

Article

Research Method for Ship Engine Fault Diagnosis Based on Multi-Head Graph Attention Feature Fusion

Zeren Ai ¹, Hui Cao ^{1,2,*}, Jihui Wang ¹, Zhichao Cui ¹, Longde Wang ¹ and Kuo Jiang ¹ 

¹ Marine Engineering College, Dalian Maritime University, Dalian 116026, China; azrzz0130@163.com (Z.A.); wjh921130@163.com (J.W.); czcdlmu@163.com (Z.C.); wld980411@163.com (L.W.); jk1120210083@dlmu.edu.cn (K.J.)

² Dalian Maritime University Smart Ship Limited Company, Dalian 116026, China

* Correspondence: lustersoft@foxmail.com

Abstract: At present, there are problems such as low fault data, insufficient labeling information, and poor fault diagnosis in the field of ship engine diagnosis. To address the above problems, this paper proposes a fault diagnosis method based on probabilistic similarity and rank-order similarity of multi-head graph attention neural networks (MPGANN) models. Firstly, the ship engine dataset is used to explore the similarity between the data using the probabilistic similarity of T_SNE and the rank order similarity of Spearman's correlation coefficient to define the neighbor relationship between the samples, and then the appropriate weights are selected for the early fusion of the two graph structures to fuse the feature information of the two scales. Finally, the graph attention neural networks (GANN) incorporating the multi-head attention mechanism are utilized to complete the fault diagnosis. In this paper, comparative experiments such as graph construction and algorithm performance are carried out based on the simulated ship engine dataset, and the experimental results show that the MPGANN outperforms the comparative methods in terms of accuracy, F1 score, and total elapsed time, with an accuracy rate of 97.58%. The experimental results show that the model proposed in this paper can still fulfill the ship engine fault diagnosis task well under unfavorable conditions such as small samples and insufficient label information, which is of practical significance in the field of intelligent ship cabins and fault diagnosis.



Citation: Ai, Z.; Cao, H.; Wang, J.; Cui, Z.; Wang, L.; Jiang, K. Research Method for Ship Engine Fault Diagnosis Based on Multi-Head Graph Attention Feature Fusion. *Appl. Sci.* **2023**, *13*, 12421. <https://doi.org/10.3390/app132212421>

Academic Editors: Ján Pitel', Tibor Krenicky and Kamil Židek

Received: 13 October 2023

Revised: 8 November 2023

Accepted: 14 November 2023

Published: 16 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: fault diagnosis; graph attention network; feature fusion; marine engines; machine learning

1. Introduction

As the main power source for ship navigation and power generation, ship engines occupy an important position in the field of ships. Due to its importance, if the engine operation failure cannot be found in time, it will cause huge economic losses or even casualties. Therefore, timely monitoring and diagnosis of the engine's working condition can effectively improve the reliability, economy, and safety of normal ship operation.

With the wave of the era of ship intelligence and the continuous development and improvement of artificial intelligence and algorithmic arithmetic power, more and more scholars apply artificial intelligence to the field of condition detection and fault diagnosis, and the fault diagnosis methods of ship engines also progress and develop. Fault diagnosis of ship engines faces challenges such as data scarcity, multiple fault types, data noise interference, and insufficient labeling information. In order to provide safety for normal ship navigation, Zhong et al. [1] introduced deep learning into ship diesel engine fault diagnosis and proposed a fault diagnosis method based on correlation distribution and deep confidence networks, which achieved good results. However, this type of method does not take into account the problem of imbalance in the ship fault data categories, resulting in some defects in the training and application of the model. Aiming at the above problems Hou et al. [2] first used principal component analysis (PCA) to reduce the

dimensionality of the data. Then, the sample size optimization algorithm (SSO) was used to solve the problem of data imbalance, and the three-dimensional Arnold mapping to the particle swarm optimization algorithm was introduced to improve the holding vector machine (SVM), which effectively improves the generalization ability of the classification model. Finally, the model is validated by example using the fault data of the fuel supply system, and the experimental results show that the method is able to alleviate the impact of unbalanced data. Zhong et al. [3] proposed a semi-supervised principal component analysis (SSPCA) to improve the diagnostic accuracy of the model by fusing labeled and unlabeled samples as an alternative to unsupervised learning, with better robustness to false alarms.

The above model has achieved good results in fault monitoring, but it can only determine whether the fault occurs or not and cannot diagnose the specific cause of the fault, which makes it difficult to play a guiding role for the subsequent operation and maintenance decision-making work. In this regard, Ren et al. [4] established a fault tree, transformed the fault tree into a Bayesian network, constructed a Bayesian network diagnostic model of the lubrication system, and, in the example validation, the model was able to accurately diagnose the cause of the failure. Han et al. [5] introduced a convolutional neural network to solve the problem of propulsion fault detection and classification of time series. Xu et al. [6] fused an artificial neural network (ANN) model, a belief rule-based reasoning (BRB) model, and an ER rule model and used a genetic algorithm to optimize the importance weights of each model to improve the overall performance of the fusion system, and finally, through the three models, joint decision-making to realize the fault diagnosis of the ship system. In recent years, algorithmic improvements have achieved significant results in the field of industrial fault diagnosis [7–11]. Samet [12] proposed a new kNN-based classifier (PFS-kNN) to find the k-nearest neighbors using Minkowski's metric of the image fuzzy soft matrix, which was validated on the UCI medical dataset, and the PFS-kNN outperformed the most state-of-the-art kNN-based algorithms. Yang et al. [13] proposed a diagnostic method based on fault sign simulation to solve the problem of a too-small sample size of faults in the fault diagnosis of a particular engine. Agrawal et al. [14] proposed a data-driven model for predictive health maintenance to monitor the health status of the equipment and take action before a fault occurs. The model will alert for any major defects in the system and can be effective in avoiding work interruptions in the production process.

A graph neural network (GNN) is a neural model that captures graph dependencies through message passing between graph nodes. Zhou et al. [15] divided GNN applications into six directions: natural language processing, computational technology, natural science research, knowledge graph, combinatorial optimization, and graph generation, and made a more comprehensive and detailed summary of GNN applications.

In recent years, variants of GNNs such as Graph Convolutional Networks (GCNs) [16] and GANN [17] have achieved breakthroughs in many deep learning tasks. Kipf et al. [18] applied GCNs to fast and scalable semi-supervised classification of nodes in graphs, which demonstrated significant superiority over a large number of datasets. Wang et al. [19] proposed a multi-graph-based GCN fault diagnosis method that fuses two feature embeddings into one combined embedding by introducing a self-attention mechanism to improve classification accuracy and stability in unbalanced datasets. Li et al. [20] proposed a domain adversarial graph convolutional network (DAGCN), which incorporates three types of information—class label, domain label, and data structure—into a deep network for modeling, thus realizing unsupervised domain adaptive (UDA) fault diagnosis. Zheng et al. [21] proposed a new quantum graph convolutional neural network (QGCN) model by drawing on quantum neural networks and graph convolutional neural networks, which utilize quantum parametric circuits and the high computational power of quantum systems to accomplish the traditional task of graph classification in machine learning. In all of the above studies, GCN has achieved good results in the field of graph structure classification due to its advantage of being able to handle non-Euclidean distance data well. However, GCN's

mean-aggregation operation in the metric matrix makes it insufficiently attentive to important nodes. Therefore, Veličković et al. [22] constructed GANN by adding a self-attention layer to traditional GCNs, which are able to set different weights for different nodes in the neighborhood, thus solving the drawback that GCNs cannot consider the importance of neighbors. On this basis, Yang et al. [23] proposed a full graph attention neural network (FGANN), which also considers the influence of nodes other than neighboring nodes and can handle the graph classification task well. Zhang et al. [24] noticed that in sentiment classification tasks, traditional models are not sensitive to syntactic structural information due to the complexity of syntactic analysis relationships, and the models lack the utilization of external sentiment knowledge. In this regard, a new graph attention neural network is proposed that realizes attention to multiple levels of syntax, semantics, and knowledge by introducing the mechanism of multi-head attention and acquires knowledge about sentiment by considering information such as syntax.

The relationship between numerous thermal parameters of ship engines is not just a simple linear relationship but is often very complex and agnostic. In this paper, to address this problem, graph learning theory is introduced to explore the relationship of geometric structure among data, and a MPGANN ship engine fault diagnosis model is proposed. Based on the ship engine failure dataset, the model uses both T_SNE and Spearman to construct the probabilistic graph structure and ordinal graph structure, selects appropriate weights for early fusion of the two graph structures, and finally adopts GANN to transfer and extract the combined features. The core of the model is that it can mine the similarity between the data from two scales and more comprehensively describe the relationship between the data and the graph structure to provide more useful information for the model. The main contributions of this paper can be summarized as follows:

- (1) Transform the ship engine dataset into two graph structures from different scales and make the two graph structures contain similarity relationships from multiple perspectives by extracting the neighbor relationships between samples so as to achieve complementation and extension of the model input information.
- (2) Introducing fusion weights, the two graph structures are structurally fused according to appropriate weights to obtain a fused graph structure that contains deeper information.
- (3) Input the obtained fusion graph structure into the GANN of fused multi-head attention for multi-channel feature extraction, and finally connect the Softmax layer to realize fault diagnosis.
- (4) The ship engine thermal parameter dataset is used to verify that the MPGANN proposed in this paper outperforms other classical algorithms and achieves higher accuracy.

The remaining sections of this paper are organized as follows: In Section 2, the theory related to the MPGANN model is introduced in detail. Section 3 presents experiments and comparisons of the construction of the model proposed in this paper, the setting of hyperparameters, and the performance of the algorithm through case studies. Section 4 then summarizes and outlooks the paper.

2. Related Work and Theory

2.1. Data Topology Construction

Due to the complexity of the work of the ship engine, there is a more complex connection between the various data. In order to fully explore the deep connection between the data, this paper will use the conditional probability and the correlation coefficient between the features of the different nodes to calculate the adjacency relationship between two different scales, as a way to construct the probability graph structure and the similarity graph structure, and the fusion of the two graph structures to get the new fusion graph structure as the input of the network for fault diagnosis.

2.1.1. Probabilistic Graph Structure

Inspired by the T-SNE method [25], in this paper, we use the conditional probability calculated by transforming the distance between the data to represent the correlation between the data, and the correlation $Cor(x_i, x_j)$ serves as the degree of correlation between the data samples x_i and x_j , and the larger $Cor(x_i, x_j)$ is, the higher the likelihood that x_i will be a neighbor of x_j .

For a dataset $X = \{x_1, x_2, x_3 \dots x_n\}$, the conditional probability (denoted as $p_{j|i}$) between any pair of sample points, x_i and x_j , is expressed by Equation (1):

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} (-\|x_i - x_k\|^2 / 2\sigma_i^2)}, \quad (1)$$

In this context, when both $p_{j|i}$ and $q_{j|i}$ are set to 0, σ_i represents the Gaussian variance centered on point x_i . It is worth noting that because data density can vary, there is no universally optimal value for σ_i across all datasets. The entropy of this distribution increases as σ_i grows larger, so a binary search approach is employed to determine the appropriate value of σ_i . Perplexity (denoted as $Per(P_i)$) is defined as the metric for computing σ_i and typically falls within the range of [5, 50]. Its definition is as follows:

$$Per(P_i) = 2^{H(P_i)}, \quad (2)$$

By specifying a perplexity range as described earlier, the information entropy $H(P_i)$ in the equation above is calculated using the following definition:

$$H(P_i) = -\sum_j p_{j|i} \log_2 p_{j|i}, \quad (3)$$

In the low-dimensional space, the symbol $q_{j|i}$ is employed to denote the conditional probability between samples and to symbolize the data correlation $Cor(x_i, x_j)$, as shown in Equation (4):

$$Cor(x_i, x_j) = q_{j|i} = \frac{\exp(-\|y_i - y_j\|^2)}{\sum_{k \neq i} (-\|y_i - y_k\|^2)}, \quad (4)$$

We utilize the KL divergence to quantify the level of concurrence between $q_{j|i}$ and $p_{j|i}$. When the KL divergence is 0, it signifies that $q_{j|i} = p_{j|i}$, signifying that y_i and y_j in the low-dimensional space precisely mirror the correct degree of correlation between data points x_i and x_j in the high-dimensional space. The KL divergence is defined as Equation (5):

$$C = KL(P_i \| Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}}, \quad (5)$$

The KL divergence is minimized through the process of gradient descent:

$$\frac{\delta C}{\delta y_i} = 2 \sum_j (p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j}) (y_i - y_j), \quad (6)$$

After the above formula, the degree of correlation between the two samples $Cor(x_i, x_j)$ is calculated, and the larger $Cor(x_i, x_j)$ is, the stronger the correlation between x_i and x_j . The probabilistic correlation $Cor(x_i, x_j)$ value between each node constitutes the probabilistic feature matrix A_{pp} :

$$A_p = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{bmatrix}, \quad (7)$$

2.1.2. Rank-Order Graph Structure

In this paper, the Spearman correlation coefficient [26] is utilized to derive the similarity graph structure among nodes. The samples in the original data are sorted in ascending order. Let x'_i and x'_j represent the positions of x_i and x_j after this sorting; these positions are referred to as the rank order of samples x_i and x_j . The Spearman correlation coefficient between two samples is calculated as Equation (8):

$$Spe(x_i, x_j) = r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, \quad (8)$$

where d is the difference between x'_i and x'_j and n represents the number of samples. The Spearman correlation coefficient values $Spe(x_i, x_j)$ between the nodes form the correlation characterization matrix A_O :

$$A_O = \begin{bmatrix} o_{11} & o_{12} & \cdots & o_{1N} \\ o_{21} & o_{22} & \cdots & o_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ o_{N1} & o_{N2} & \cdots & o_{NN} \end{bmatrix}, \quad (9)$$

2.1.3. Feature Graph Fusion

Multi-scale feature fusion is mainly used to fuse the probabilistic graph structure and rank-order graph structure constructed earlier and fully excavate the intrinsic feature connection between each graph structure. The more widely used feature fusion methods are early fusion [27], late fusion, and attention feature fusion. Attention feature fusion is able to effectively deal with multimodal data, and it has strong adaptability [28,29]. Late fusion can capture the temporal correlation and interaction effects between data sources well and has better performance in time series tasks [30]. However, the above two methods have the shortcomings of a complex model structure, a high demand for computational resources, and a high cost of training time when the number of samples is small [31]. Therefore, for the discrete small-sample ship engine failure sample set, this paper chooses early fusion [32] as the method of fusion of the multi-graph structure, sets the feature weight W_i and neighbor threshold T of each graph to construct the fusion graph structure, and its formula is as follows:

$$A_F = \sum_{i=1}^K W_i Nol(A_i), \quad (10)$$

$$Nol(x) = \frac{x - \min(x)}{\max(x) - \min(x)}, \quad (11)$$

$$A_T = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix}, \quad (12)$$

$$a_{ij} = \begin{cases} 1, & A_F^{ij} \geq T \\ 0, & A_F^{ij} < T \end{cases} \quad (13)$$

where A_F is the fusion feature matrix; $Nol(A_i)$ represents the feature matrix of each graph after normalization; A_i is the original feature matrix; A_T signifies the adjacency matrix after multi-graph fusion; and a_{ij} denotes the elements of the adjacency matrix in the i th row and j th column of A_F after applying the adjacency threshold T .

The larger the value of the element A_F^{ij} in A_F , the stronger the similarity between node i and node j . How to set the threshold T to define the adjacency matrix A_T is the key. In this

paper, we use quartiles to determine the threshold. The calculated A_F^{ij} values are arranged from small to large and divided into four equal parts. The three values in the numerical split point position are defined as Q1, Q2, and Q3, and Q3 is set to be the critical value with T . Compare A_F^{ij} with T . When A_F^{ij} is greater than T , it means that node i and node j are connected, and $a_{ij} = 1$. When A_F^{ij} is smaller than T , it means that i -node and j -node are not connected, then $a_{ij} = 0$.

2.2. Graphical Attention Neural Network

Both GANN and GCN studied in this paper are variants of GNN. Different from the Laplacian smoothing of all directly adjacent nodes at each node of GCN [33], GANN uses the attention mechanism to calculate the edge weight coefficient. The attention mechanism should be to the operation of aggregating neighbors of GCN. According to the three elements of the attention mechanism: query, source, and attention value [34], we will determine to set query, which is the feature vector of the current central node, source to the feature vector of all neighbors, and attention value to the new feature vector of the center node after the aggregation operation.

Consider the feature vector corresponding to node x_i in layer 1 of the graph as $h_i (h_i \in R^{d^{(l)}})$, where $d^{(l)}$ denotes the feature length of the node. After applying an aggregation operation with an added attention mechanism, the output is a new feature vector $h'_i (h'_i \in R^{d^{(l+1)}})$ for each node, with $d^{(l+1)}$ denoting the length of the output feature vector. We refer to this aggregation operation as the Graph Attention Layer (GAL).

Define the central node as x_i , then the correlation from neighboring nodes x_j to x_i is as follows:

$$e_{ij} = a(Wh_i, Wh_j), \quad (14)$$

The weight parameter of the feature transformation of the nodes in the layer is denoted as $W \in R^{d^{(l+1)} \times d^{(l)}}$. The function $a(\cdot)$ calculates the correlation between two nodes, and in this paper, a single fully connected layer is chosen:

$$e_{ij} = \text{LeakyReLU} \left(\vec{a}^T \left[W \vec{h}_i \parallel W \vec{h}_j \right] \right), \quad (15)$$

The weight vector $a \in R^{2d^{(l+1)}}$ is used along with the activation function LeakyReLU with a negative slope $\alpha = 0.2$. Here, " $\parallel$ " denotes a vector splicing operation. To better assign the weights, the computed correlation is normalized using Softmax:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})}, \quad (16)$$

α is the weight coefficient, which is calculated in the equation above to ensure that the aggregated weight coefficients of all neighboring nodes of the central node sum to one. Figure 1 illustrates the computation process. Wh_i and Wh_j are first parameterized by a , then activated by LeakyReLU, and finally normalized by Softmax to obtain the weight parameters. Here's the complete formula for calculating the weight coefficients:

$$\alpha_{ij} = \frac{\exp \left(\text{LeakyReLU} \left(\vec{a}^T \left[W \vec{h}_i \parallel W \vec{h}_j \right] \right) \right)}{\sum_{k \in N_i} \exp \left(\text{LeakyReLU} \left(\vec{a}^T \left[W \vec{h}_i \parallel W \vec{h}_k \right] \right) \right)}, \quad (17)$$

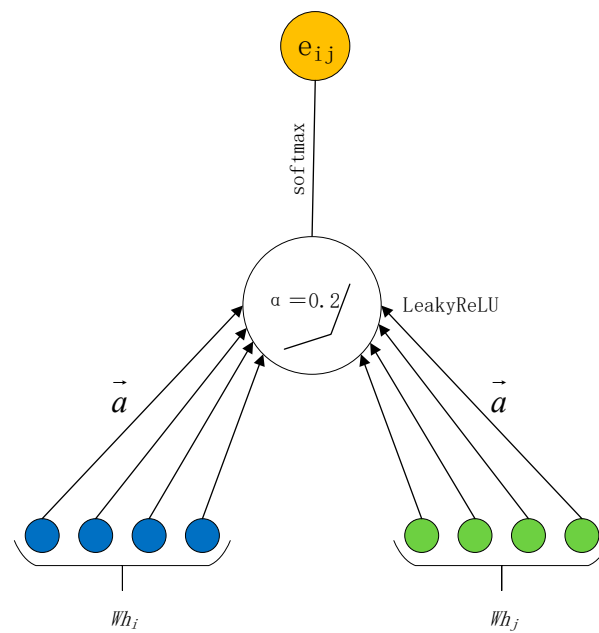


Figure 1. Weighting factors of the calculation process.

The update formula for the feature vector of the node x_i is calculated as follows based on the weight coefficients obtained from the above formula:

$$h^{(l+1)}_i = \sigma \left(\sum_{k \in N_i} \alpha_{ij} W h^{(l)}_j \right), \quad (18)$$

where $h^{(l+1)}_i$ is the updated feature vector of node x_i in layer $l + 1$; σ represents the activation function (commonly used is LeakyReLU); N_i denotes the neighbors of node x_i ; α_{ij} is the weight coefficient calculated using the previously mentioned weight coefficient formula; and W is the weight matrix parameter. This formula calculates the updated feature vector for each node by aggregating information from its neighbors, with the aggregation weights determined by the attention mechanism.

To enhance the expressive capacity of the attention layer, this paper employs the multi-head attention mechanism, similar to the approach utilized by Vaswani et al. [35], to enhance the model's expressive ability. The typical method for combining the outputs involves concatenating the outputs from k sets of mutually independent attention mechanisms in the equation provided above:

$$h^{(l+1)}_i = \parallel_{k=1}^K \sigma \left(\sum_{j \in N_i} \alpha_{ij}^{(k)} W^{(k)} h^{(l)}_j \right), \quad (19)$$

This equation represents the concatenation operation. Each $\alpha_{ij}^{(k)}$ is the weight coefficient computed by the k th set of attention mechanisms, and $W^{(k)}$ corresponds to the associated learnable parameter. This approach combines the outputs from different attention heads, each with its own set of weights and parameters, allowing the model to capture diverse patterns and relationships in the data.

2.3. Algorithmic Process

Figure 2 shows a schematic diagram of the structure of the MPGANN model proposed in this paper. Figure 3 is a schematic diagram of the algorithm flow of the MPGANN model in this paper. Firstly, the values of the detection indexes of the engine in different states are collected by sensors, and the data are normalized. Furthermore, each sample

is regarded as a node, and by constructing a probabilistic graph structure and an ordinal graph structure, the two structural graphs are fused as inputs to the multi-head GANN model by the multi-graph accumulation method, so that the model can learn the sample features at multiple scales. The features of neighboring nodes are aggregated in the model and computed iteratively to maximize the classification accuracy between the extracted feature values and the labeled values.

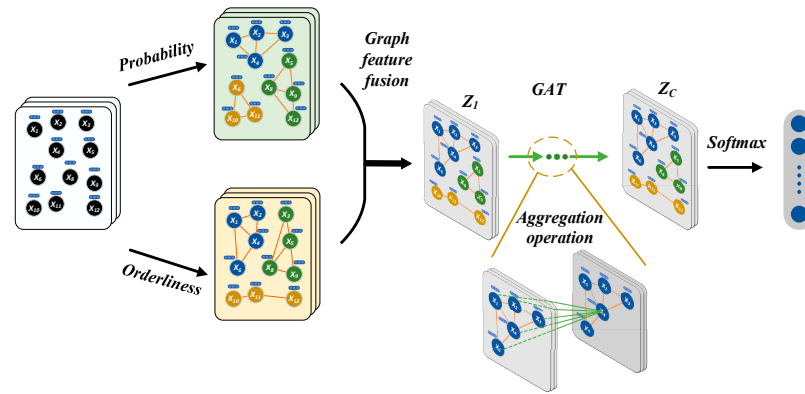


Figure 2. Structure of the MPGANN model.

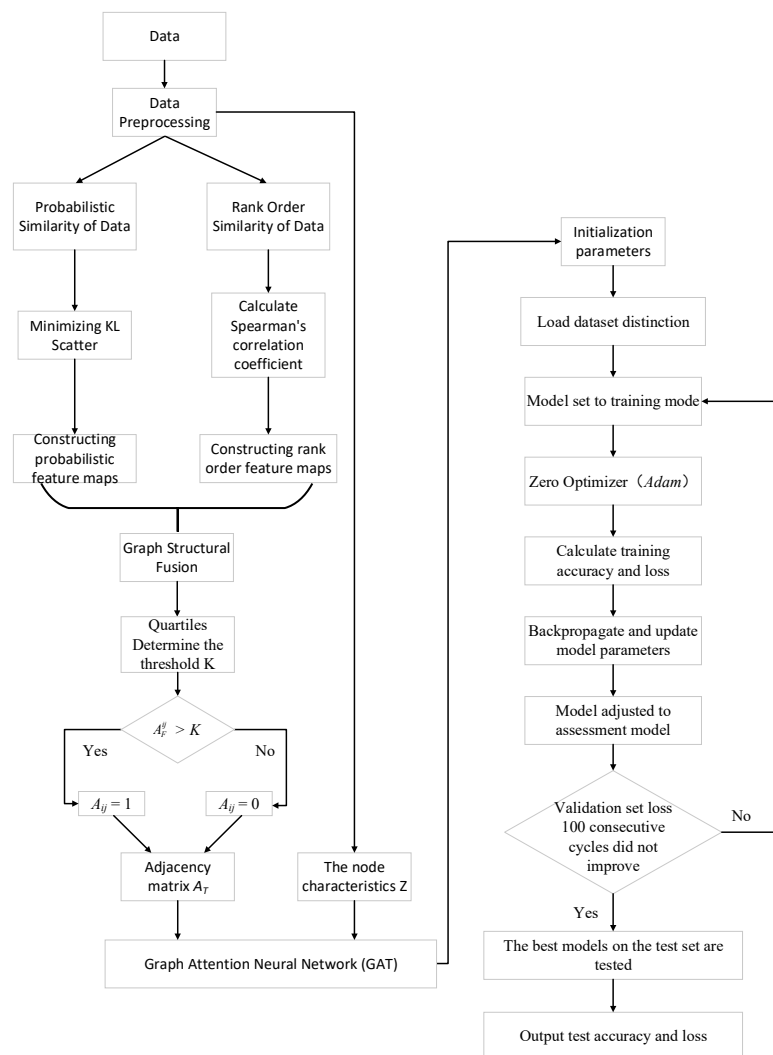


Figure 3. Algorithm flow chart.

3. Case Study

3.1. Data Description

The subject of this experiment is a Wärtsilä 9L34DF marine engine. The real ship failure simulation is costly, dangerous, and destructive; therefore, the engine failure simulation performed by AVL-BOOST (R2019.2) software is used in this paper. The simulation model of the marine engine is shown in Figure 4, where SB is the system boundary; CL is the air filter; TC is the supercharger; CO is the air cooler; PL is the intake and exhaust manifold; C is the cylinder; MP is the arrangement of the detection points; and the black line is the corresponding connecting pipe.

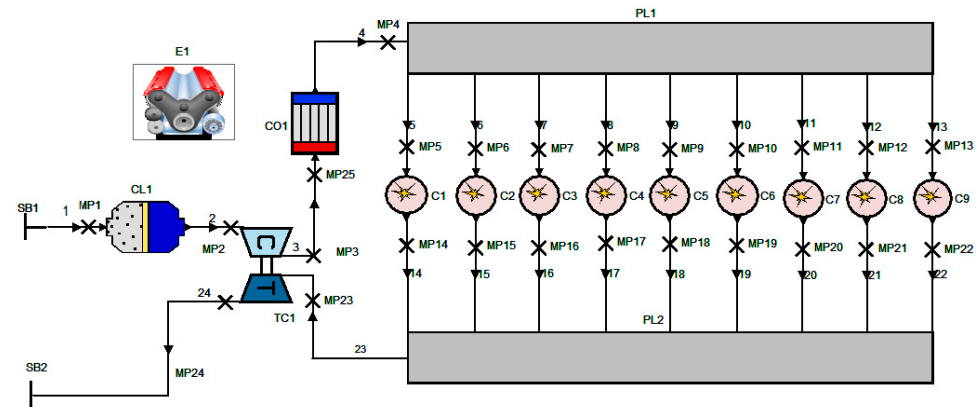


Figure 4. Simulation model of the ship's engine.

Before the fault, simulation needs to verify the effectiveness of the model. The text for the parameters under different working loads is verified for the comparison of measured values and simulation values. Error analysis and comparison results are shown in Table 1.

As can be seen from Table 1, the error between the measured and real values of each main performance parameter under different loads is less than 4%. Especially under 100% load, the error of each main performance parameter is within 2%, and the error is within the acceptable range of requirements, so this paper chooses the model under 100% working conditions for fault simulation and fault sample collection and simulates five states, namely, injection timing advance (F1), injection timing delay (F2), supercharger efficiency decline (F3), air cooler efficiency decline (F4), and normal operation (F5). Five states are simulated, and the fault simulation scheme is shown in Table 2.

Table 1. Comparison results between model simulation data and measured data.

Working Condition	Form	Power/kW	Booster Outlet Pressure/Bar	Air Cooler Outlet Temperature/°C	Fuel Consumption/g/Cycle
100% load	simulation value	4104.99	4.6732	46.6	4.14
	measured value	4078	4.528	46	4.12
	inaccuracies/%	0.6618	1.1815	1.3043	0.4854
75% load	simulation value	3393.19	3.2067	42.9	3.17
	measured value	3375	3.74	42	3.18
	inaccuracies/%	0.5390	−2.1471	2.143	−0.3145
50% load	simulation value	2295.03	2.3018	36.7	2.23
	measured value	2250	2.57	38	2.24
	inaccuracies/%	2.001	−3.4319	−3.4211	−0.4464
25% load	simulation value	1164.88	1.277	28.4	1.21
	measured value	1125	1.42	29	1.26
	inaccuracies/%	3.5449	−3.0282	−2.0689	−3.9683

Table 1. *Cont.*

Working Condition	Form	Inlet Flow/g/s	Average Effective Pressure/bar	Exhaust Temperature/°C
100% load	simulation value	8370.9823	20.0947	455.3
	measured value	8300	19.96	460
	inaccuracies/%	0.8552	0.6748	−0.1022
75% load	simulation value	6477.0966	17.4753	438.75
	measured value	6800	16.52	445
	inaccuracies/%	−3.2780	3.9667	−1.4045
50% load	simulation value	4249.7621	11.4935	437.8
	measured value	4600	11.01	430
	inaccuracies/%	−3.266	3.4518	1.8140
25% load	simulation value	2371.3959	5.7910	427.7
	measured value	2600	5.51	410
	inaccuracies/%	−3.9848	3.5957	3.5593

Table 2. Engine failure simulation scheme.

Malfunction Code	Failure Characteristics	Parameterization				
		Regular Value	Degree of Failure			
			LV1	LV2	LV3	LV4
F1	Injection timing advance	−13.1 deg	−14.1	−15.1	−16.1	−17.1
F2	Delayed injection timing	−13.1 deg	−12.1	−11.1	−10.1	−9.1
F3	Decline in supercharger efficiency	100%	95%	90%	85%	80%
F4	Reduced air cooler efficiency	100%	95%	90%	85%	80%

After the above fault simulation scheme, in order to fully explore and analyze the relationship between internal engine parameters and faults and realize the diagnosis of engine faults, a total of 400 sets of samples of 23-dimensional detection index parameters were selected, of which 80 sets of samples for each state were used to carry out the fault diagnosis study. The dataset is as specific as Table 3, and the collected detection indexes are shown in Table 4.

Table 3. Engine failure dataset.

Working Condition	Dimension	Sample Size
Injection timing advance (F1)	23	80
Delayed injection timing (F2)	23	80
Decline in supercharger efficiency (F3)	23	80
Reduced air cooler efficiency (F4)	23	80
Normal operation (F5)	23	80

Table 4. Detection indicator indexes.

Notation	Monitoring Indicators	Unit	Notation	Monitoring Indicators	Unit
P _{mp3}	Booster outlet pressure	bar	T _{mp23}	Exhaust manifold temperature	°C
V _{mp3}	Supercharger outlet flow	m/s	P _{mp24}	Turbine outlet pressure	bar
T _{mp3}	Supercharger inlet temperature	°C	V _{mp24}	Turbine Outlet Flow	m/s
P _{mp4}	Air cooler inlet pressure	bar	T _{mp24}	Turbine outlet temperature	°C
V _{mp4}	Air cooler outlet flow	m/s	F	Air intake	kg/s
T _{mp4}	Air cooler outlet temperature	°C	g	Fuel consumption rate	g/kW·h
P _{mp5}	Cylinder inlet pressure	bar	P _z	Maximum burst pressure	bar
V _{mp5}	Cylinder inlet flow	m/s	λ	Maximum voltage rise	bar/deg
T _{mp5}	Cylinder inlet temperature	°C	N _e	Power (output)	kW
T _{mp14}	Cylinder exhaust temperature	°C	P _I	IMEP	bar
P _{mp23}	Exhaust manifold pressure	bar	P _B	BMEP	bar
V _{mp23}	Exhaust manifold flow	m/s	/	/	/

3.2. Comparison of Graph Structures

In this section, by comparing and analyzing the commonly used correlation coefficients, including Pearson correlation coefficient, Kendall correlation coefficient, and Spearman correlation coefficient, as well as their constructed correlation graph structure and probability graph structure, respectively, we compare the model accuracy and loss and select the graph structure construction method that is suitable for the proposed model in this paper. The activation function selected for the model is ReLU, the learning rate is 0.01, the number of iterations is 1000, and the number of multi-attention heads is 16. The ratio of the model training set, validation set, and test set is 7:1:2, which is selected randomly. Each graph structure construction method was subjected to 20 experiments, and the average accuracy of the 20 ornaments was used as the basis for judgment.

As can be seen in Figure 5, the average accuracy of the Spearman correlation coefficient in the correlation graph structure is better than that of the Pearson correlation coefficient and Kendall correlation coefficient, and the average accuracy of the fusion graph obtained after the fusion of the graph structures is improved as a model input. This is due to the fact that different graph construction methods extract information from different aspects of the data, and the fused graph structure contains information from both aspects of the data, which can be input into the neural network with more information than a single graph structure. Especially when the probabilistic graph structure and the correlation graph structure constructed by the Spearman correlation coefficient are fused, the average accuracy of the model is greatly improved, reaching 96.36%. It can be seen that the model performance is optimal when the correlation graph structure is constructed with the Spearman correlation coefficient, which is chosen to construct the graphs in all subsequent experiments.

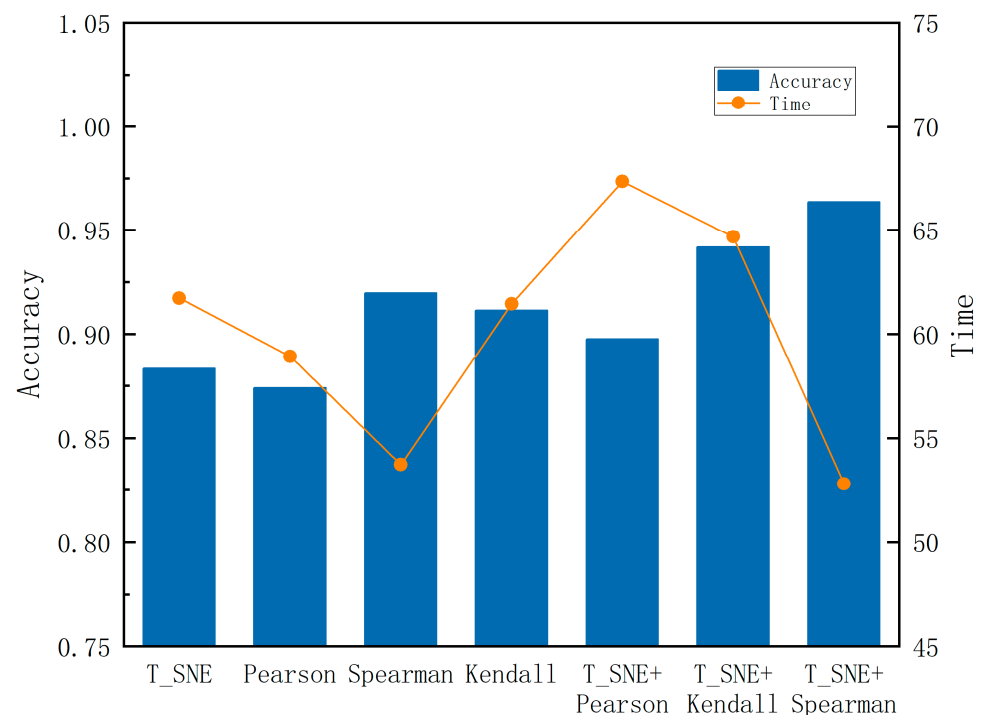


Figure 5. Test accuracy for each graph structure and total elapsed time.

Figure 6 shows the fused graph structure after fusion using the T_SNE probabilistic graph structure and the Spearman rank order graph structure. The edges between the nodes indicate that the nodes are neighbors to each other, as can be seen from the figure that the node relationships are only roughly delineated in a single graph structure, and after fusion of the graph structures, the neighbor relationships between the nodes have been clearly described.

After constructing the neighbor relationship, the labels of some nodes are removed; nodes with color in Figure 7 indicate that the node has label information, while nodes without color indicate that the node has no label information. During the graph learning process, these unlabeled nodes do not provide labeling information, but they can aggregate the neighbor's node information to be used as auxiliary data in order to provide the information used to explore the similarity structure between the nodes. The information of the nodes is used to aggregate the neighboring information through multi-head attention, and after the graph learning, the updated new nodes are classified, and all the samples are tagged with labels. The above figure shows the classification results, and it can be seen that although the model is not given all the labeling information, it can still correctly label the unlabeled nodes.

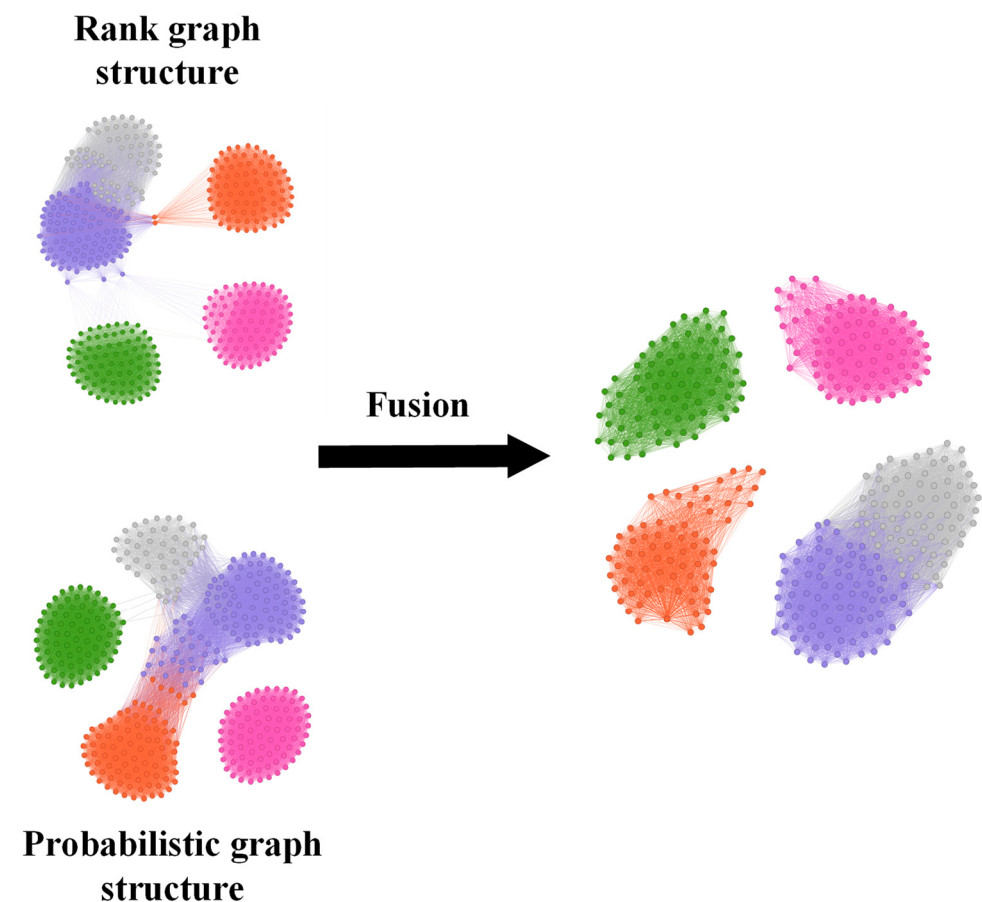


Figure 6. Fusion of probabilistic graph structures with rank-order graph structures.

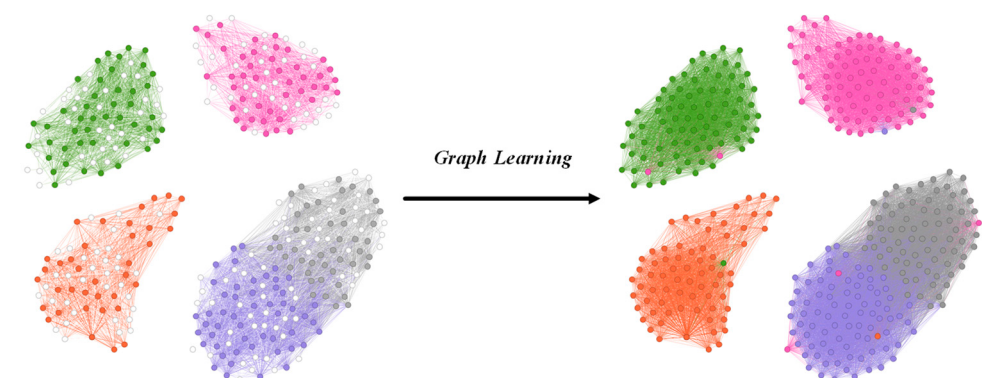


Figure 7. Graph learning process.

3.3. Network Parameter Setting

In this subsection, we will delve into the critical hyperparameters of the neural network model for fault diagnosis. These include the activation function, the number of attention heads, and the model learning rate. We maintain the ratio of the training set, validation set, and test set at 7:1:2, conducting a total of 20 experiments. The results from these experiments will be averaged to determine the average accuracy rate, serving as the basis for the final assessment and guiding the selection of the most suitable network structure.

3.3.1. Activation Function

The activation function is a very important part of the neural network structure, which can nonlinearly transform the input data. In this paper, five typical activation functions, Sigmoid, ReLU, ELU, ReLU6, and Tanh, as well as three novel activation functions, SupEx [36], α SechSig, and α TanhSig [37], are selected for comparison, and the number of iterations is set to be 1000 and the learning rate is set to be 0.01. Figure 8 shows the total time consumed and the average accuracy in the test set for the models with different activation functions selected. of the total elapsed time and the average accuracy of the test set.

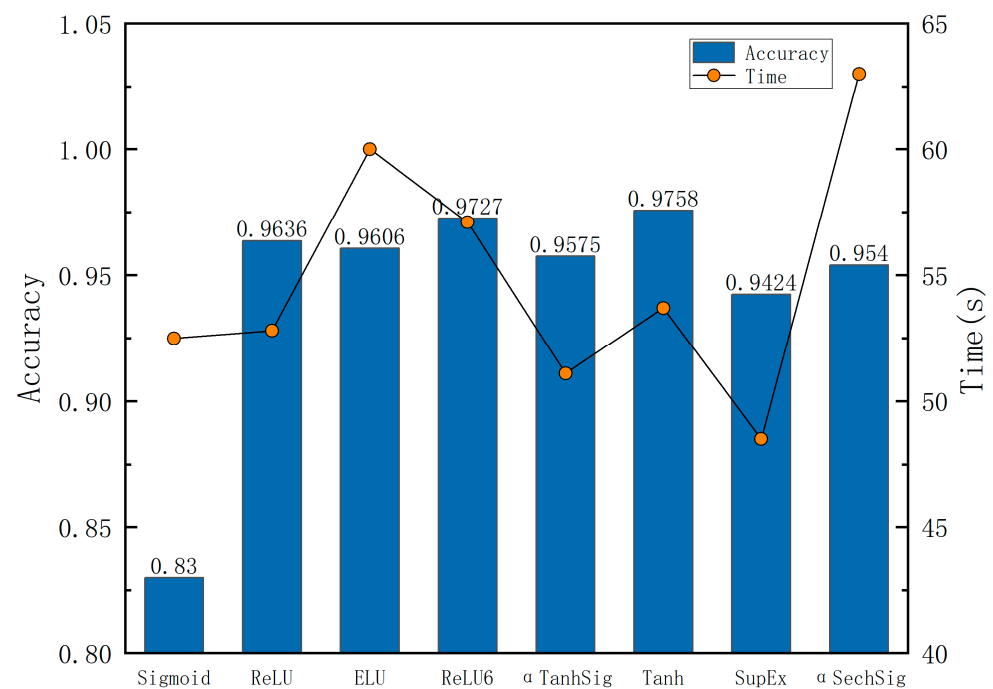


Figure 8. Average accuracy and total elapsed time for different activation functions.

In Figure 8, it can be seen that Sigmoid has the lowest accuracy when it is used as an activation function, and the total time consumed by α SechSig and ELU is significantly higher than the other activation functions, and the total time consumed by the rest of the activation functions, even though the total time consumed by SupEx and α TanhSig is less, the total time consumed is not long in the data of the small samples in the first place, so the accuracy is used as the main criterion and the total time consumed is used as the auxiliary criterion, and the model's classification accuracy for the test set is over 0.97 when Tanh and ReLU6 are used as activation functions, the classification accuracy of the model on the test set exceeds 0.97, reaching 0.9758 and 0.9727, of which Tanh not only has high accuracy but also the total time consumed is less than ReLU6.

In Figure 9, we can intuitively see that Tanh, as the activation function, has the highest average accuracy; the fluctuation range of the experimental results is also the smallest; and most of the experimental results are in the 0.975 or so range. The synthesis of the

above two results can be seen in this paper's model Tanh, as the activation function is the optimal effect.

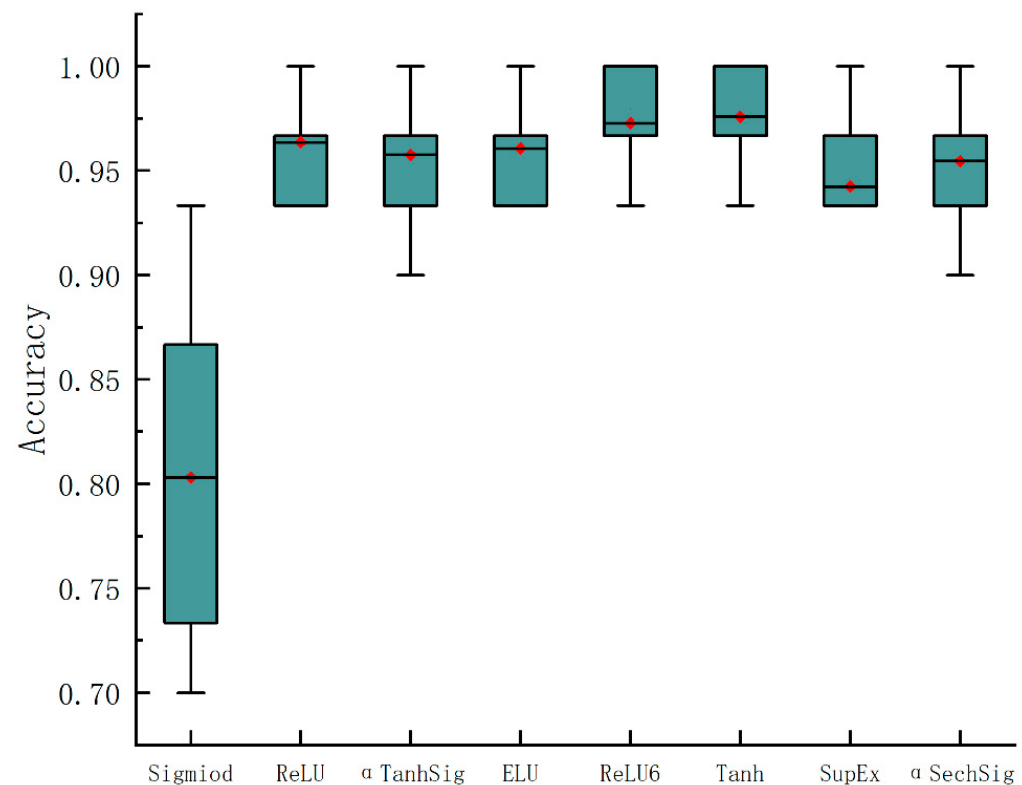


Figure 9. Boxplot of the accuracy of different activation functions.

3.3.2. Number of Attention Heads

Multi-head attention allows the model to focus on different channels of the input at the same time, thus improving the expressiveness and performance of the model. It should be noted that while multi-head attention increases the complexity of the model, it also increases the computational overhead of training and inference, so the optimal number of heads needs to be determined experimentally to achieve the best performance of the model. The activation function selected in the experiment is Tanh, the value range of H is set to $[1, 50]$, and other settings are the same as in the previous experiment. The final comparison of the average accuracy of the test set under different numbers of heads and the experimental results is shown in the following figure.

Based on Figure 10, it is evident that the model achieves higher accuracy within the range of head counts $H = 16$, $H = 22$, and $H = 30$. The highest accuracy is attained at a learning rate of $Lr = 0.01$ with $H = 16$. However, it is worth noting that excessively high head counts lead to longer running times and increased computational costs. Therefore, the optimal configuration of the model is $Lr = 0.01$ and $H = 16$.

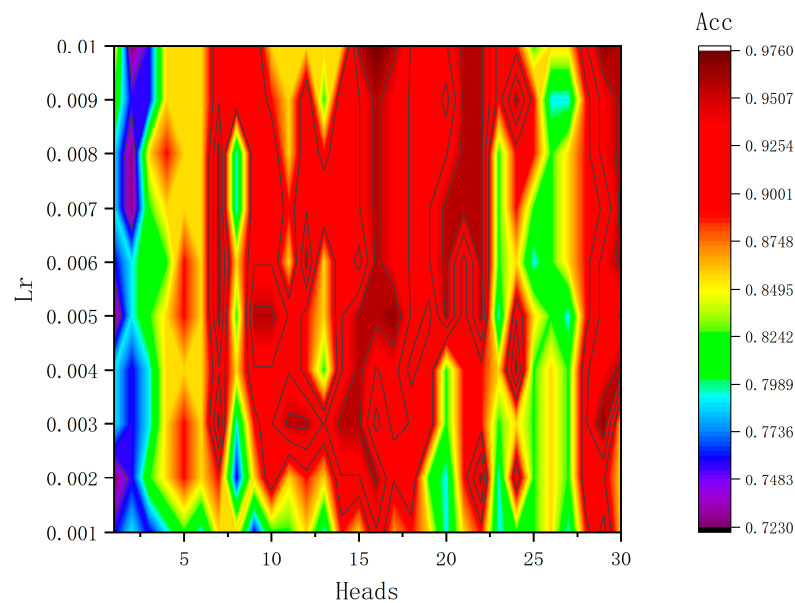


Figure 10. Heat map of model accuracy at different learning rates and head counts.

3.3.3. Algorithm Performance Comparison

In order to verify the validity of the proposed model in this paper, GCN, CNN, SVM, and BPNN algorithms are selected in the experiments in this subsection to compare the algorithm performance with the MPGANN model proposed in this paper. The parameters of GCN, such as the neighbor graph construction method, the number of hidden layers, the number of neurons in the hidden layer, and the dropout rate, are all the same as those of MPGANN. CNN has a total of four convolutional layers. The size of the convolution kernel is 5-3-2-2, the stride is 1-1-1-1, and each convolution layer is connected to an average pooling layer after each convolution layer. The window size is 2-2-2-2, padding is 0, and stride is 1. The kernel function of SVM is the Gaussian kernel function; set the penalty factor $c = 1.0$ and set the gamma to auto. In the BPNN hidden layer, the number of neurons is set to 16, the number of iterations for all algorithms is set to 1000, and the learning rate is set to 0.01.

Accuracy is an overall performance indicator; if it is used as a performance evaluation index for diagnostic algorithms, it will likely mask the specific classification of diagnostic algorithms, so in order to solve this type of problem, this paper introduces precision, recall, and the F1 score, calculated as:

$$R = \frac{TP}{TP + FN}, \quad (20)$$

$$P = \frac{TP}{TP + FP}, \quad (21)$$

$$F = \frac{2}{\frac{1}{P} + \frac{1}{R}} = \frac{2RP}{R + P}, \quad (22)$$

where R is the recall rate; P is the precision rate; F is the F1 score; TP is the number of samples that classify positive samples as positive samples; FN is the number of positive samples classified as negative samples, and FP is the number of negative samples classified as positive samples.

The precision and recall of each algorithm used in this paper are shown in Table 5. From the table, it can be seen that each algorithm can basically correctly classify the normal working conditions, but the complexity of the engine fault data leads to a relatively poor classification effect for the traditional diagnostic algorithms. In this experiment, the precision rate and recall rate of BPNN are lower than other diagnostic algorithms, and

although the precision rate and recall rate of GCN and CNN diagnosis for each condition are better than those of SVM and BPNN, the diagnostic effect is still unsatisfactory in F1 and F2 conditions. The MPGANN proposed in this paper achieves 100% accuracy and recall in all working conditions except for 0.89 accuracy and 0.88 recall in F2 and F1 working conditions, which is an obvious advantage over other diagnostic algorithms and is more in line with practical industrial needs.

Table 5. Precision and recall for each algorithm.

Algorithms	Precision					Recall				
	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5
MPGANN	1	0.89	1	1	1	0.88	1	1	1	1
GCN	0.88	0.884	0.94	1	1	0.88	0.94	1	0.94	0.94
CNN	0.8	0.78	1	1	1	0.75	0.88	1	1	0.94
SVM	0.71	0.76	1	1	1	0.75	0.81	0.94	1	0.94
BPNN	0.65	0.72	0.88	0.88	1	0.69	0.81	0.88	0.88	0.81

From Figure 11, it can be seen that fault F1 is misclassified in all diagnostic algorithms, while after the feature mining of probabilistic graph relationships and rank order graph structure, MPGANN and GCN only have two F1 fault samples misclassified, and the MPGANN proposed in this paper only has two samples misclassified, which shows that the MPGANN model can mine more information between the data to improve the diagnostic algorithm's classification accuracy and stability and is better than other models.

Figure 12 shows the average accuracy, average precision, average recall, and F1 score of each diagnostic algorithm. The performance of each diagnostic algorithm can be seen more clearly by comparing the figures. CNN, as a network model for deep learning, is widely used in the field of mechanical fault diagnosis and has achieved excellent results. However, due to the fact that there is no large amount of data for model training in this experiment, the training effect of CNN is not good, and it cannot make correct classification judgments. GCN is limited by its own information propagation rules, and even though it adopts the same topology structure for fault diagnosis, the diagnostic effect is still not good. However, the MPGANN proposed in this paper uses the multi-head attention mechanism in the aggregation operation, which makes the feature aggregation more reasonable and increases the learning efficiency of the model, and the average precision, average accuracy, average recall, and F1 score reach 0.978, 0.976, 0.975, and 0.976.

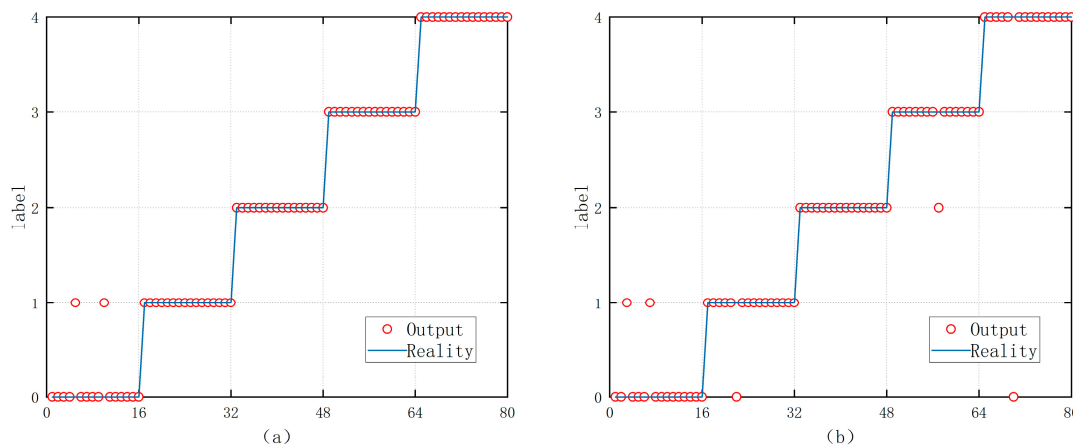


Figure 11. Cont.

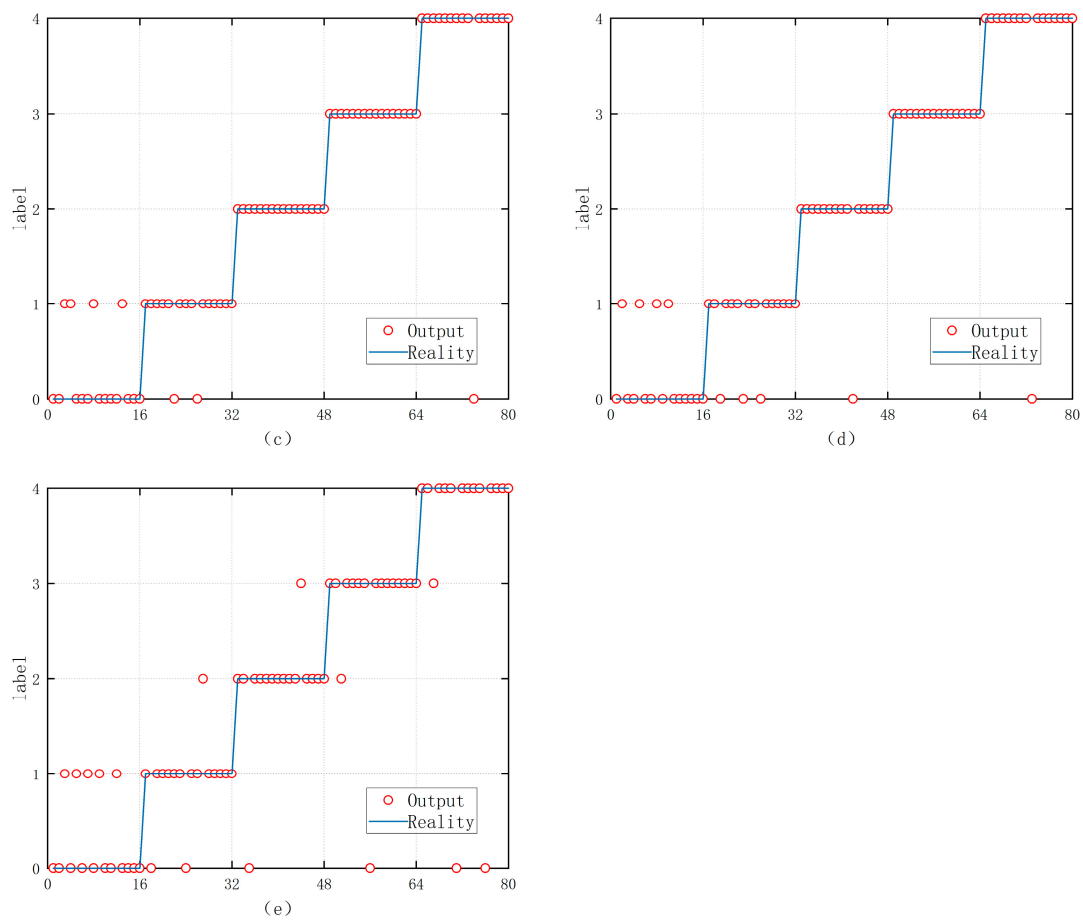


Figure 11. Categorization of test set failures: (a) MPGAT, (b) GCN, (c) CNN, (d) SVM, and (e) BPNN.

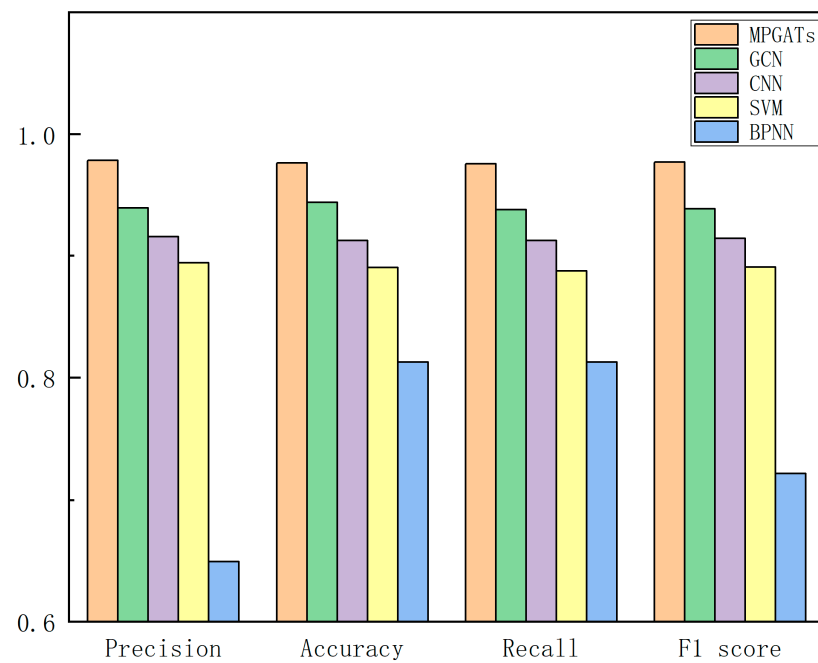


Figure 12. Algorithm performance comparison.

4. Conclusions

In this paper, a new MPGANN model is proposed, and in the experiments on the ship engine dataset, firstly, different graph structure construction methods are compared, and the fused graph structure enhances the expressive ability, especially the fusion of probabilistic and ordinal graph structures, which dramatically improves the diagnostic accuracy of the model. Secondly, the accuracy of the model under different activation functions, head counts, and learning rates is greatly different, which shows that the influence of hyperparameters on the model is also not negligible. Finally, when MPGANN is compared with other algorithms, it outperforms the various methods compared in diagnosis in terms of accuracy, precision, recall, and F1 score. Therefore, the MPGANN proposed in this paper has a certain practical significance in the field of intelligent ship fault diagnosis. In summary, the research for this paper is summarized as follows:

- (1) The two graph structure construction methods in the MPGANN model can effectively obtain probabilistic similarity and ordinal similarity between data samples and transform the data into probabilistic graph structure and ordinal graph structure.
- (2) Early fusion is employed to combine the probability map structure and rank-order map structure with the incorporation of feature weights. This integration process effectively amalgamates information from samples at various scales.
- (3) The multi-head attention mechanism is applied to conduct multi-channel feature screening on the fusion graph structure, extracting feature information with higher relevance to enhance the diagnostic performance of the model.
- (4) The model effect was validated using the ship engine fault dataset, and compared with other models in terms of accuracy, precision, recall, and F1 score, MPGANN was the most effective, with a diagnostic accuracy as high as 97.58% and an F1 score of 97.6%.

The proposed diagnostic model in this paper fuses features from different aspects of the data, mining the information of the dataset at a deeper level, but it also has some limitations, such as: the fused features contain only probabilistic similarity and rank-order similarity; the fusion weights need to be set in advance; and so on.

There are three goals for future work: (1) Physical connections in real space will also imply information, and the actual structural connections of the engine will be transformed into graph structures to mine the correlation information between the samples. (2) Choose more graph fusion methods or assign meanings to edge features so that the model has more information for training to improve its diagnostic accuracy. (3) Introduce the attention mechanism into the fusion process to achieve adaptive allocation of fusion weights.

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

Funding: This research was funded by the project: Development of Ship Operation Condition Monitoring and Simulation Platform (funding number: 1638882993269) and the research and application of the Smart Ship Digital Twin Information Platform (funding number: 2022JH1/10800097).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: Author Hui Cao was employed by Dalian Maritime University Smart Ship Limited Company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Zhong, G.Q.; Wang, H.Y.; Zhang, K.Y.; Jia, B.Z. Fault diagnosis of Marine diesel engine based on deep belief network. In Proceedings of the 2019 Chinese Automation Congress (CAC), Hangzhou, China, 22–24 November 2019; pp. 3415–3419. [\[CrossRef\]](#)
2. Hou, L.; Zhang, J.; Du, B. A Fault Diagnosis Model of Marine Diesel Engine Fuel Oil Supply System Using PCA and Optimized SVM. *J. Phys. Conf. Ser.* **2020**, *1576*, 012045. [\[CrossRef\]](#)
3. Zhong, K.; Li, J.B.; Wang, J.; Han, M. Fault Detection for Marine Diesel Engine Using Semi-supervised Principal Component Analysis. In Proceedings of the 9th International Conference on Information Science and Technology (ICIST), Hulunbuir, China, 2–5 August 2019; pp. 146–151.
4. Ren, D.P.; Zeng, H.; Wang, X.L.; Pang, S.; Wang, J.D. Fault Diagnosis of Diesel Engine Lubrication System Based on Bayesian Network. In Proceedings of the 6th International Conference on Control, Automation and Robotics (ICCAR), Electr Network, Singapore, 20–23 April 2020; pp. 423–429.
5. Han, P.H.; Li, G.Y.; Skulstad, R.; Skjong, S.; Zhang, H.X. A Deep Learning Approach to Detect and Isolate Thruster Failures for Dynamically Positioned Vessels Using Motion Data. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–11. [\[CrossRef\]](#)
6. Xu, X.J.; Zhao, Z.Z.; Xu, X.B.; Yang, J.B.; Chang, L.L.; Yan, X.P.; Wang, G.D. Machine learning-based wear fault diagnosis for marine diesel engine by fusing multiple data-driven models. *Knowl. Based Syst.* **2020**, *190*, 105324. [\[CrossRef\]](#)
7. Cao, H.; Zhang, J.D.; Cao, X.; Li, R.; Wang, Y.R. Optimized SVM-Driven Multi-Class Approach by Improved ABC to Estimating Ship Systems State. *IEEE Access* **2020**, *8*, 206719–206733. [\[CrossRef\]](#)
8. Wang, R.H.; Chen, H.; Guan, C. Random convolutional neural network structure: An intelligent health monitoring scheme for diesel engines. *Measurement* **2021**, *171*, 108786. [\[CrossRef\]](#)
9. Memis, S.; Arslan, B.; Aydin, T.; Enginoglu, S.; Camci, C. Distance and Similarity Measures of Intuitionistic Fuzzy Parameterized Intuitionistic Fuzzy Soft Matrices and Their Applications to Data Classification in Supervised Learning. *Axioms* **2023**, *12*, 463. [\[CrossRef\]](#)
10. Choo, K.B.; Cho, H.; Park, J.H.; Huang, J.F.; Jung, D.; Lee, J.; Jeong, S.K.; Yoon, J.; Choo, J.; Choi, H.S. A Research on Fault Diagnosis of a USV Thruster Based on PCA and Entropy. *Appl. Sci.* **2023**, *13*, 3344. [\[CrossRef\]](#)
11. Zhao, N.Y.; Mao, Z.W.; Wei, D.H.; Zhao, H.P.; Zhang, J.J.; Jiang, Z.N. Fault Diagnosis of Diesel Engine Valve Clearance Based on Variational Mode Decomposition and Random Forest. *Appl. Sci.* **2020**, *10*, 1124. [\[CrossRef\]](#)
12. Memis, S. Picture Fuzzy Soft Matrices and Application of Their Distance Measures to Supervised Learning Picture Fuzzy Soft k-Nearest Neighbor (PFS-kNN). *Electronics* **2023**, *12*, 4129. [\[CrossRef\]](#)
13. Yang, K.; Fan, H.Y. Research on Fault Diagnosis Method of Diesel Engine Thermal Power Conversion Process. In Proceedings of the 3rd International Forum on Energy, Environment Science and Materials (IFEESM), Shenzhen, China, 25–26 November 2017; pp. 1445–1450.
14. Agrawal, S.K.; Banerjee, S.; Sinha, A.; Das, D. SafeEngine: Fault Detection with Severity Prediction for Diesel Engine. In Proceedings of the 2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC), Hyderabad, India, 16–18 September 2022; pp. 216–220.
15. Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; Sun, M. Graph neural networks: A review of methods and applications. *AI Open* **2020**, *1*, 57–81. [\[CrossRef\]](#)
16. Nie, J.; Xu, Y.; Huang, Y.; Li, J. The review of image processing based on graph neural network. In Proceedings of the International Conference on Intelligent Robotics and Applications, Yantai, China, 22–25 October 2021; pp. 534–544.
17. Zhou, A.Z.; Li, Y.F. Structural attention network for graph. *Appl. Intell.* **2021**, *51*, 6255–6264. [\[CrossRef\]](#)
18. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, arXiv:1609.02907.
19. Wang, R.; Chen, H.; Guan, C. DPGCN Model: A Novel Fault Diagnosis Method for Marine Diesel Engines Based on Imbalanced Datasets. *IEEE Trans. Instrum. Meas.* **2022**, *72*, 1–11. [\[CrossRef\]](#)
20. Li, T.; Zhao, Z.; Sun, C.; Yan, R.; Chen, X. Domain adversarial graph convolutional network for fault diagnosis under variable working conditions. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–10. [\[CrossRef\]](#)
21. Zheng, J.; Gao, Q.; Lü, Y. Quantum graph convolutional neural networks. In Proceedings of the 2021 40th Chinese Control Conference (CCC), Shanghai, China, 26–28 July 2021; pp. 6335–6340.
22. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv* **2017**, arXiv:1710.10903.
23. Yang, F.; Zhang, H.Y.; Tao, S.M. Semi-supervised classification via full-graph attention neural networks. *Neurocomputing* **2022**, *476*, 63–74. [\[CrossRef\]](#)
24. Zhang, S.; Gong, H.; She, L. An aspect sentiment classification model for graph attention networks incorporating syntactic, semantic, and knowledge. *Knowl. Based Syst.* **2023**, *275*, 110662. [\[CrossRef\]](#)
25. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
26. Zhang, W.-Y.; Wei, Z.-W.; Wang, B.-H.; Han, X.-P. Measuring mixing patterns in complex networks by Spearman rank correlation coefficient. *Phys. A Stat. Mech. Appl.* **2016**, *451*, 440–450. [\[CrossRef\]](#)
27. Wang, Y.; Qin, Y.; Guo, J.; Cao, Z.; Jia, L. Multi-point short-term prediction of station passenger flow based on temporal multi-graph convolutional network. *Phys. A Stat. Mech. Appl.* **2022**, *604*, 127959. [\[CrossRef\]](#)
28. Wang, L.; Hu, Z.; Kong, Q.; Qi, Q.; Liao, Q. Infrared and Visible Image Fusion via Attention-Based Adaptive Feature Fusion. *Entropy* **2023**, *25*, 407. [\[CrossRef\]](#) [\[PubMed\]](#)

29. Luo, Z.; Li, J.; Zhu, Y. A deep feature fusion network based on multiple attention mechanisms for joint iris-periocular biometric recognition. *IEEE Signal Process. Lett.* **2021**, *28*, 1060–1064. [[CrossRef](#)]
30. Yang, J.; Yang, J.-Y.; Zhang, D.; Lu, J.-F. Feature fusion: Parallel strategy vs. serial strategy. *Pattern Recognit.* **2003**, *36*, 1369–1381. [[CrossRef](#)]
31. Wang, Y.; Xu, X.; Yu, W.; Xu, R.; Cao, Z.; Shen, H.T. Comprehensive Framework of Early and Late Fusion for Image–Sentence Retrieval. *IEEE MultiMedia* **2022**, *29*, 38–47. [[CrossRef](#)]
32. Boulahia, S.Y.; Amamra, A.; Madi, M.R.; Daikh, S. Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition. *Mach. Vis. Appl.* **2021**, *32*, 121. [[CrossRef](#)]
33. Ruiz, L.; Gama, F.; Ribeiro, A. Graph neural networks: Architectures, stability, and transferability. *Proc. IEEE* **2021**, *109*, 660–682. [[CrossRef](#)]
34. Brauwiers, G.; Frasincar, F. A general survey on attention mechanisms in deep learning. *IEEE Trans. Knowl. Data Eng.* **2021**, *35*, 3279–3298. [[CrossRef](#)]
35. Weston, J.; Chopra, S.; Bordes, A. Memory networks. *arXiv* **2014**, arXiv:1410.3916.
36. Kiliçarslan, S.; Közkurt, C.; Bas, S.; Elen, A. Detection and classification of pneumonia using novel Superior Exponential (SupEx) activation function in convolutional neural networks. *Expert Syst. Appl.* **2023**, *217*, 119503. [[CrossRef](#)]
37. Közkurt, C.; Kiliçarslan, S.; Bas, S.; Elen, A. α SechSig and α TanhSig: Two novel non-monotonic activation functions. *Soft Comput.* **2023**, *27*, 18451–18467. [[CrossRef](#)]

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