a certain extent. In summary, the proposed AAKNNG can be described in Eq. (12).

$$G^A = \{V, E^k, w_{ij}\}. \quad (12)$$

3.2 Fault diagnosis based on dynamic-weighted MHGAT

A dynamic-weighted MHGAT is developed with the input of AAKNNG to decrease the influence of noise interference on the input graph. This network comprises two GALs. The first GAL utilizes larger attention heads to capture global information, whereas the second GAL employs smaller attention heads to focus on local fault features. Each GAL integrates features extracted from various attention heads to enhance feature representation capabilities. Subsequently, the node features are flattened and fed into a fully connected layer (FCL). Finally, the fault labels are derived from the softmax classifier.

Despite the adjustment of redundant edge connections on the input graph through the proposed adaptive edge adjustment method, some inappropriate edge connections are still unavoidable. Therefore, we introduce a dynamic weighting strategy to mitigate the interference of improper edges on feature extraction by adjusting edge weights, thereby decreasing the impact of noise interference on the input graph. Specifically, in the model training process, the edge weights would be adjusted on the basis of the output high-level features periodically. As shown in Fig. 5, the detailed implementation steps are as follows:

(1) With the initial input graph $G^A = \{V, E^k, W_{ij}\}$, the model training is started and the parameters will be automatically updated using the backpropagation algorithm.

(2) The fully connected output features F^A are recorded for every t_1 epoch during the training process.

(3) Every t_2 epoch, the collected sets of F^A (the amount is t_2/t_1) are input into the edge construction process and

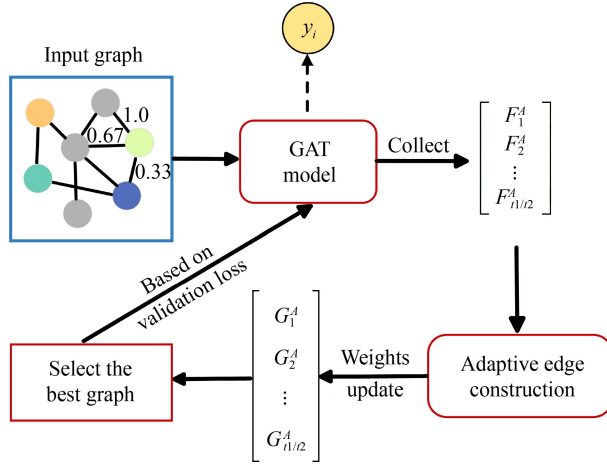


Fig. 5 Dynamic weight updating process.

recalculate the weights, thereby obtaining t_2/t_1 updated graphs. The calculations are conducted using Euclidean distance to minimize the computational time required for F^A similarity calculation, with the edge construction and weight calculation methods aligned with those outlined in Section 3.1.2. The updated weights can be presented as

$$W_i^s = \{w_{i1}, w_{i2}, \dots, w_{ik_i}\}, \quad (13)$$

where W_i^s denotes the weight of the central node i in the s th output features F^A .

(4) The obtained t_2/t_1 updated graphs are fed back into DWGAT, and the best one is chosen according to the validation loss. Then, the model training will continue until the maximum iterations t are reached.

Notably, during the model training process, we only adjust the edge weights, which does not change the structure of the input graph. This approach also avoids the frequent graph construction process and saves computation time. The pseudocode of the proposed AAKNN-DWGAT is presented in Algorithm 1.

4 Case study

Two experiments are conducted to validate the effectiveness of the proposed AAKNN-DWGAT, using the axial flow pump dataset and the XJTUGearbox dataset, respectively. All algorithms in this study are written in Python 3.7 and run with PyTorch 2.0.1 on an Intel® Core™ i7-10875H processor with 16 GB of RAM. The learning rate, batch size, and training epochs are set to 0.005, 32, and 500, respectively. Additionally, the Adam optimizer is utilized.

4.1 Fault diagnosis on the axial flow pump

4.1.1 Data description

The test bench for the axial flow pump primarily comprises a pump body (700ZLB-70), an electromotor (750 r·min⁻¹/160 kW/380 V), a hydrophone (RHS-30), an acquisition box, and a circulating pool with a volume of 10000 m³, as shown in Fig. 6. A hydrophone is installed near the underwater inlet pipe of the axial flow pump collect useful signals, with the sampling frequency being

Algorithm 1 Proposed fault diagnosis method based on AAKNN-DWGAT

Input: raw signals X .

Output: fault types Y .

1. Calculate the feature matrix F : $F = |\text{FFT}(X)|^2$.
 2. Calculate the DFW matrix H : $H_{ij} = \|F_i, F_j\|_2$.
 3. Define the threshold for abnormal samples by the 3σ rule.
 4. Obtain outliers T : $T_i = \{H_{i1}^N - H_{j1}^N > \mu + 3\sigma\}$.
 5. Calculate the second-order difference S : $S_i = H'_i(x+1) + H'_i(x-1) - 2H'_i(x)$.
 6. Find the first local minimum indices k_i .
 7. Construct the adaptive edges of the nodes E^k .
 8. Calculate the weight of edges w_{ij} .
 9. Construct $G^A = \{V, E^k, w_{ij}\}$.
 10. Establish DWGAT model;
 11. Train DWGAT model:
 12. **for** t in epochs **do**
 13. **for** t_1 in epochs **do**
 14. Obtain fully connected output features F^A : $F^A = \text{FCL}(\text{GAL2}(\text{GAL1}(X)))$.
 15. **for** t_2 in epochs **do**
 16. Update weights.
 17. Obtain Y : $Y = \text{softmax}(F^A)$.
 18. Compute loss: $\text{CE} = -\sum Y \log \hat{Y}$.
 19. Update with back propagation.
 20. **end for**
 21. **Output** Y .
-

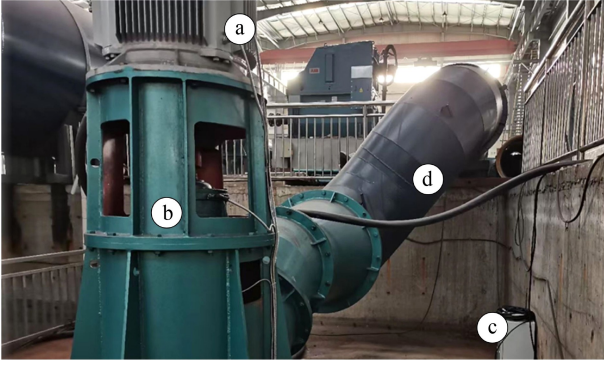


Fig. 6 Test platform of the axial flow pump: a. Electromotor; b. Pump body; c. Acquisition box; d. Outlet pipe.

8192 Hz. To simulate real-world scenarios, we manually create five different types of faults: loose base, rotor imbalance, rotor misalignment, blade cracks, and impeller winding. The details are as follows: 1) The loose base fault is constructed by intentionally loosening bolts. 2) The rotor unbalance faults are created by appending unbalance blocks to the rotor. 3) The rotor misalignment fault is induced by adjusting the deviation angle between the centerlines of the two half couplings. 4) Blade cracks are introduced through the utilization of wire-cutting technology. 5) The impeller winding fault develops through the process of winding foam products. As a result, we acquire raw signals from six distinct conditions, including the normal state and the previously mentioned five fault types. These signals are all used for the construction of the sample set.

4.1.2 Construction of sample set and settings of model hyperparameters

With a sample length of 1024, 1400 samples are obtained. The amount ratio of the normal state and each fault state is 2:1 to simulate the situation of unbalanced data. All samples are divided randomly into three parts: the training set, validation set, and testing set, with an amount

ratio of 4:1:5. Subsequently, the samples are transformed into frequency-domain information through FFT, leading to the calculation of the similarity matrix. Then, the second-order difference is calculated. The results for two of the samples are shown in Fig. 7.

Subsequently, the proposed AAKNNG and DWGAT are constructed on the basis of the steps of Section 3. In the DWGAT model, the parameters t_1 , t_2 , and t are set to 10, 100, and 500, respectively. The graph feature learning hyperparameters of DWGAT are shown in Table 1.

4.1.3 Fault diagnosis based on AAKNN-DWGAT

To demonstrate the superiority of the proposed method under the same graph structural complexity, we conducted performance comparisons between AAKNNG and KNNG with the same edge amount. With adjusted hyperparameters of k_{\max} and k , comparative experiments were conducted at the scale of approximately 12600, 14000, 15400, 16800, and 18200 edges, with corresponding k_{\max} values of 11, 12, 13, 14, and 15, and k values of 9, 10, 11, 12, and 13, respectively. The classification models were all the proposed DWGAT. The time consumed involves graph construction, model training, validation, and testing.

Figure 8 compared AAKNNG and KNNG under different edge numbers. The accuracy for AAKNNGs averages 94.26%, 94.61%, 94.83%, 95.13%, and 95.42%, respectively, indicating significant improvements in diagnostic capabilities over KNNG. This improvement is attributed to the proposed AAKNNG's optimization of graph structure and reduction of redundant edge connections. In terms of time consumption for graph construction, AAKNNG requires approximately 1.53 s, whereas KNNG consumes 0.21 s only, leading to higher time consumption for the proposed AAKNN-DWGAT, as illustrated in Fig. 8(b). However, the time difference is almost negligible as the time spent on graph construction constitutes only a small part of the total time. Moreover, with an increase in the number of edge connections, the

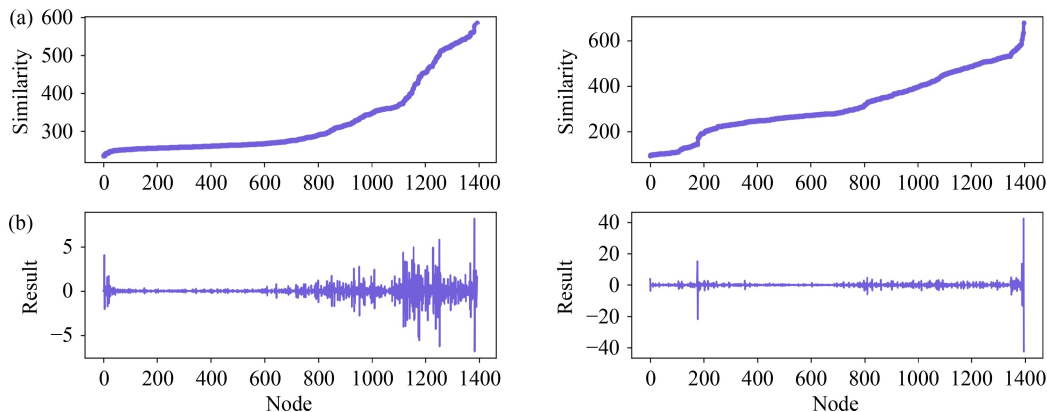


Fig. 7 Distribution of second-order difference for two of the samples: (a) ordered FSM curve and (b) 2nd-order difference curve.

Table 1 Model structure of DWGAT for the dataset

Component	Layer	Attention heads	Input size	Output size
Graph feature learning	GAL_1	16	1400×1024	1400×256
	GAL_2	8	1400×256	1400×64
	FCL	–	1400×64	1400×6

diagnosis accuracies for both graphs are all improved, but the calculation time is significantly increased. To further describe the differences in performance between these two graph construction methods, Fig. 9 gives the confusion matrices for using AAKNNG and KNNG with two amount settings of edge connections.

4.1.4 Effect of the head numbers on the model

To investigate the impact of the number of heads on the proposed DWGAT, additional experiments were carried out. For GAL_1, the head numbers were set as [8,10,16,32]; for GAL_2, the head numbers were set as [6,7,8,9,10]. The number of edges in the graph is estimated to be approximately 11200. Figures 10 and 11 illustrate the outcomes of the supplementary experiments with varied head configurations, revealing that the model achieves optimal performance when GAL_1 is set to 16 and GAL_2 to 8. With fewer heads, there is limited fusion of information across different scales, whereas an excessive number of heads results in an overload of information fusion, leading to reduced accuracy and prolonged training time. Therefore, we opted for head numbers 16 and 8 as the model's hyperparameters.

4.1.5 Comparison between two similarity calculation methods

A simple experimental comparison was conducted in this section to further validate the superiority of the proposed DFW over Euclidean distance. The similarity of node edges was calculated by DFW and Euclidean distance

respectively, with a total of 16800 edges. As illustrated in Fig. 12, the edge similarity calculation strategy employing DFW demonstrates superior accuracy, reduced loss, and accelerated convergence when compared with the Euclidean distance method. This finding suggests that the application of DFW in computing frequency-domain signals more effectively captures fault features and reflects the improvement effect of the proposed method.

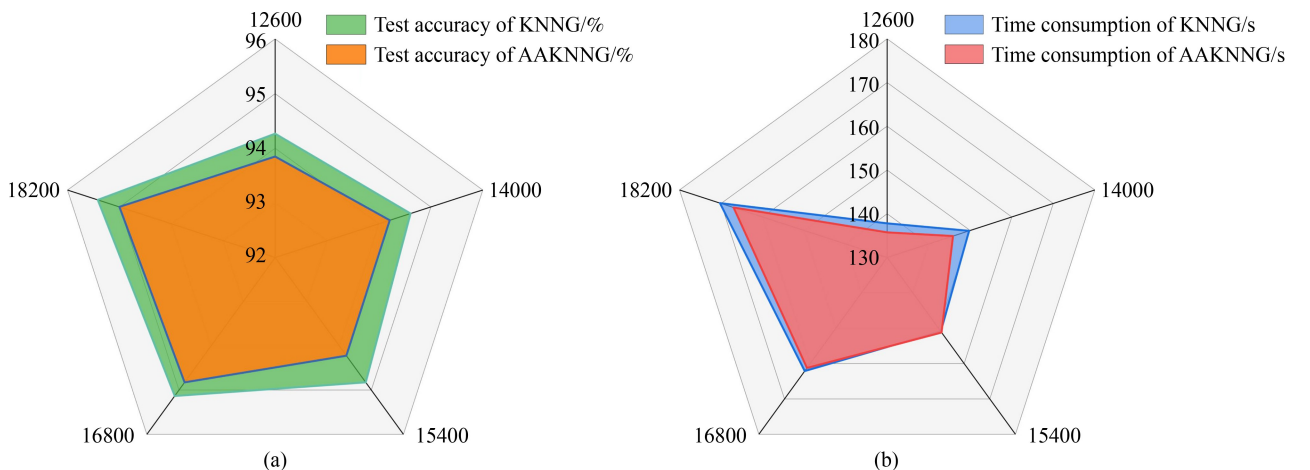
4.1.6 Comparison with other graph construction methods

Some state-of-the-art (SOTA) graph construction methods under varying signal-to-noise ratios (SNR) were used for comparison to explore the noise robustness of the proposed graph construction method. The definition of SNR is provided in Eq. (14). Five graph construction methods were used for comparison, namely, KNNG [28], dynamic-weighted graph (DWG) [36], affinity graph (AG) [37], the SuperGraph [17], and the proposed AAKNNG. They all employed DWGAT for feature learning and classification. As depicted in Fig. 13, at higher SNRs (e.g., 5 and 10 dB), the accuracies of AAKNNG and KNNG are very close, while the other three methods exhibit lower accuracy compared with these two methods. However, as the SNR decreases further, the advantage of AAKNNG becomes more pronounced, clearly surpassing the other four graph construction methods. Thus, the noise robustness of the proposed AAKNNG is verified.

$$\text{SNR}_{\text{db}} = 10 \log_{10} \frac{P_{\text{signal}}}{P_{\text{noise}}}. \quad (14)$$

4.1.7 Comparison with SOTA fault diagnosis methods

A simple comparative experiment was conducted in this section to verify the improvement effect of the proposed method compared to other SOTA fault diagnosis

**Fig. 8** Fault diagnostic results of two graph construction methods: (a) comparison of test accuracy; (b) comparison of consuming time.

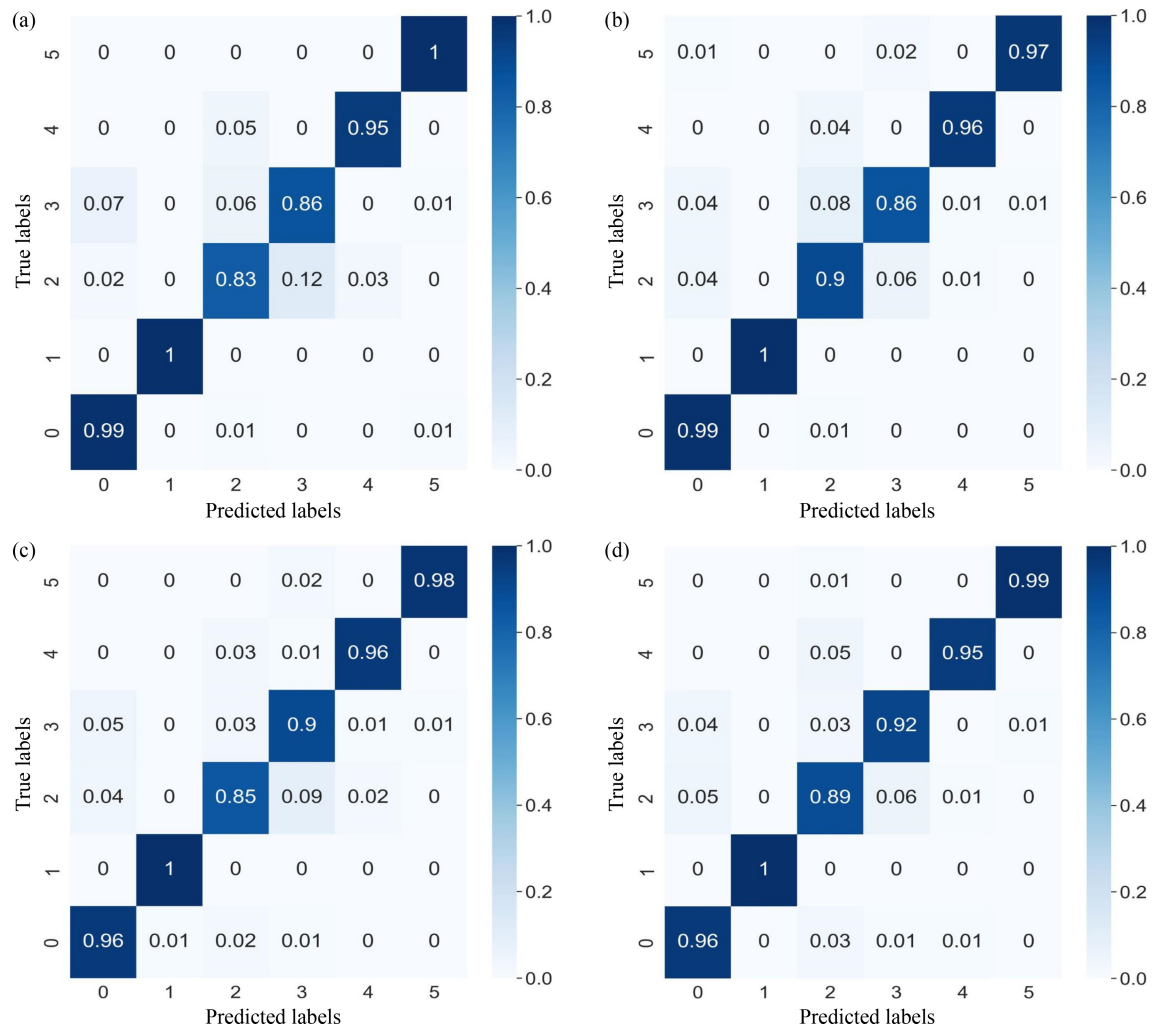


Fig. 9 Confusion matrices under different numbers of edge constructions: (a) KNNG with 14000 edges; (b) KNNG with 18200 edges; (c) AAKNNG with 14000 edges; (d) AAKNNG with 18200 edges.

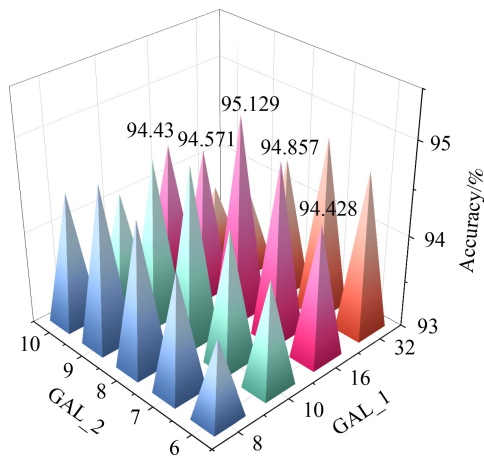


Fig. 10 Diagnosis accuracy with different number of heads.

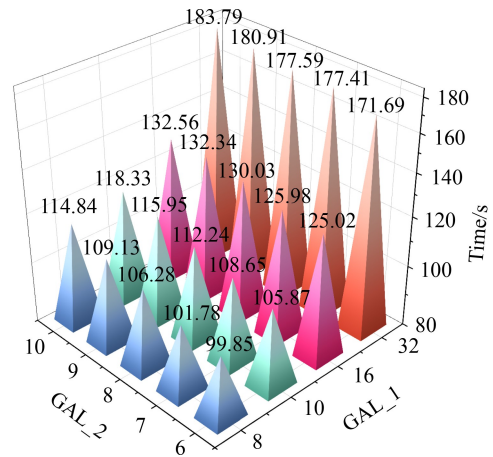


Fig. 11 Time consumption with different number of heads.

methods. The proposed method was compared with AG+MRF-GCN [37], SuperGraph+GCN [17], KNNG+MHGAT [28], and FDKNN-DGAT [31]. The detailed

hyperparameter settings are shown in Table 2, and K is the size of Chebyshev filter in GCN. As illustrated in Fig. 14, most of these methods related to GNN demonstrate

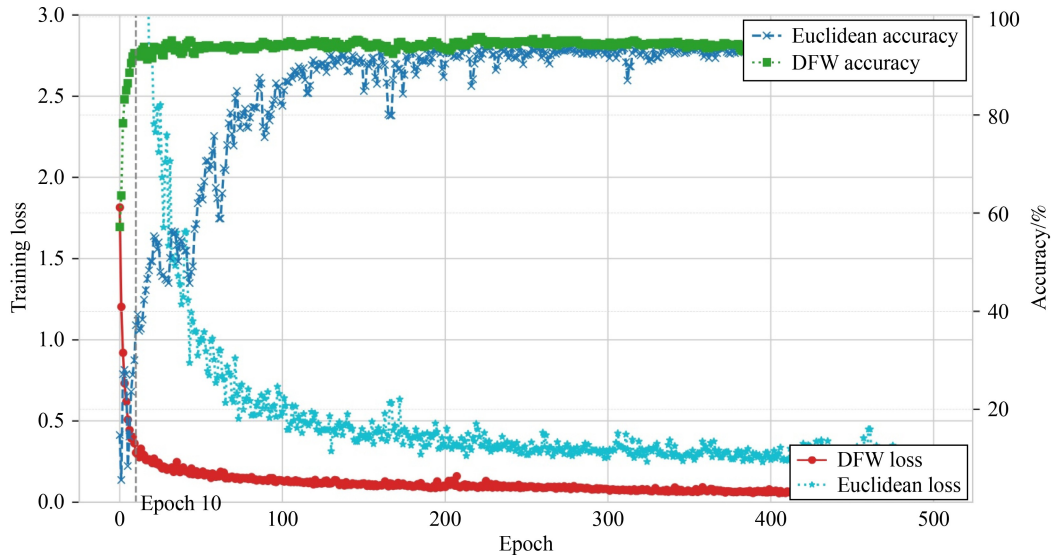


Fig. 12 Comparison of two similarity calculation methods (DFW and Euclidean distance).

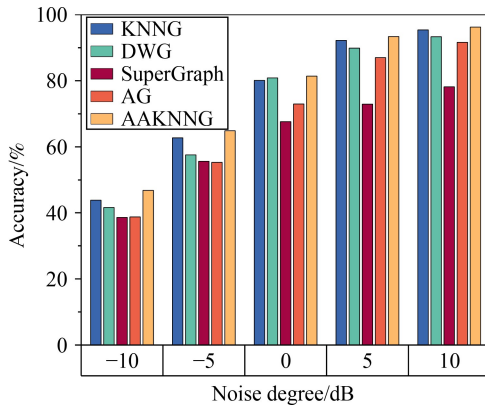


Fig. 13 Comparison of graph construction methods under different noisy conditions.

high accuracy. In comparison with other fault diagnosis methods, the proposed fault diagnosis method achieves the highest test accuracy of more than 95.0% and exhibits a lower variance, reflecting its superiority.

4.1.8 Comparison with other classical fault diagnosis methods

Some classical deep learning methods were used for comparison to validate the superiority of the proposed

AAKNN-DWGAT. They involved one-dimensional convolutional neural network (1DCNN) [38], GCN [37], multi-layer perceptron (MLP), and single-head GAT (SHGAT). In addition, four sample sets with different amount ratios of faulty samples were constructed to validate the diagnosis performance of these methods. The detailed settings of the sample sets and hyperparameters of methods are shown in Tables 3 and 4, respectively.

The comparison result from Fig. 15 indicates that, across different sample sets, the proposed DWGAT model outperforms other deep learning models. Additionally, the four GNN-related methods demonstrate significant advantages over the other two methods. Moreover, compared with KNNG, the GNN-related methods employing AAKNNG as input show better fault diagnosis results for imbalanced data. Notably, as the imbalance level of the data sets increases, the decreased rates of accuracy for GNN-related methods are significantly slower than that of MLP and 1DCNN, which also reflects the advantages of GNNs in handling imbalanced data.

4.2 Fault diagnosis on the XJTUGearbox

4.2.1 Data description

The experimental setup, as shown in Fig. 16, comprises

Table 2 Hyperparameter settings of the different fault diagnosis methods

Method	Parameter configuration
AG+MRF-GCN	Input size: 1400×512 , $\varepsilon = 0.994$, structure of GCN: 512-300-100, fully connected layer: 100-6, $K = 3, 2, 1$
SuperGraph+GCN	Input size: 1400×33 , structure of GCN: 33-30-20, fully connected layer: 20-6, $K = 3, 3, 3$
FDKNN-DGAT	Input size: 1400×64 , $k = 10$, structure of GAT: 64-32, attention heads: 4, 4, fully connected layer: 32-6
KNNG+MHGAT	Input size: 1400×1024 , $k = 10$, structure of GAT: 64-32, attention heads: 16, 8, fully connected layer: 32-6
AAKNN-DWGAT	Input size: 1400×1024 , $k_{\max} = 12$, structure of GAT: 64-32, attention heads: 16, 8, fully connected layer: 32-6, t_1, t_2, t : 10, 100, 500

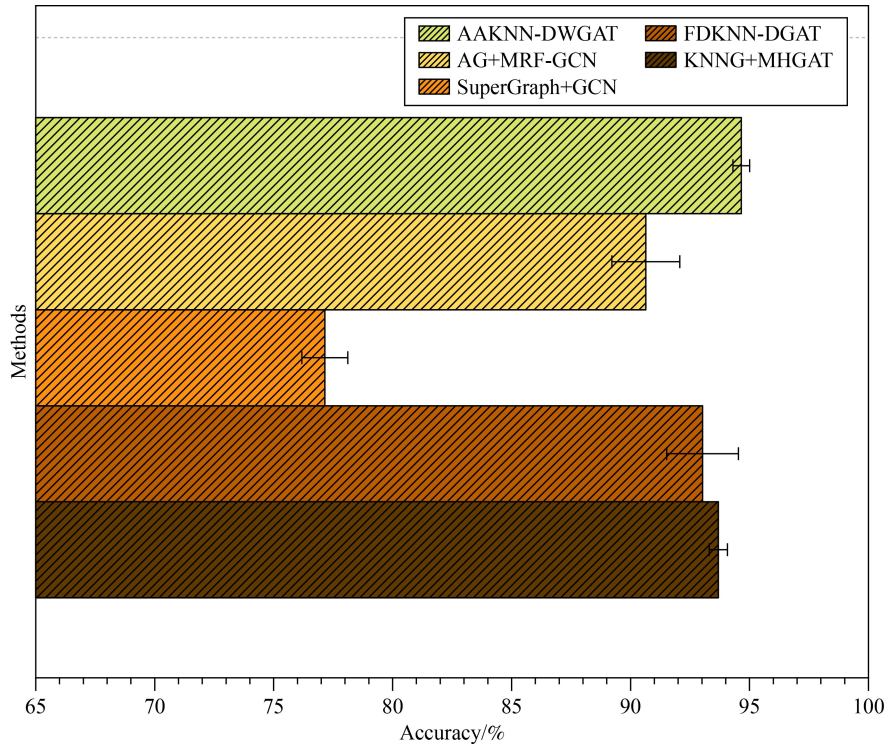


Fig. 14 Comparison under different fault diagnosis methods.

Table 3 Number of samples for four sample sets^{a)}

Sample set	Training		Validation	Test
	Normal	Fault		
S0	160	80 × 5	140	700
S1	160	40 × 5	140	700
S2	160	20 × 5	140	700
S3	160	16 × 5	140	700

a): “80 × 5” denotes that there are 80 samples for each of the five fault states.

Table 4 Hyperparameter settings of the different fault diagnosis methods

Method	Parameter configuration
KNNG+DWGAT	$k = 10$, t_1 , t_2 , t : 10, 100, 500, attention heads: 16, 8
1DCNN	num_convolution_kernel = 3, pool_size = 2
MLP	hidden_shape: 100-50
KNNG+SHGAT	$k = 10$, attention heads = 1
KNNG+GCN	$k = 10$, structure of GCN: 512-300-100
AAKNN-DWGAT	$k_{max} = 12$, t_1 , t_2 , t : 10, 100, 500, attention heads: 16, 8

components such as a drive motor, a controller, a planetary gearbox, a parallel gearbox, and a brake. The motor is a three-phase 3 HP motor, powered by three-phase AC (230 V, 60/50 Hz) [39]. It operates at a speed of 1800 r/min, with a sampling frequency of 20480 Hz. To collect the useful fault signals, four failure modes of planetary gears and four failure modes of bearings are intentionally constructed on the planetary reducer during the experiment. The bearing faults are pre-set on the first

stage planetary gear of the planetary reducer, while the gear faults are pre-set on the second stage planetary gear of the planetary reducer. The gear failures include tooth surface wear, missing teeth, root cracks, and tooth broken. Bearing faults include ball faults, inner race faults, outer race faults, and a combination of the three types of bearing faults mentioned above. Therefore, nine types of vibration signals are collected, including the normal state and the eight types of faults mentioned above. The partial fault states are depicted in Fig. 17.

4.2.2 Construction of sample set and fault recognition results

The XJTUGearbox dataset contains a total of 450 samples, with 50 samples for each state. They were divided randomly into three sets: the training set, validation set, and testing set, with an amount ratio of 4:1:5. Unlike in Section 4.1.2, we selected different sample lengths, with sample sizes of 512, 1024, and 2048, respectively. After the dataset division, the proposed AAKNN-DWGAT was constructed and compared with 1DCNN, MLP, SHGAT, GCN, and MHGAT. The hyperparameter settings of the different methods are referred to in Table 4, and the output results are shown in Table 5.

As shown in Table 5, the proposed AAKNN-DWGAT consistently achieves the highest diagnosis accuracy with different signal lengths, with an average accuracy exceeding 85%, 94%, and 99%, respectively.

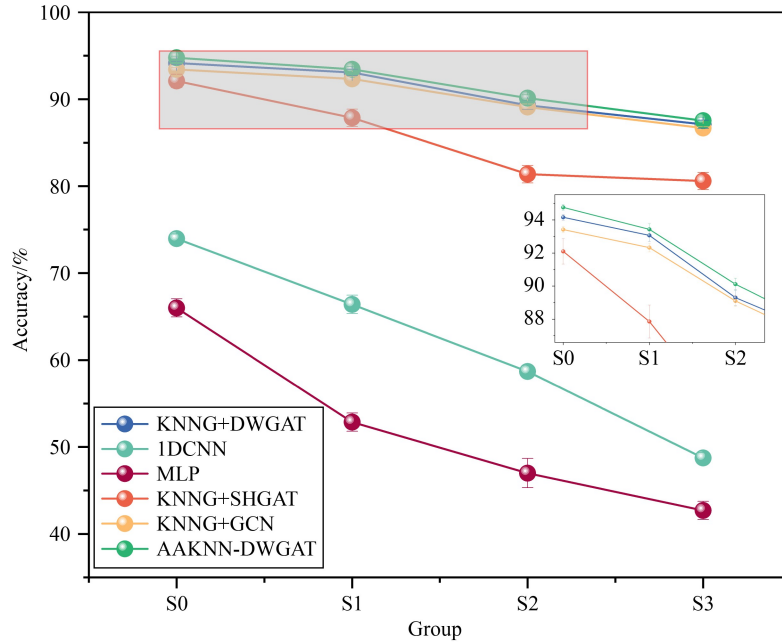


Fig. 15 Classification accuracy under different training samples.

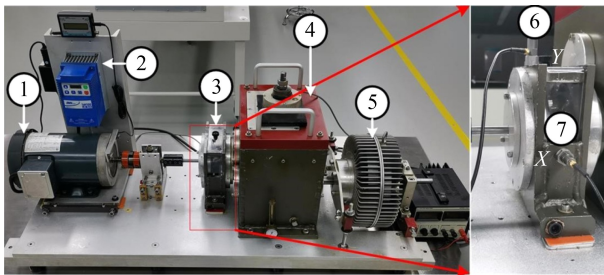


Fig. 16 Test rig of XJTUGearbox dataset. 1: motor; 2: controller; 3: planetary gearbox; 4: parallel gearbox; 5: brake; 6: accelerometer in X-direction; 7: accelerometer in Y-direction.

Furthermore, the proposed method exhibits lower variance, indicating the stability of its diagnostic results. Additionally, the time consumption of AAKNN-DWGAT is slightly longer than that of the other method, but it is still within the acceptable range. Table 5 demonstrates that an improper sample length of 512 significantly diminishes diagnosis accuracy, whereas an excessively lengthy sample size of 2048 overly captures fault features of nodes, making it difficult to reveal the superiority of different fault diagnosis methods. Compared with traditional IDCNN and MLP, GNN-related methods achieve better diagnosis performance, especially for the proposed AAKNN-DWGAT. As a result, the

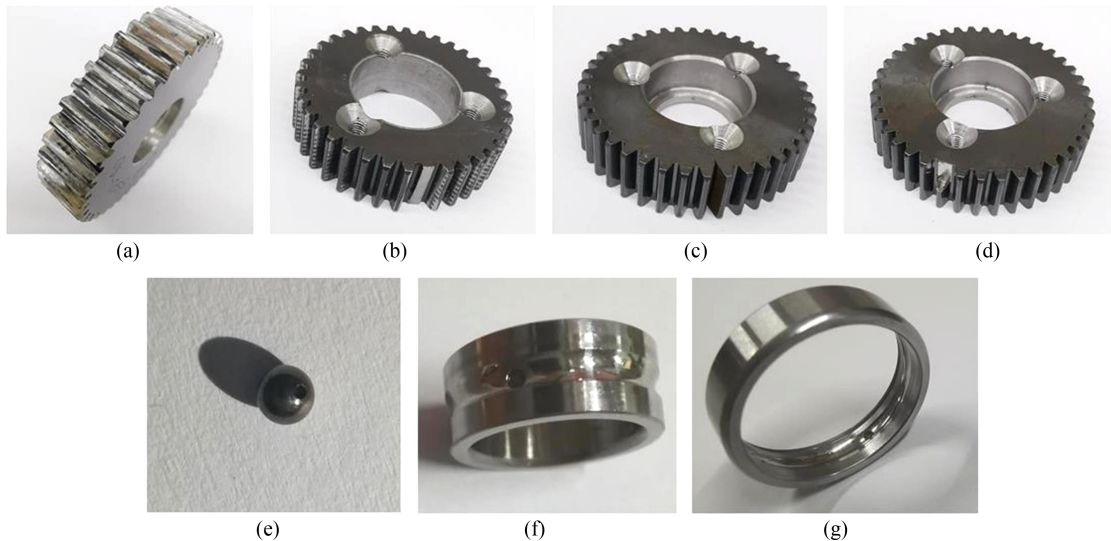


Fig. 17 Fault states of gears and bearings: (a) tooth surface wear, (b) missing teeth, (c) root cracks, (d) tooth broken, (e) ball faults, (f) inner race faults, and (g) outer race faults.

Table 5 Results on the XJTUGearbox dataset^{a)}

Graph	Structure Model	Time/s	Test accuracy/%		
			512	1024	2048
KNNG	SHGAT	20.03	82.846 ± 1.62	91.982 ± 0.89	98.236 ± 0.94
–	IDCNN	12.32	77.776 ± 1.18	77.110 ± 2.37	78.666 ± 1.41
–	MLP	0.68	74.400 ± 1.49	72.266 ± 1.59	72.534 ± 1.89
KNNG	GCN	14.47	83.646 ± 0.48	90.156 ± 0.54	98.668 ± 0.54
KNNG	MHGAT	80.93	83.922 ± 1.10	93.236 ± 0.46	99.062 ± 0.11
AAKNNG	DWGAT	85.20	85.010 ± 0.50	94.312 ± 0.36	99.199 ± 0.20

a): Time denotes the consuming time with the sample length of 1024.

effectiveness and superiority of the proposed AAKNN-DWGAT are verified.

5 Conclusions

This study proposes AAKNN-DWGAT for rotating machinery fault diagnosis. First, a DFW method is proposed for the assessment of node similarity, incorporating a second-order difference strategy to identify the inflection points of similarity distribution, thus establishing a weighted topological graph structure with adaptive edge adjustment. Second, a DWGAT model is constructed to adjust the edge weights on the basis of the output high-level features, thereby reducing the impact of incorrect edge connections to some extent. Finally, the effectiveness of the proposed method is validated by utilizing the axial flow pump dataset and the XJTUGearbox dataset. In comparison with other conventional and SOTA deep learning methods, the proposed method demonstrates superior fault recognition and noise robustness.

However, during the process of constructing the AAKNNG, the issue of edge connection adaptive adjustment unavoidably increases the computational time while improving the diagnosis accuracy. Such adjustment may introduce extra problems in some real-time fault diagnosis scenarios and should be further optimized in future research.

Nomenclature

Abbreviations

AAKNNG	Adaptive adjustment k -nearest neighbor graph
AAKNN-DWGAT	Adaptive adjustment k -nearest neighbor graph-driven dynamic-weighted graph attention network
AG	Affinity graph
ANN	Artificial neural networks
DFW	Dynamic frequency warping
DWG	Dynamic-weighted graph

DWGAT	Dynamic-weighted graph attention network
FCL	Fully connected layer
FFT	Fast Fourier transformation
FSM	Frequency similarity matrix
GAL	Graph attention layer
GAT	Graph attention network
GCN	Graph convolutional network
GNN	Graph neural network
IFD	Intelligent fault diagnosis
KNNG	k -nearest neighbor graph
MLP	Multi-layer perceptron
MHGAT	Multi-head graph attention network
SHGAT	Single-head graph attention network
SNR	Signal-to-noise ratios
SOTA	State-of-the-art
IDCNN	One-dimensional convolutional neural network

Acknowledgements This study was financially supported by the National Natural Science Foundation of China (Grant No. 51905396). We would like to express our sincere gratitude to the reviewers who contributed valuable comments to this paper.

Conflict of Interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution, and reproduction in any medium or format, as long as appropriate credit is given to the original author(s) and source, a link to the Creative Commons license is provided, and the changes made are indicated.

The images or other third-party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Visit <https://creativecommons.org/licenses/by/4.0/> to view a copy of this license.

References

1. Shao H D, Zhou X D, Lin J, Liu B. Few-shot cross-domain fault

- diagnosis of bearing driven by task-supervised ANIL. *IEEE Internet of Things Journal*, 2024, 11(13): 22892–22902
2. Lei Y G, Yang B, Jiang X W, Jia F, Li N P, Nandi A K. Applications of machine learning to machine fault diagnosis: a review and roadmap. *Mechanical Systems and Signal Processing*, 2020, 138: 106587
 3. Liu Z P, Zhang L. A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings. *Measurement*, 2020, 149: 107002
 4. Tang S N, Yuan S Q, Zhu Y. Deep learning-based intelligent fault diagnosis methods toward rotating machinery. *IEEE Access*, 2020, 8: 9335–9346
 5. Zhang T C, Chen J L, Pan T Y, Zhou Z T. Towards intelligent fault diagnosis under small sample condition via a signals augmented semi-supervised learning framework. In: *Proceedings of 2020 IEEE 18th International Conference on Industrial Informatics*. Warwick: IEEE, 2020, 669–672
 6. Luo J J, Shao H D, Lin J, Liu B. Meta-learning with elastic prototypical network for fault transfer diagnosis of bearings under unstable speeds. *Reliability Engineering & System Safety*, 2024, 245: 110001
 7. Fu D X, Liu J, Zhong H, Zhang X, Zhang F. A novel self-supervised representation learning framework based on time-frequency alignment and interaction for mechanical fault diagnosis. *Knowledge-Based Systems*, 2024, 295: 111846
 8. Fu Y, Liu Y, Yang Y. Multi-sensor GA-BP algorithm based gearbox fault diagnosis. *Applied Sciences*, 2022, 12(6): 3106
 9. Yang Z X, Wang X B, Wong P K, Zhong J H. ELM based representational learning for fault diagnosis of wind turbine equipment. In: *Proceedings of ELM-2015 Volume 2*. Cham: Springer, 2016, 169–178
 10. Kadam V, Kumar S, Bongale A, Wazarkar S, Kamat P, Patil S. Enhancing surface fault detection using machine learning for 3d printed products. *Applied System Innovation*, 2021, 4(2): 34
 11. Abdul Z K, Al-Talabani A K. Highly accurate gear fault diagnosis based on support vector machine. *Journal of Vibration Engineering & Technologies*, 2023, 11(7): 3565–3577
 12. Zhang X, Wang H F, Wu B, Zhou Q, Hu Y M. A novel data-driven method based on sample reliability assessment and improved cnn for machinery fault diagnosis with non-ideal data. *Journal of Intelligent Manufacturing*, 2023, 34(5): 2449–2462
 13. Li W H, Lan H, Chen J B, Feng K, Huang R Y. WavCapsNet: an interpretable intelligent compound fault diagnosis method by backward tracking. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 1–11
 14. Xing S B, Lei Y G, Wang S H, Jia F. Distribution-invariant deep belief network for intelligent fault diagnosis of machines under new working conditions. *IEEE Transactions on Industrial Electronics*, 2021, 68(3): 2617–2625
 15. Tang Y, Zhang X F, Zhai Y J, Qin G J, Song D Y, Huang S D, Long Z. Rotating machine systems fault diagnosis using semisupervised conditional random field-based graph attention network. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 1–10
 16. Chen H, Wang X B, Yang Z X. Adaptive semi-supervise graph neural network for fault diagnosis of tunnel ventilation systems. In: *Proceedings of 5th International Conference on System Reliability and Safety*. Palermo: IEEE, 2021, 53–57
 17. Yang C Y, Zhou K B, Liu J. SuperGraph: spatial-temporal graph-based feature extraction for rotating machinery diagnosis. *IEEE Transactions on Industrial Electronics*, 2022, 69(4): 4167–4176
 18. Wu Z H, Pan S R, Chen F W, Long G D, Zhang C Q, Yu P S. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(1): 4–24
 19. Li D P, Chen J X, Huang R Y, Chen Z Y, Li W H. Sensor-aware CapsNet: towards trustworthy multisensory fusion for remaining useful life prediction. *Journal of Manufacturing Systems*, 2024, 72: 26–37
 20. Salamat A, Luo X, Jafari A. HeteroGraphRec: a heterogeneous graph-based neural networks for social recommendations. *Knowledge-Based Systems*, 2021, 217: 106817
 21. Li Q M, Han Z C, Wu X M. Deeper insights into graph convolutional networks for semi-supervised learning. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. New Orleans: AAAI, 2018, 3538–3545
 22. Zhang D C, Stewart E, Entezami M, Roberts C, Yu D J. Intelligent acoustic-based fault diagnosis of roller bearings using a deep graph convolutional network. *Measurement*, 2020, 156: 107585
 23. Chen Z W, Xu J M, Peng T, Yang C H. Graph convolutional network-based method for fault diagnosis using a hybrid of measurement and prior knowledge. *IEEE Transactions on Cybernetics*, 2022, 52(9): 9157–9169
 24. Zhao L, Song Y J, Zhang C, Liu Y, Wang P, Lin T, Deng M, Li H F. A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 21(9): 3848–3858
 25. Chen Z W, Ke H B, Xu J M, Peng T, Yang C H. Multichannel domain adaptation graph convolutional networks-based fault diagnosis method and with its application. *IEEE Transactions on Industrial Informatics*, 2023, 19(6): 7790–7800
 26. Ai Z R, Cao H, Wang J H, Cui Z C, Wang L D, Jiang K. Research method for ship engine fault diagnosis based on multi-head graph attention feature fusion. *Applied Sciences*, 2023, 13(22): 12421
 27. Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. 2017, arXiv preprint arXiv:1710.10903
 28. Jiang L, Li X J, Wu L, Li Y B. Bearing fault diagnosis method based on a multi-head graph attention network. *Measurement Science & Technology*, 2022, 33(7): 075012
 29. Ding C Y, Sun S L, Zhao J. MST-GAT: a multimodal spatial-temporal graph attention network for time series anomaly detection. *Information Fusion*, 2023, 89: 527–536
 30. Zhang X, Liu J, Zhang X, Lu Y L. Multiscale channel attention-driven graph dynamic fusion learning method for robust fault diagnosis. *IEEE Transactions on Industrial Informatics*, 2024, 20(9): 11002–11013
 31. Tao H F, Shi H J, Qiu J E, Jin G H, Stojanovic V. Planetary gearbox fault diagnosis based on FDKNN-DGAT with few labeled data. *Measurement Science & Technology*, 2024, 35(2): 025036
 32. Yu X X, Tang B P, Deng L. Fault diagnosis of rotating machinery based on graph weighted reinforcement networks under small

- samples and strong noise. *Mechanical Systems and Signal Processing*, 2023, 186: 109848
33. Xie Z L, Chen J L, Feng Y, He S L. Semi-supervised multi-scale attention-aware graph convolution network for intelligent fault diagnosis of machine under extremely-limited labeled samples. *Journal of Manufacturing Systems*, 2022, 64: 561–577
 34. Zhang X, Jiang L, Wang L, Zhang T A, Zhang F. A pruned-optimized weighted graph convolutional network for axial flow pump fault diagnosis with hydrophone signals. *Advanced Engineering Informatics*, 2024, 60: 102365
 35. Xiao Y M, Shao H D, Wang J, Yan S, Liu B. Bayesian variational transformer: a generalizable model for rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing*, 2024, 207: 110936
 36. Zhang X, Hu Y M, Liu J, Zhang X, Wu B. Robust rotating machinery diagnosis using a dynamic-weighted graph updating strategy. *Measurement*, 2022, 202: 111895
 37. Li T F, Zhao Z B, Sun C, Yan R Q, Chen X F. Multireceptive field graph convolutional networks for machine fault diagnosis. *IEEE Transactions on Industrial Electronics*, 2021, 68(12): 12739–12749
 38. Ince T, Kiranyaz S, Eren L, Askar M, Gabbouj M. Real-time motor fault detection by 1-D convolutional neural networks. *IEEE Transactions on Industrial Electronics*, 2016, 63(11): 7067–7075
 39. Li T F, Zhou Z, Li S N, Sun C, Yan R Q, Chen X F. The emerging graph neural networks for intelligent fault diagnostics and prognostics: a guideline and a benchmark study. *Mechanical Systems and Signal Processing*, 2022, 168: 108653