

CMAGIN: A Channel-aware Multi-scale Adaptive Graph Interaction Network for Multivariate Time Series Forecasting

Weipeng Liu^{1*}, Bowen Shi², Jixuan Wang^{3*}, Yuxiang Hu⁴, Xiaofei Xue⁵, Haixing Zhu⁶, and Xiangrui Gong⁷

^{1,2}*School of Electrical Engineering, Hebei University of Technology, Tianjin 300401, China*

liuweipeng@hebut.edu.cn

shibowenr@tju.edu.cn

^{1,3,4,5,6,7}*School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China*

wangjixuan@hebut.edu.cn

HuYX0126@163.com

xuexiaofei@hebut.edu.cn

zhx930419@163.com

18822633034@163.com

^{1,2,3,4,5,6,7}*State Key Laboratory of Intelligent Power Distribution Equipment and System, Hebei University of Technology, Tianjin 300401, China*

ABSTRACT

Multivariate time series (MTS) data is widely utilized in industrial manufacturing, equipment maintenance, and health monitoring. However, the high dimensionality, dynamic nature, and heterogeneity characteristics bring significant challenges for modeling. Traditional deep learning algorithms based on sequential modeling struggle to capture the complex structural relationships between different time series variables, making it difficult to uncover interaction patterns and potential dependencies. To address the dynamic and complex dependencies among variables in MTS data and further balance the importance distribution across multiple temporal feature channels, this work proposes a channel-aware multi-scale adaptive graph interaction network (CMAGIN) for MTS forecasting. The proposed framework integrates a dynamic and adaptive graph constructor with local awareness and global attention (DAGC-LAGA) and a channel-wise adaptive center enhancement (CACE) mechanism. The design of DAGC-LAGA captures sparse neighborhood relations through a multi-view local dynamic graph constructor and further leverages a global attention graph enhancer to model semantic correlations. The results effectively display dynamic dependencies among variables. The introduction of the CACE module dynamically enhances key node features by calculating the node importance at the channel level. In addition, applying the centrality-aware

attention mechanism improves the sensitivity of the model to crucial temporal patterns. Furthermore, the results are verified via the C-MAPSS dataset for aircraft engine degradation prediction. Experimental results demonstrate that the CMAGIN model outperforms comparative methods in both RMSE and Score metrics, and exhibits robust performance under complex operating conditions and multiple-fault scenarios. Future research could investigate scalable applications of CMAGIN across diverse industrial scenarios to enable field deployment of intelligent operation and maintenance systems.

1. INTRODUCTION

In industrial scenarios, the widespread deployment of sensors in control systems leads to the continuous generation of large volumes of multivariate time series (MTS) data. These data can predict system states through historical behavior analysis and hold significant application value in fields such as transportation and energy (Feng, Shao, Wang, Zhang, & Wen, 2025; Huo et al., 2023). The MTS data consists of sequential measurements collected from multiple sensors over time and is widely used to represent the evolving states of complex systems or processes. The accurate modeling and forecasting of MTS data become essential for timely decision making and fault prevention in industrial applications (Casolaro, Capone, Iannuzzo, & Camastra, 2023; H. Wang, Zhang, Yang, & Xiang, 2023). However, the heterogeneous sources and dynamic characteristics result in the industrial MTS data being dependent on strong spatiality and temporal dependency. These properties make it difficult for traditional

Weipeng Liu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
<https://doi.org/10.36001/IJPHM.2025.v16i2.4571>

Abbreviation	Full Name	Description
CMAGIN	Channel-aware multi-scale adaptive graph interaction network	Proposed framework
DAGC-LAGA	Dynamic and adaptive graph constructor with local awareness and global attention	Dynamic graph module
CACE	Channel-wise adaptive center enhancement	Channel-level modulation module
MV-LDGC	Multi-view local dynamic graph constructor	Local sparse subgraph builder
GAGE	Global attention graph enhancer	Global dependency modeler
RUL	Remaining Useful Life	Target variable
PHM	Prognostics and Health Management	Application field

Table 1. List of abbreviations

methods to capture the underlying patterns and interactions. To further capture spatiotemporal dependencies, model nonlinear interactions, and further handle uncertainties in industrial environments, it is necessary to develop novel forecasting methods.

Deep learning methods have achieved significant success in MTS forecasting. Some of these approaches, such as the long- and short-term time-series network (Lai, Chang, Yang, & Liu, 2018) and the temporal pattern attention for MTS forecasting (Shih, Sun, & Lee, 2019), effectively capture nonlinear temporal patterns. However, these approaches still face challenges in learning spatial dependencies among variables and accurately modeling pairwise interactions. Meanwhile, graph neural networks (GNNs) have demonstrated strong potential in capturing spatial relationships by utilizing graph-structured data, where entities are abstracted as nodes and their interactions as edges (Wu et al., 2021). Compared to traditional methods, GNNs demonstrate superior performance in MTS forecasting by effectively capturing spatial correlations through their graph structure, dynamically modeling variable interactions via message passing mechanisms, and flexibly adapting to changing sensor network configurations (Jin et al., 2024). These indicate that GNNs could be a powerful solution for addressing the critical challenge of spatiotemporal dependency modeling in complex industrial systems

Although GNN-based approaches have significantly advanced MTS forecasting, several critical challenges persist in modeling complex and dynamic spatiotemporal dependencies. Adaptive graph convolutional recurrent network (Bai, Yao, Li, Wang, & Wang, 2020) employs node-adaptive embeddings to implicitly learn spatial relationships, yet it depends on a shared and temporally invariant graph construction. While recent efforts have addressed temporal adaptivity, they still face limitations in fusing heterogeneous graph signals or capturing both local and global dynamic interactions comprehensively (N. Xu, Kosma, & Vazirgiannis, 2023). These models rely on static or single-scale graph structures, limiting the ability to capture time-varying high-order dependencies. STGSL-Balanced (W. Chen et al., 2023) applies shared processing across variables,

limiting the model's capacity to distinguish informative channels from redundant ones. The parameter-sharing paradigm weakens the ability to adapt to dynamic variations and hampers the extraction of critical features, ultimately degrading predictive performance (Liu et al., 2022). These methods adopt uniform transformation and aggregation operations across channels, overlooking inherent heterogeneity and temporal dynamics present in multivariate signals

To address the aforementioned issues, this paper proposes a channel-aware multi-scale adaptive graph interaction network (CMAGIN) for modeling dynamic dependencies and channel-specific heterogeneity. CMAGIN comprises two key modules: a dynamic and adaptive graph constructor with local awareness and global attention (DAGC-LAGA) and a channel-wise adaptive center enhancement (CACE) module. In addition, DAGC-LAGA builds graph structures from two views: the multi-view local dynamic graph constructor (MV-LDGC) captures evolving sparse dependencies by tracking adjacency changes over time, while the global attention graph enhancer (GAGE) constructs a global graph using cross-channel semantic correlations to model long-range dependencies. CACE enhances central node representations by computing channel-level importance scores based on temporal strength and consistency, generating attention vectors to guide information aggregation from critical channels. This enables precise modeling of key variables across time and channels.

In summary, our main contributions are as follows:

- 1) To address the complex and dynamic dependencies among variables in MTS data and the uneven distribution of importance across temporal feature channels, this work proposes a CMAGIN for MTS forecasting, which incorporates the DAGC-LAGA and CACE.
- 2) This paper presents the DAGC-LAGA, which integrates the MV-LDGC and GAGE submodules to dynamically capture sparse local dependencies and global semantic correlations in MTS data. Simultaneously, the CACE module is introduced to establish a channel-wise node importance

measurement mechanism, enabling dynamic enhancement of the expressive power of critical node features.

3) To validate the effectiveness of the proposed model, remaining useful life (RUL) prediction experiments are conducted on four subsets of the C-MAPSS dataset. The results demonstrate that the proposed approach outperforms state-of-the-art methods in terms of RMSE and Score, showing robust performance under diverse operating conditions and compound fault scenarios.

2. RELATED WORK

2.1. From Statistical Models to Deep Sequential Learning

Traditional time series forecasting depends heavily on statistical methods such as ARIMA, which assumes linearity and stationarity (Tarmanini, Sarma, Gezevin, & Ozgonenel, 2023). While effective for simple patterns, these methods struggle to model complex nonlinearities and long-range dependencies (Hyndman & Athanasopoulos, 2018). The emergence of deep learning introduces recurrent architectures such as long short-term memory (LSTM) (Guo, He, & Wang, 2024) and gated recurrent unit (GRU) (Bacanin et al., 2023), which capture temporal dependencies more effectively through gating mechanisms. In addition, 1D convolutional neural networks (Kim, Oh, Kim, & Choi, 2023) leverage parameter sharing and local receptive fields, showing competitive performance in short-term time series modeling.

Subsequently, the Transformer architecture (Vaswani et al., 2017) revolutionizes sequence modeling with self-attention, enabling parallel computation and superior long-range dependency capture. Efficient variants, such as FEDformer (T. Zhou et al., 2022) and Crossformer (S. Li & Cai, 2024) further enhanced performance through frequency-domain representation and cross-variable attention. Nevertheless, these sequence-based models inherently lack mechanisms to explicitly represent dynamic inter-variable dependencies, which are crucial for multivariate forecasting tasks.

2.2. Graph Neural Networks for Time Series Forecasting

GNNs have emerged as a powerful paradigm for modeling structured dependencies among time series variables. Foundational works such as the Graph Convolutional Network (Kipf & Welling, 2016) and subsequent surveys (Wu et al., 2021) have established the mathematical underpinnings and practical applications of GNNs across multiple domains. Seminal models such as spatio-temporal graph convolutional networks (STGCN) (B. Yu, Yin, & Zhu, 2018) and Graph WaveNet (Wu, Pan, Long, Jiang, & Zhang, 2019) combined spectral graph convolution with temporal modules, introducing adaptive adjacency mechanisms to learn latent spatial relationships. Dynamic graph approaches, such as EvolveGCN (Pareja et al., 2020), employ recurrent

units to evolve graph structures over time. Meanwhile, the attention mechanism pioneered by the graph attention network (GAT) (Veličković et al., 2017) has been incorporated into spatiotemporal models (e.g., Kim et al., 2023; Ding et al., 2023) to capture global correlations. Despite these advances, existing approaches often rely on homogeneous aggregation and fixed-scale graphs, limiting the ability to adapt to distribution shifts across channels or to capture locally sparse yet globally connected dependencies.

In the prognostics and health management (PHM) domain, modeling dynamic multi-scale interactions remains a key challenge that constrains performance improvement in RUL prediction and fault diagnosis (Lei et al., 2018). Recent studies highlight a growing trend toward integrating GNNs, attention mechanisms, and domain knowledge to enhance interpretability and robustness in health assessment systems (Su, H., & Lee, J., 2023; Kumar et al., 2024). This integration underscores the unique value of adaptive graph reasoning in complex industrial monitoring tasks. Motivated by these challenges, this paper proposes CMAGIN for spatiotemporal time series forecasting in RUL prediction applications. The model comprises two core components: a DAGC-LAGA and a CACE module. The former enriches structural information for node representation by integrating both locally sparse and globally attentive graph structures, while the latter enhances feature stability and discriminability by introducing a global center node to guide cross-channel feature fusion.

3. DESIGN OF THE CMAGIN FRAMEWORK

To address the complex and dynamic dependencies among variables in MTS data and the uneven distribution of importance across temporal feature channels, this work proposes a CMAGIN for MTS forecasting. The overall architecture is shown in Figure 1. The input MTS data first flows through the DAGC-LAGA module, which dynamically constructs and refines graph structures to capture both local and global dependencies among variables. Within this module, MV-LDGC models sparse dependencies that change over time, while GAGE complements these with long-range cross-channel interactions. The resulting representations are then fed into the CACE module, which assesses the node importance along the channel dimension and selectively enhances key channel features via a central node guidance mechanism. Together, these modules enable CMAGIN to effectively capture complex temporal dynamics while adaptively emphasizing critical channel features.

3.1. The Construction of DAGC-LAGA

The spatio-temporal dependencies among variables in MTS data are complex and dynamically evolve over time. Traditional methods often rely on static or single-scale graph structures, making it difficult to accurately capture multi-order dependencies and multi-scale interactions that change over time, thereby limiting their modeling capabilities. To

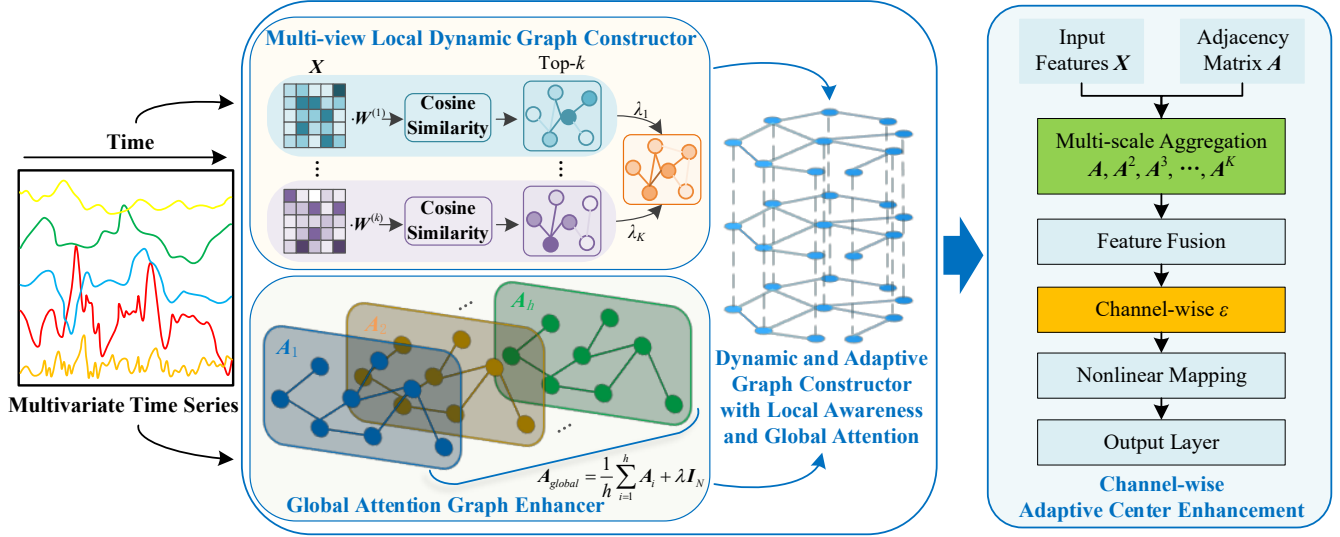


Figure 1. Overall structure of CMAGIN

address this issue, this paper proposes a DAGC-LAGA. This module consists of two components: the MV-LDGC and the GAGE. The MV-LDGC dynamically constructs multiple sparse graph structures based on variable similarity within local temporal windows, aiming to capture the evolving characteristics of short-term neighborhood relationships. Meanwhile, GAGE employs a cross-channel attention mechanism to model global semantic dependencies among variables, generating graph structures that reflect such global interactions. On this basis, the model introduces learnable graph fusion weights to adaptively integrate both local and global graph structures, enabling dynamic modeling of multi-scale and time-varying dependencies.

3.1.1. Multi-view Local Dynamic Graph Constructor

In MTS modeling, the dependencies among variables not only evolve over time but also exhibit diverse local interaction patterns. However, existing methods typically rely on static and single-perspective graph structures to model variable relationships, which limits the ability to effectively capture locally evolving interactions. To address this issue, this work proposes the MV-LDGC, which dynamically constructs multiple sparse graph structures from different perspectives. These graphs are then adaptively integrated through a learnable fusion mechanism, enabling the model to capture diverse local interactions among multivariate variables. This approach enhances the modeling capacity for short-term local dependencies and improves the robustness to noise in complex and dynamic MTS data.

Inspired by multi-view graph learning methods, the node feature matrix $X \in \mathbb{R}^{N \times d}$ and K subgraph spaces are introduced. This paper maps the input features into multiple subspaces via a set of linear transformations, generating feature representations $X^{(k)} = XW^{(k)}$, $k = 1, 2, 3, \dots, K$

corresponding to each subgraph. Specifically, $W^{(k)} \in \mathbb{R}^{d \times d}$ denotes the learnable projection matrix for the k -th subgraph, and $X^{(k)} \in \mathbb{R}^{N \times d}$ represents the node features in the k -th subspace. The multi-space projection generates multiple feature subspaces in parallel through linear transformations, capturing diverse latent relationship patterns in MTS. Subsequently, the topological relationships of nodes within each subgraph space are computed independently, where an improved cosine similarity is employed to calculate the association strength between node i and node j . Thus, the adjacency matrix $A^{(k)}$ of the k -th subgraph is defined as

$$A^{(k)} = \sigma \left(\frac{X^{(k)} \cdot (X^{(k)})^\top}{\|X^{(k)}\|_2 \cdot (\|X^{(k)}\|_2)^\top + \epsilon} \right) \quad (1)$$

Where $\|X^{(k)}\|_2$ denotes the operation of computing the L2 norm row-wise (i.e., for each node's feature vector). ϵ is the small constant added to avoid division by zero, and $\sigma(\cdot)$ denotes a LeakyReLU activation to enhance nonlinear representation capability and ensure numerical stability.

To reduce computational complexity and suppress noise interference, while alleviating the over-smoothing problem in graph structures, a Top- k sparsification strategy is applied to the adjacency matrix of each subgraph. Specifically, for each node, only the connections to the top k most similar nodes are retained, and the remaining connections are set to zero. This approach preserves significant local connections while reducing the computational complexity from $O(N^2)$ to $O(kN)$. The computation is formulated as follows.

$$\hat{A}_{ij}^{(k)} = \begin{cases} A_{ij}^{(k)} & \text{if } j \in N_k(i) \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where the neighbor set $N_k(i)$ of node i is selected using the Top- k method: $N_k(i) = \text{argtopk}\{\mathcal{A}_{ij}^{(k)} \mid j \neq i\}$.

To capture the diverse local dependency patterns characterized by different subgraphs and dynamically balance each view's contribution to the overall graph structure, this paper introduces learnable graph weight parameters $\lambda \in \mathbb{R}^k$ to perform a weighted fusion of multiple subgraph adjacency matrices $\hat{\mathcal{A}}^{(k)}$. This fusion captures the local interaction patterns across subgraphs, resulting in an adjacency matrix $\mathcal{A}_{local}$ that represents diverse local dependencies.

$$\mathcal{A}_{local} = \sum_{k=1}^K \lambda_k \hat{\mathcal{A}}^{(k)} \quad (3)$$

where λ_k is initialized with a uniform distribution $\lambda_k = \frac{\exp(\omega_k)}{\sum_{l=1}^K \exp(\omega_l)}$, and $\{\omega_k\}_{k=1}^K$ is the trainable parameter optimized jointly with the network through backpropagation.

3.1.2. Global Attention Graph Enhancer

In MTS modeling, variables exhibit complex and global semantic dependencies that often extend beyond local neighborhoods. Traditional local graph structures struggle to effectively capture these dependencies. Existing methods primarily focus on local adjacency relationships and lack systematic modeling of global semantic correlations, resulting in models that struggle to fully understand and leverage long-range dependencies and global interactions among variables.

To address this limitation, this paper designs the GAGE module. This module dynamically uncovers global semantic correlations among variables through a cross-channel self-attention mechanism, constructing a graph structure that reflects the global dependencies of multivariate data. Specifically, the module computes similarity weights between all node pairs, strengthening interactions between distant nodes and effectively supplementing the global information that local graphs may miss. Leveraging multi-head attention, the GAGE module establishes global node connections, resolving long-range dependencies and latent relationships potentially overlooked during dynamic multi-graph construction. This provides a global topology correction for the graph structure. Figure 2 illustrates the working mechanism of the graph attention enhancement.

First, the input features are linearly projected through a shared linear transformation layer to generate three feature matrices: Query ($\mathcal{Q}$), Key ($\mathcal{K}$), and Value ($\mathcal{V}$). Then, these three matrices are split into h heads by a chunking operation, forming multiple sets of feature representations required for the multi-head attention mechanism. For each attention head,

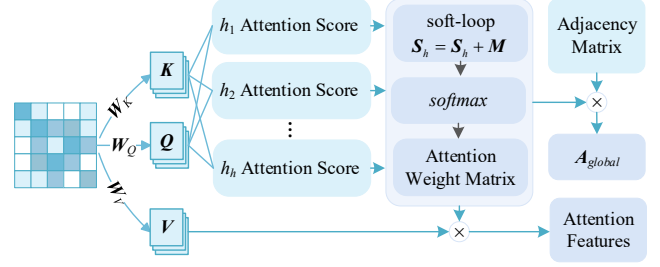


Figure 2. Working mechanism of GAGE

the scaled dot-product attention score $\mathcal{S}_h$ is independently computed as

$$\mathcal{S}_h = \frac{\mathcal{Q}_h \mathcal{K}_h^T}{\sqrt{d/h}} \quad \forall h \in \{1, \dots, h\} \quad (4)$$

where the scaling factor $\sqrt{d/h}$ is used to control gradient stability and prevent the gradient vanishing problem caused by excessively large variance in the dot product results. As the dimension d/h increases, the variance of the dot product also increases. the scaling operation helps stabilize the training process.

In practical attention computation, a node often exhibits the highest relevance when querying itself, leading to an attention distribution that is heavily concentrated on self-connections. This reduces the model's ability to capture potential global dependencies among different nodes. Therefore, to prevent local self-connections from dominating the relational learning and to encourage the model to focus more on cross-node interactions, this paper introduces a self-loop suppression mechanism by constructing a Boolean mask matrix $\mathcal{M} \in \{0, -\infty\}^{N \times N}$, where the diagonal elements are set to $-\infty$ and the remaining elements are set to 0. After suppressing the diagonal elements in the attention score matrix $\mathcal{S}_h$, the masked score matrix $\mathcal{S}_h = \mathcal{S}_h + \mathcal{M}$ is obtained. In implementation, the $-\infty$ mask is incorporated into the attention logits before normalization, effectively nullifying the contribution of self-loop terms during global dependency modeling. Applying the softmax operation to $\mathcal{S}_h$ produces the normalized global attention weight matrix $\mathcal{A}_h = \text{softmax}(\mathcal{S}_h)$, in which the diagonal elements tend toward zero after softmax. Here, $\mathcal{A}_h$ denotes the global dependency strength of node i on node j .

To ensure that each node retains at least its information propagation path and to enhance the structural stability of the attention graph, this paper explicitly introduces self-loop connections after aggregating multi-head attention weights. This yields the final globally enhanced attention adjacency matrix, denoted as

$$\mathcal{A}_{global} = \frac{1}{h} \sum_{i=1}^h \mathcal{A}_i + \lambda \mathcal{I}_N \quad (5)$$

where, A_i denotes the global attention matrix from the i -th attention head. I_N is the identity matrix, and λ is the self-loop strength coefficient used to control the weight proportion of self-loop connections.

3.1.3. Adaptive Fusion Mechanism

In MTS data, a single graph construction approach often struggles to balance the fine-grained modeling of local structures and the comprehensive representation of global dependencies. The former emphasizes the sparsity and temporal variability of adjacency relations, while the latter focuses on global semantic consistency. To alleviate this modeling bias, this paper proposes an adaptive fusion mechanism designed to dynamically integrate the structural information from both local dynamic graphs and global attention graphs, thereby achieving unified global-local modeling.

The sparse local adjacency matrix $A_{local} \in \mathbb{R}^{N \times N}$ output by the multi-view local dynamic graph constructor and the dense global matrix $A_{global} \in \mathbb{R}^{N \times N}$ generated by the attention-based enhancement module are adaptively fused. The resulting dynamic adaptive adjacency matrix constructed in this work is defined as

$$A_f = \text{Norm}(\alpha \cdot A_{local} + (1-\alpha) \cdot A_{global}) \quad (6)$$

where the normalization operation $\text{Norm}(\cdot)$ is calculated as $\text{Norm}(A) = D^{-1/2} A D^{-1/2}$, $D_{ii} = \sum_{j=1}^n A_{ij}$, ensuring numerical stability during graph signal propagation and alleviating the problem of gradient explosion. The fused adjacency A_f is subsequently used for graph convolution in the following layer. $\alpha \in [0,1]$ denotes a learnable balancing parameter which is initialized to 0.5 and optimized via backpropagation to dynamically adjust the contribution of local dependencies and global information. The gradient update rule is

$$\frac{\partial L}{\partial \alpha} = \sum_{i,j} \frac{\partial L}{\partial A_{ij}} (A_{local,ij} - A_{global,ij}) \quad (7)$$

This formula indicates that if the relationship between a pair of nodes in the local graph contributes more to reducing the overall loss, the corresponding gradient will drive an increase in α , thereby enhancing the weight of the local graph in the fusion. Otherwise, the contribution of the global graph will be increased.

At this point, the dynamic adaptive graph construction is complete. This module combines dynamic multi-graphs with attention graphs to eliminate the perspective bias inherent in a single graph construction method, achieving a unified representation of local precision and global completeness. As a result, the graph neural network can simultaneously capture both the local structural features and the global semantic relationships of the graph data.



Figure 3. Channel enhancement mechanism

3.2. Channel-wise Adaptive Center Enhancement

In MTS graph modeling, different feature channels often exhibit statistical disparities, such as varying unit scales, inconsistent dynamic change frequencies, or uneven distributions of semantic contributions. This heterogeneity can lead to issues like uneven overfitting, feature confusion, or redundant aggregation when graph neural networks treat the central node features of all channels uniformly. Traditional graph isomorphism networks (GIN) (K. Xu, Hu, Leskovec, & Jegelka, 2018) typically use a single globally shared scalar parameter ϵ to control the weight of the node's features during information aggregation, which fails to fully express the personalized importance of each channel and limits the model's expressive capacity.

To address these issues, this paper proposes the CACE mechanism. This module integrates the channel-adaptive characteristics of GIN with a multi-scale neighborhood aggregation strategy. Through learnable channel-level parameters ϵ , it adaptively weights the central nodes across feature dimensions, enabling the model to independently learn the enhancement degree of the center node for each feature dimension. As a result, it dynamically adjusts the influence of center nodes in different channels, better accommodating the heterogeneous distributions of multivariate features.

3.2.1. Multi-scale Neighborhood Aggregation

In graph neural networks, the update of node representations is typically based on propagation through adjacency matrices of fixed orders (e.g., first-order or second-order), which limits the model's ability to express graph data with varying structural complexities. Especially in dynamic systems or heterogeneous structures, fixed multi-hop neighbor aggregation strategies may overlook certain crucial structural semantic information.

To enhance the model's adaptability to structural diversity, the module captures the l -hop neighborhood information through the power of the adjacency matrix A^l . Each order adjacency matrix is symmetrically normalized with self-loops as $A^{(l)} = \text{Norm}(A^l + I)$. Then, for each scale $l \in \{1, \dots, L\}$, a feature transformation is performed, and the

aggregated neighbor features based on the k -th order adjacency matrix are given by

$$\mathbf{H}^{(l)} = \Theta^{(l)}(\mathbf{A}^{(l)}\mathbf{X}) \quad (8)$$

where $\Theta^{(k)}$ is the learnable linear transformation for the k -th order aggregation.

3.2.2. Multi-scale Feature Fusion

Multi-scale neighborhood information can be fused through either concatenation or summation.

$$\mathbf{H} = \begin{cases} \text{concat}(\mathbf{H}^{(1)}, \mathbf{H}^{(2)}, \dots, \mathbf{H}^{(L)}) \\ \sum_{l=1}^L \mathbf{H}^{(l)} \end{cases} \quad (9)$$

Additionally, to maintain consistency of feature dimensions, a projection layer Φ_{proj} is introduced to perform a projection transformation on the input features, ensuring that features of different scales can be effectively fused within the same space.

3.2.3. Channel-wise GIN Enhancement Mechanism

The traditional GIN uses a globally shared scalar ε to adjust the information fusion ratio between the central node and its neighboring nodes. However, this design cannot capture the varying degrees of dependency on the central node across different feature channels, limiting the model's expressive power on heterogeneous graphs.

To enhance the model's capability on structurally heterogeneous graphs, the module introduces channel-wise learnable ε parameters combined with multi-hop neighborhood information fusion, thereby achieving adaptive multi-scale information integration. Figure 3 illustrates the structure of this channel enhancement mechanism.

The original GIN aggregation formula is $\mathbf{H}^{(l+1)} = \text{MLP}\left((1 + \varepsilon) \cdot \mathbf{H}^{(l)} + \sum_{j \in N(i)} \mathbf{H}_j^{(l)}\right)$, where ε is the globally shared scalar parameter. This design cannot adapt to the varying degrees of dependency on the central node across different feature channels, which limits the model's expressive power. To enhance the model's representation capability, this work introduces a multi-scale neighborhood aggregation mechanism, extending the single-hop neighborhood summation to a weighted sum of multiple adjacency matrices $\sum_{l=1}^L \mathbf{W}_l \mathbf{H}^{(l)}$. Furthermore, the scalar ε is generalized to the diagonal matrix $\text{diag}(\varepsilon)$, i.e., $(1 + \varepsilon)\mathbf{X} \rightarrow (\mathbf{I} + \text{diag}(\varepsilon)) \odot \mathbf{X}\mathbf{W}_{proj}$, enabling channel-wise information modulation.

$$\mathbf{H}_{out} = \text{MLP}\left((\mathbf{I} + \text{diag}(\varepsilon)) \odot \mathbf{X}\mathbf{W}_{proj} + \sum_{l=1}^L \mathbf{W}_l \mathbf{H}^{(l)}\right) \quad (10)$$

where $\varepsilon \in \mathbb{R}^{d_{out}}$ is the learnable channel-adaptive vector optimized through backpropagation, enabling each channel to independently control the influence of its center node. This design improves interpretability by highlighting the relative importance of different sensor channels. For example, in the C-MAPSS dataset, channels with larger ε values may correspond to temperature and pressure sensors that dominate the degradation trend. The symbol $\odot$ denotes element-wise multiplication. $\mathbf{W}_{proj} \in \mathbb{R}^{d_{in} \times d}$ is a projection matrix used to learn differentiated interaction patterns of the central node features at various distances. $\mathbf{W}_l$ is the linear transformation matrix for the l -th order neighborhood features, and L denotes the maximum propagation order.

The final output stage is processed by an MLP layer with batch normalization and nonlinear activation, achieving stable feature modeling and enhanced expressive capability. By incorporating channel-wise parameterization and multi-scale aggregation mechanisms, this module retains the strong expressive power of GIN while improving the model's adaptability to complex graph structures and heterogeneous node relationships.

4. EXPERIMENT

4.1. Dataset Description and Experimental Setup

This study utilizes the commercial modular aero-propulsion system simulation (C-MAPSS) dataset (Saxena, Goebel, Simon, & Eklund, 2008) to conduct remaining useful life (RUL) prediction for aero-engines, thereby validating the effectiveness and robustness of the proposed method in real-world industrial scenarios. The C-MAPSS dataset, provided by NASA Ames Research Center, simulates the operational and degradation processes of aircraft engines under various operating conditions and fault modes, and has been widely used in intelligent predictive maintenance research. The dataset comprises four subsets: FD001, FD002, FD003, and FD004. Each subset corresponds to different numbers of operating conditions and fault types, with progressively increasing task complexity. In each subset, 21 sensors are deployed to comprehensively monitor the operational status of each engine, collecting and recording key physical parameters in real-time, such as temperature, pressure, and rotational speed. Basic information of the C-MAPSS dataset is summarized in Table 2.

To comprehensively assess the performance of CMAGIN on the RUL prediction task, this study follows prior work (Y. Wang, Y. Xu, et al., 2023) and adopts two commonly used regression metrics: root mean square error (RMSE) and Score function.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (11)$$

Datasets	FD001	FD002	FD003	FD004
Training	100	260	100	249
Testing	100	295	100	248
Fault Types	1	1	2	2
Sensors	21	21	21	21
Conditions	1	6	1	6

Table 2. Basic information of the C-MAPSS dataset

$$\text{Score} = \sum_{i=1}^n \xi_i, \xi_i = \begin{cases} e^{\frac{\hat{y}_i - y_i}{13}} - 1, & \text{if } \hat{y}_i - y_i < 0 \\ e^{\frac{\hat{y}_i - y_i}{10}} - 1, & \text{if } \hat{y}_i - y_i \geq 0 \end{cases} \quad (12)$$

where n is the number of samples. $\hat{y}_i$ and y_i denote the predicted and true RUL of the i -th sample respectively. RMSE measures the average deviation between predicted and true values, reflecting overall prediction accuracy. The Score function applies weighted penalties to prediction errors, imposing stronger penalties on early or late predictions, thereby aligning more closely with practical engineering requirements. Lower RMSE and Score values indicate better predictive performance and greater model robustness.

The experiments consist of two parts: comparative studies against state-of-the-art (SOTA) methods and ablation studies on the key modules of the proposed model. To ensure result stability and reliability, each model is independently run 10 times, and the average performance metrics are reported. All experiments are conducted on a system equipped with an NVIDIA GeForce RTX A6000 GPU. The model is implemented using the PyTorch 1.9 framework. The Adam optimizer is used for training, with a minimum of 20 epochs. Other training hyperparameters are tuned based on validation set performance. For fair comparison, all baseline models were reimplemented or fine-tuned using identical training configurations, including optimizer type, learning rate, batch size, and early stopping criteria. Where available, official implementations were employed and trained under the same experimental conditions to ensure methodological consistency.

4.2. Comparison with SOTA Models

To assess the effectiveness of CMAGIN in temporal graph modeling tasks, this subsection presents a systematic comparison with representative SOTA methods, including those based on graph neural networks, graph convolution techniques, graph pooling strategies, and conventional temporal deep learning architectures. The methodological details, performance results, and RMSE box plots are shown in Table 3 and Figure 4, respectively.

The experimental results demonstrate that the proposed method consistently outperforms all baseline models across all four C-MAPSS sub-datasets (FD001–FD004), achieving

superior performance in both prediction accuracy (RMSE) and cost-sensitive metric (Score). On the FD001 dataset, characterized by a single operating condition and a single fault mode, CMAGIN achieves the lowest RMSE of 10.99 and Score of 206, representing a 9.0% and 18.6% improvement over the second-best model FCSTGNN (Y. Wang et al., 2024). This advantage likely results from FCSTGNN's limited ability to integrate temporal dynamics in graph modeling. The proposed DAGC-LAGA and CACE modules effectively enhance the central node's representation and dynamically capture multi-scale neighbor relationships, leading to higher prediction accuracy in simpler fault scenarios.

The FD003 dataset involves a single condition and compound faults. CMAGIN continues to demonstrate strong robustness, achieving the best RMSE (11.09) and Score (199) among all compared models. This further indicates that CMAGIN has the capability to handle complex fault patterns. Even on the more challenging multi-condition datasets FD002 and FD004, CMAGIN maintains a clear advantage. On FD002, which involves multiple operating conditions and a single fault mode, CMAGIN achieves the lowest RMSE (12.80) and Score (711). It outperforms FCSTGNN (RMSE = 13.27, Score = 777), suggesting a strong ability to capture degradation patterns under varying conditions with high accuracy and cost awareness. On FD004, featuring multiple conditions and compound faults, CMAGIN again achieves the best results (RMSE = 13.58, Score = 759). It reduces the Score by approximately 15.6% compared to FCSTGNN (RMSE = 14.06, Score = 899), further confirming its robustness under high-noise, high-dimensional, and heterogeneous input scenarios.

CMAGIN outperforms existing representative approaches across multiple evaluation metrics, demonstrating its effectiveness and advancement in addressing the problem of RUL prediction for industrial equipment. The temporal modeling approach AConvLSTM (Xiao et al., 2021) excels in capturing temporal dependencies during the degradation process but fails to explicitly model spatial dependencies among sensors, making it difficult to account for structural heterogeneity. Classical graph convolutional methods, such as GCN (Kipf & Welling, 2016) and HAGCN (T. Li et al., 2021), exhibit certain capabilities in modeling spatial dependencies among sensors. However, the overall predictive performance remains limited. One possible reason is the reliance on static graph structures, which struggle to adapt to dynamically evolving dependencies during the degradation process. Another reason may be the insufficient ability to model temporal evolution, leading to inadequate integration of time-series information under non-stationary conditions. In contrast, the DAGC-LAGA module in CMAGIN dynamically and adaptively combines sparse local structures with global relational modeling, enabling the graph structure to update dynamically during training and effectively capturing node interactions and latent

Datasets	FD001		FD002		FD003		FD004	
Index	RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score
GCN	12.68±0.29	245±27	13.77±0.21	833±54	12.01±0.17	227±41	14.39±0.39	986±72
iPool	12.35±0.24	261±37	13.17±0.21	775±43	12.33±0.28	258±47	14.52±0.22	1063±102
TAP	12.37±0.16	<u>221±26</u>	13.18±0.18	<u>747±260</u>	12.40±0.26	246±23	14.41±0.25	<u>899±37</u>
HAGCN	13.42±0.25	302±33	14.55±0.27	1035±55	13.47±0.28	374±59	14.69±0.37	995±87
AConvLSTM	12.42±0.31	279±28	13.21±0.51	845±161	12.66±0.46	311±38	15.72±1.89	1516±790
HierCorrPool	12.18±0.13	250±15	<u>13.08±0.13</u>	754±9	12.04±0.07	<u>218±14</u>	14.23±0.20	928±48
MAGNN	12.79±0.22	256±22	13.35±0.21	798±33	12.28±0.18	302±41	14.46±0.29	1045±79
FCSTGNN	<u>12.08±0.17</u>	253±18	13.27±0.24	777±51	<u>11.96±0.26</u>	254±37	<u>14.06±0.28</u>	899±54
CMAGIN	10.99±0.18	206±17	12.80±0.18	711±26	11.09±0.15	199±15	13.58±0.24	759±36

Table 3. Comparison with SOTA models on the C-MAPSS dataset

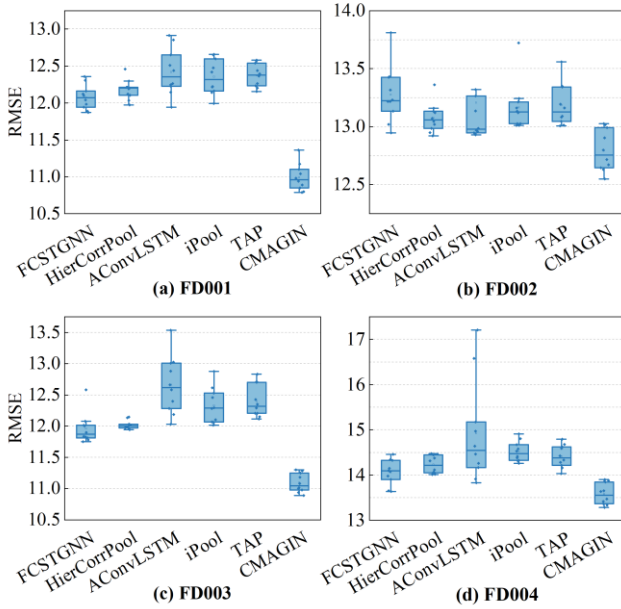


Figure 4. Comparative box plots of model performance

dependencies under non-stationary and variable operating conditions.

Advanced GNN-based methods, such as MAGNN(L. Chen et al., 2023) and FCSTGNN, integrate structural and temporal information to some extent. MAGNN enhances spatial modeling by incorporating heterogeneous semantic information, but its complex architecture is prone to overfitting on small- and medium-scale industrial datasets. Graph pooling-based methods, including iPool (X. Gao et al., 2021), TAP (Gao, Liu, & Ji, 2021), and HierCorrPool (Y. Wang, Wu, Li, Xie, & Chen, 2023), achieve better performance than GCN in some scenarios, but still fail to stably model complex structures. This limitation is mainly due to the inevitable loss of critical node information during graph pooling, which becomes particularly pronounced under multi-operating conditions or compound fault scenarios, leading to performance degradation. Although FCSTGNN

demonstrates strong spatiotemporal modeling capabilities, its static graph construction lacks contextual adaptability, making it difficult to dynamically capture evolving inter-sensor relationships and constraining its performance under complex working conditions. The CACE module in CMAGIN enhances the expressive power of central nodes across different feature channels, enabling precise characterization of critical node information across time and channels.

In summary, experimental results across multiple C-MAPSS subsets demonstrate that the proposed model consistently surpasses compared methods in terms of prediction accuracy and engineering adaptability, particularly under complex operational scenarios. Future work may further explore its performance in cost-sensitive tasks and investigate structural compression and lightweight deployment strategies to support efficient application in real-world industrial environments.

To validate the reliability of the observed performance improvement, paired t-tests are conducted between CMAGIN and the best-performing baseline method (FCSTGNN) across 10 independent experimental runs. The results show statistically significant improvements in both RMSE and Score metrics ($p < 0.05$), confirming that the performance gains are not due to random variations.

4.3. Ablation Study

This section conducts ablation experiments to verify the effectiveness of the key modules proposed in this work, including the MV-LDGC module, the GAGE module, and the CACE module. Under the same experimental settings, each of these modules is sequentially removed or replaced to observe its impact on the overall model performance. The specific configurations are as follows.

(1) “w/o GAGE” denotes the removal of the GAGE module, retaining only the MV-LDGC module for graph structure

Datasets	FD001		FD002		FD003		FD004	
Index	RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score
w/o GAGE	11.38±0.21	<u>214±17</u>	<u>12.98±0.19</u>	752±22	11.40±0.19	228±19	13.76±0.17	<u>762±45</u>
w/o MV	<u>11.28±0.19</u>	219±15	13.01±0.21	791±18	11.47±0.18	297±18	<u>13.68±0.20</u>	811±39
w/o CACE	11.47±0.20	233±18	13.03±0.24	819±20	<u>11.24±0.17</u>	<u>224±15</u>	13.89±0.19	871±42
w/o DA&CA	11.87±0.23	248±22	13.15±0.22	<u>720±24</u>	11.75±0.21	232±20	13.94±0.22	893±51
CMAGIN	10.99±0.18	206±17	12.80±0.18	711±26	11.09±0.15	199±15	13.58±0.24	759±36

Table 4. Ablation Study on the C-MAPSS Dataset

building, to evaluate the role of the global attention mechanism in capturing long-range node dependencies.

(2) “w/o MV” indicates the removal of the MV-LDGC module while keeping the GAGE module, aimed at examining the model performance in the absence of local perceptual capability.

(3) “w/o CACE” refers to retaining both the MV-LDGC and GAGE modules, but replacing the CACE module with a conventional graph convolutional network, to assess the contribution of CACE in characterizing key node representations across channels.

(4) “w/o DA&CA” involves simultaneously removing the MV-LDGC module, the GAGE module, and the CACE module, substituting them with dot-product attention-based graph construction and standard graph convolutional networks, to evaluate the overall effectiveness and synergistic benefits of the proposed method.

Table 4 presents the comparative results of these model variants against the full model across multiple evaluation metrics.

The ablation results for “w/o GAGE” show that removing the GAGE module leads to varying degrees of performance degradation across all subsets compared to the full model. Notably, on the multi-fault subset FD003, the Score metric increases from 199 to 228, a rise of approximately 14.6%. This indicates that the GAGE module plays a significant role in modeling long-range dependencies across nodes, thereby enhancing the model’s representational capacity under complex fault conditions. A similar performance decline is also observed on the multi-condition, multi-fault subset FD004, further demonstrating the positive impact of global structure modeling on overall model performance.

Analysis of the “w/o MV” results reveals that the Score values on the multi-condition datasets FD002 and FD004 increase from 711 and 759 to 791 and 811, respectively, suggesting that removing the MV-LDGC module weakens the model’s adaptability to varying operating conditions. This indicates the module’s contribution to modeling heterogeneity across conditions. In contrast, performance changes on the single-condition datasets FD001 and FD003 are relatively minor, implying that this module has a lesser impact on tasks with limited condition variability.

Additionally, on FD001, the “w/o MV” configuration slightly outperforms “w/o GAGE” in terms of RMSE, suggesting that the global attention mechanism may be more beneficial in simpler, single-condition scenarios.

In the “w/o CACE” experiment, replacing the CACE mechanism with a conventional graph convolution results in overall performance decline across all subsets. For instance, the Score on FD004 rises from 759 to 871, an increase of about 14.7%. This result indicates that the CACE module, through channel-wise modulation and multi-scale aggregation, enhances the model’s ability to emphasize important node features, thereby improving overall prediction accuracy. Although the RMSE on FD003 is slightly better than some other variants, the overall trend confirms the module’s advantage in modeling complex spatiotemporal features.

When the GAGE, MV-LDGC, and CACE modules are all removed simultaneously, the model’s performance deteriorates substantially, achieving the worst results across all subsets. For example, on FD001, the Score increases from 206 to 248, highlighting the indispensable role of these key modules in structural modeling and feature learning. Compared with the “w/o DA&CA” configuration, the whole model achieves a 20.4% improvement in Score on FD001 and attains the best performance across all datasets, confirming the effectiveness and stability of the multi-module integrated design under diverse scenarios.

In summary, the three key modules proposed in this study demonstrate notable effectiveness in MTS modeling tasks.

(1) The GAGE mechanism introduces global contextual information, thereby enhancing the modeling of long-range dependencies among nodes. This is particularly beneficial in subsets characterized by multiple fault interferences and complex feature correlations, as it improves the model’s capability to capture critical fault-related information.

(2) The MV-LDGC module contributes to the model’s adaptability to varying operating conditions, offering advantages in scenarios involving diverse working environments.

(3) The CACE mechanism, through the integration of channel-weight modulation and multi-scale aggregation

strategies, further strengthens the representation of key nodes across channels during the feature fusion process.

Collectively, the integration of these three modules enables the proposed model to achieve stable and accurate prediction performance under varying fault types and operational conditions.

4.4. Parameter Sensitivity Analysis

To further investigate the impact of key parameters on the performance of CMAGIN, this work conducts a systematic sensitivity analysis using the C-MAPSS aircraft engine degradation datasets (FD001–FD004). Specifically, the effects of three parameters are examined: the number of subgraphs (default value: 3), the number of attention heads (default value: 4), and the top- k sparsification degree (default value: 5). Under a controlled experimental setup where all other configurations remain unchanged, each parameter is varied individually following a univariate control strategy. The model's performance on each subset is recorded across different parameter settings to evaluate the sensitivity and identify the optimal configuration for each parameter.

4.4.1. Number of Subgraphs

To evaluate the effect of the number of subgraphs on model performance within the dynamic multi-graph construction module, the number of subgraphs is systematically varied. Here, “1 graph” corresponds to a single global graph, while “3/5 graphs” represent the construction of fused multi-graph structures. Under consistent training settings, comparative experiments are conducted on the four C-MAPSS subsets (FD001–FD004), using RMSE and Score as evaluation metrics to assess predictive performance. The grouped column chart in Figure 5 reveals the optimal performance variation patterns of the model under different subgraph configurations.

It can be observed that the influence of subgraph quantity on model performance is closely related to the operational complexity and degradation characteristics of the dataset. On datasets with relatively simple operating conditions, such as FD001 and FD003, the model achieves the best predictive performance when three subgraphs are used, with RMSE values of 10.79 and 10.89 and Score values of 193.3 and 183.35, respectively, outperforming both the single-graph and five-graph configurations. This indicates that moderate multi-graph fusion effectively captures multiple degradation patterns exhibited by the engines under various conditions, including short-term local correlations, long-term dependency trends, and cross-channel commonalities.

However, when the number of subgraphs increases to five, a noticeable performance drop is observed on FD001, which may be attributed to the introduction of redundant or noisy information, thereby increasing the risk of overfitting. Notably, on the more complex FD002 dataset, performance

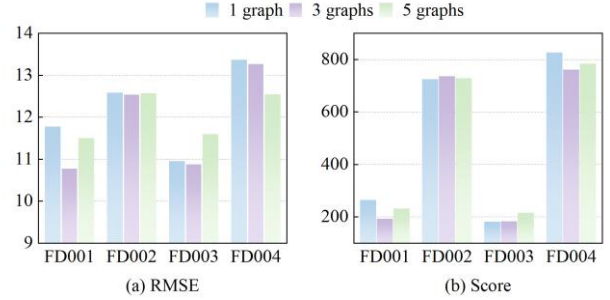


Figure 5. Sensitivity analysis of the number of subgraphs

differences among different subgraph configurations are marginal, which aligns with the relatively stable sensor relationships in this subset. In the larger FD004 dataset, RMSE and Score exhibit inconsistent trends: while five subgraphs yield the lowest RMSE, the best Score is still obtained with three subgraphs. This discrepancy may stem from the varying sensitivity of these metrics to different degradation stages.

Overall, the experimental results suggest that using three subgraphs provides a balanced and effective configuration across the datasets employed in this study. Therefore, this setting is adopted as the default in the experiments conducted in this work.

4.4.2. Number of Attention Heads

A systematic sensitivity analysis was conducted to investigate the impact of the number of attention heads, which is a key hyperparameter, on the model's performance in aircraft engine degradation prediction. As illustrated in Figure 6, comparative experiments were performed on the four C-MAPSS subsets (FD001 to FD004) under three configurations with 2, 4, and 8 attention heads, respectively. The predictive performance was evaluated using both RMSE and Score metrics.

Experimental results reveal a clear non-linear relationship between the number of attention heads and model performance. On the FD001 dataset, the 4 heads configuration achieved the best results, with an RMSE of 10.79 and a Score of 193.3, showing 2.0% and 3.6% improvements respectively compared to the 2 heads baseline. However, further increasing the number to 8 heads led to a significant performance drop, with RMSE and Score deteriorating by 6.3% and 9.4%, respectively. This suggests that a moderate number of attention heads can effectively model multi-dimensional feature interactions, while an excessive number may introduce redundancy and increase the risk of overfitting.

In contrast, the FD002 dataset exhibited a different pattern. While the lowest RMSE was still achieved with 4 heads (12.55), the 8 heads configuration yielded the lowest Score (657.96), outperforming the 4-head setting by approximately 10.6%. This inconsistency between RMSE and Score may

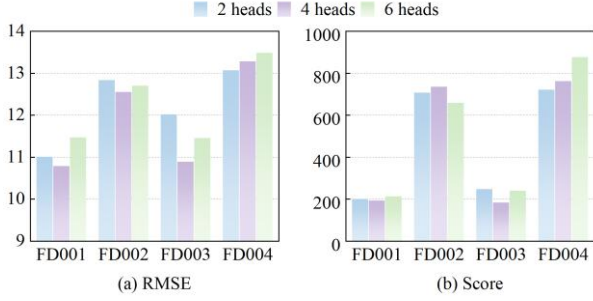


Figure 6. Sensitivity analysis of the number of attention heads

stem from the dataset's specific operational characteristics, where a greater number of attention heads helps capture more complex feature dependencies, thereby improving overall predictive quality.

The FD003 dataset showed a similar trend to FD001, with the 4-head configuration delivering the best results. Compared to the 2 heads setup, RMSE and Score improved by 9.4% and 25.8%, respectively, further confirming the modeling advantages of a moderate number of attention heads for tasks with medium-level complexity. However, on the FD004 dataset, model performance deteriorated consistently as the number of attention heads increased. In particular, the 8 heads configuration resulted in the worst Score (875.97), likely due to the higher noise levels and lower effective feature dimensionality in FD004, which causes multi-head attention to introduce redundant or interfering information.

In summary, the optimal number of attention heads appears to be closely related to the complexity and feature characteristics of the dataset. For the C-MAPSS subsets used in this study, the 4 heads configuration consistently demonstrated superior performance and is therefore adopted as the default setting in the experiments conducted in this work.

4.4.3. Top- k Sparsification

To investigate the impact of the top- k sparsification strategy in the dynamic graph construction module on model performance, this study systematically evaluates the prediction results under three different sparsity levels (Top-3, Top-5, and Top-10) on the four C-MAPSS sub-datasets from FD001 to FD004. As illustrated in Figure 7, the experimental results reveal that the degree of sparsity in the constructed graph structure significantly influences degradation prediction performance, and the effect varies depending on dataset characteristics.

On the FD001 dataset, the Top-5 sparsification configuration achieved optimal performance with an RMSE of 10.79 and a Score of 193.3, demonstrating 6.6% and 4.7% improvements over the Top-3 setting, respectively. However, further increasing k to 10 resulted in performance degradation. This suggests that a moderate sparsification level can effectively

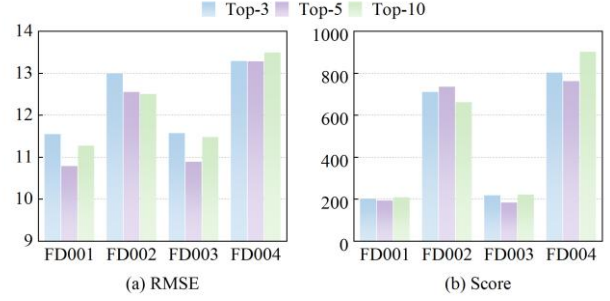


Figure 7. Sensitivity analysis of top- k sparsification

retain critical connections while filtering out noisy edges, thereby enhancing the model's representational capacity. In contrast, excessive sparsity (Top-3) or density (Top-10) could impair performance.

Interestingly, the FD002 dataset exhibited a different trend, where Top-10 sparsification yielded the best Score of 661.77, showing a 10.1% improvement over the Top-5 setting. This may be attributed to the dataset's more complex global dependencies, which benefit from preserving a greater number of edges to capture intricate relationships.

For the FD003 dataset, results were consistent with those observed on FD001. The Top-5 configuration again delivered superior performance, with RMSE and Score improvements of 5.9% and 16.1% over Top-3, respectively, confirming the advantage of moderate sparsification under moderately complex operating conditions.

However, for the FD004 dataset, model performance declined consistently as k increased, with the Top-10 setting resulting in the worst Score of 900.86. This indicates that the dataset may contain substantial noise or ineffective inter-sensor relationships, and that denser graph structures may introduce redundant or distracting information, thereby degrading model stability.

The experimental results demonstrate that the Top- k sparsification strategy significantly influences prediction performance, with its optimal configuration depending on the structural characteristics of the graph and the operational complexity of the dataset. In most scenarios, the Top-5 setting achieves a favorable balance by retaining informative connections while reducing noise, making it the default choice in this study. However, adjusting the sparsification level based on dataset-specific properties may further enhance performance in certain cases. These findings reveal an apparent "moderation effect" between sparsification and model performance, where both excessively sparse and overly dense graph structures degrade accuracy. This underscores the importance of carefully designing and tuning sparsification strategies in dynamic graph construction modules.

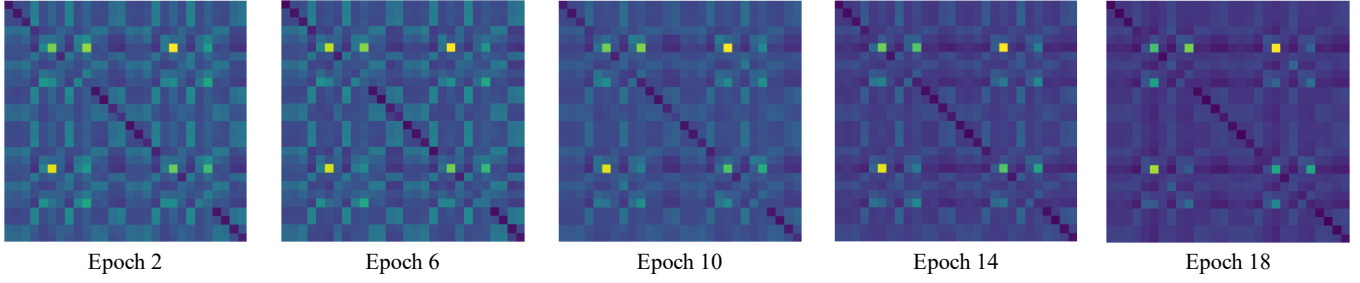


Figure 8. Visualization of learned adjacency matrices at different training epochs

4.5. Computational Complexity Analysis

The computational complexity of CMAGIN mainly arises from the DAGC-LAGA and the CACE modules.

(1) In the MV-LDGC of DAGC-LAGA, building K subgraphs requires $O(KN^2)$ pairwise similarity computations. After applying the Top- k sparsification, this cost is reduced to $O(KN \cdot k)$. In the GAGE, the multi-head attention mechanism introduces $O(hN^2d/h) = O(N^2d)$ operations for h attention heads, where d is the feature dimension. Thus, the overall time complexity of DAGC-LAGA per layer is approximately $O(kKN + N^2d)$, and the corresponding space complexity is $O(N^2 + Nd)$.

(2) In the CACE, the channel-wise enhancement mainly involves multi-scale aggregation and MLP transformations, whose cost is $O(LNd^2)$, where L is the number of scales.

(3) Combining the above, the total computational cost of CMAGIN per layer can be expressed as $O(kKN + N^2d + LNd^2)$, while the memory footprint remains $O(N^2 + Nd)$. Given that the C-MAPSS dataset contains only 21 sensor variables, CMAGIN operates efficiently and is easily scalable to larger systems. In practical industrial PHM scenarios, adjusting k , h , or L provides a linear trade-off between efficiency and accuracy, ensuring real-time applicability.

4.6. Interpretability Analysis

4.6.1. Channel-wise Interpretability

Compared with other SOTA graph-based forecasting models, CMAGIN provides stronger interpretability at both structural and channel levels. The CACE module employs a learnable vector ε to explicitly regulate each channel's central node influence. Analyzing the trained ε values quantifies the relative importance of different sensor channels. This offers physical insights into which variables dominate degradation processes. For example, channels with larger ε_i values in the C-MAPSS dataset typically correspond to temperature or pressure sensors exhibiting stronger degradation trends.

In contrast, most baseline methods like Graph WaveNet, AGCRN, and MTGNN use shared attention or weight matrices across all channels. These methods capture inter-

variable correlations but lack explicit interpretability for individual features. The channel-wise enhancement of CMAGIN thus provides a clearer, more physically meaningful explanation of model behavior while maintaining competitive prediction accuracy. Furthermore, visualizing the learned adjacency matrices from the DAGC-LAGA module reveals the evolution of local-to-global dependency pattern during training, highlighting how the model progressively refines its structural understanding of the system dynamics.

4.6.2. Visualization of Graph Structure Evolution

To further enhance the interpretability, the graph evolution during training is visualized by recording the learned adjacency matrices at selected epochs (epoch 2, 6, 10, 14, and 18). As shown in Figure 8, the adjacency matrices gradually evolve from sparse and disordered connections in the early stages to more structured and semantically meaningful patterns as training progresses. This indicates that the DAGC-LAGA progressively captures both local dependencies and global correlations among sensor nodes. The observed changes in graph topology provide an intuitive understanding of how the model refines its structural perception of the system over time.

Moreover, the learned adjacency matrices exhibit distinct block-like patterns, where dense submatrices emerge among specific groups of nodes. In the heatmap visualization, brighter colors (approaching yellow) represent stronger learned connection weights between two sensors, while darker regions (closer to blue or purple) indicate weaker or negligible interactions. Therefore, the emergence of bright, compact pixel blocks implies the formation of sensor clusters with strong mutual dependencies, suggesting that CMAGIN successfully identifies subsystem-level structures within the overall system.

This observation aligns with physical intuition. For instance, sensors measuring similar physical quantities (e.g., temperature and pressure within the same engine module) tend to exhibit higher correlations, forming bright sub-blocks in the adjacency matrix. The interpretable block structures highlight the CMAGIN's ability to uncover hierarchical and physically meaningful relationships among system variables.

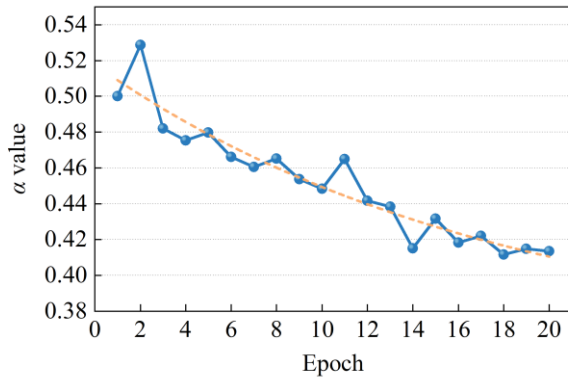


Figure 9. Evolution of α during training

4.6.3. Dynamics of the Fusion Coefficient α

To further interpret how CMAGIN balances local and global dependencies during training, the evolution of the learnable fusion coefficient α is monitored.

As shown in Figure 9, α gradually decreases from approximately 0.5 to around 0.41, following an exponential decay trend. Initially, $\alpha = 0.5$ indicates that both local sparse graphs and global attention graphs make comparable contributions. As training proceeds, α steadily decreases and stabilizes after about 15 epochs, implying that the model progressively refines its dependency structure and transitions toward globally informed representations.

This observation reveals that CMAGIN increasingly relies on globally refined attention graphs rather than local sparse connections in later training stages, where global contextual relationships become more influential in predicting system degradation trends, while local correlations remain complementary. Such adaptive behavior reflects a dynamic balance between global contextual learning and local structural refinement, demonstrating that the fusion mechanism effectively calibrates the information flow between heterogeneous graph structures.

The overall trend of α evolution provides intuitive interpretability for how the model adjusts its reliance on different structural priors during learning.

5. CONCLUSION

This paper proposes a CMAGIN model for time series prediction. The model combines a dynamic adaptive graph construction module with local perception and global attention, and incorporates a channel-wise adaptive center enhancement mechanism. Specifically, the DAGC-LAGA module combines multi-view local dynamic graph construction with a global attention enhancement mechanism, enabling dynamic modeling of temporal graph structures and improving the capacity to capture both local and global dependencies. The CACE module enhances the

representation of key nodes across multiple channels by introducing channel-level node importance estimation and centrality-aware attention mechanisms, thereby strengthening the model's ability to capture and represent critical degradation information.

Experimental results on the four standard C-MAPSS sub-datasets demonstrate that the proposed model outperforms baseline methods in both RMSE and Score metrics. Specifically, on the FD001 subset, CMAGIN achieves 9.0% and 18.6% improvements in RMSE and Score compared to the suboptimal FCSTGNN model, exhibiting superior prediction accuracy and robustness. Through systematic comparative analysis, three major limitations of existing approaches are identified: while AConvLSTM excels at temporal modeling, it neglects spatial dependencies; static graph methods (GCN/HAGCN) fail to adapt to dynamic dependencies during degradation processes; and graph pooling methods (e.g., iPool) suffer from critical node information loss. To address these deficiencies, CMAGIN introduces innovative solutions: the DAGC-LAGA module employs a dynamic adaptive graph construction strategy that integrates local awareness with global attention to effectively capture node interactions and latent dependencies under non-stationary and variable operating conditions. Simultaneously, the CACE mechanism enhances the expressive power of central nodes across different feature channels for precise characterization of critical node information.

These findings collectively validate the effectiveness and advancement of the proposed method for complex temporal graph modeling tasks. Future research could further explore extending the application of CMAGIN to broader industrial scenarios to facilitate the practical deployment of intelligent operation and maintenance systems.

ACKNOWLEDGEMENT

This work was supported in part by the National Key Research and Development Program of China under Grant Nos. 2024YFB3310900 and 2022YFB3303803

REFERENCES

- Bacanin, N., Jovanovic, L., Zivkovic, M., Kandasamy, V., Antonijevec, M., Deveci, M., & Strumberger, I. (2023). Multivariate energy forecasting via metaheuristic tuned long-short term memory and gated recurrent unit neural networks. *Information Sciences*, 642. doi:10.1016/j.ins.2023.119122
- Bai, L., Yao, L., Li, C., Wang, X., & Wang, C. (2020). *Adaptive Graph Convolutional Recurrent Network for Traffic Forecasting*. Advances in Neural Information Processing Systems, 33, 17804-17815.
- Casolaro, A., Capone, V., Iannuzzo, G., & Camastra, F. (2023). Deep Learning for Time Series Forecasting: Advances and Open Problems. *Information*, 14(11), 598. doi:10.3390/info14110598

- Chen, L., Chen, D., Shang, Z., Wu, B., Zheng, C., Wen, B., & Zhang, W. (2023). Multi-scale adaptive graph neural network for multivariate time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 35(10), 10748-10761.
- Chen, W., Wang, Y., Du, C., Jia, Z., Liu, F., & Chen, R. (2023). *Balanced Spatial-Temporal Graph Structure Learning for Multivariate Time Series Forecasting: A Trade-off between Efficiency and Flexibility*. In Proceedings of the Asian Conference on Machine Learning, 189, 185-200.
- Ding, C., Sun, S., & Zhao, J. (2023). MST-GAT: A multimodal spatial-temporal graph attention network for time series anomaly detection. *Information Fusion*, 89, 527-536. doi:10.1016/j.inffus.2022.08.011
- Feng, C., Shao, L., Wang, J., Zhang, Y., & Wen, F. (2025). Short-term Load Forecasting of Distribution Transformer Supply Zones Based on Federated Model-Agnostic Meta Learning. *IEEE Transactions on Power Systems*, 40(1), 31-45. doi:10.1109/TPWRS.2024.3393017
- Gao, H., Liu, Y., & Ji, S. (2021). Topology-aware graph pooling networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12), 4512-4518.
- Gao, X., Dai, W., Li, C., Xiong, H., & Frossard, P. (2021). ipool—information-based pooling in hierarchical graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 33(9), 5032-5044.
- Guo, Q., He, Z., & Wang, Z. (2024). Monthly climate prediction using deep convolutional neural network and long short-term memory. *Scientific Reports*, 14(1), 17748. doi:10.1038/s41598-024-68906-6
- Huo, G., Zhang, Y., Wang, B., Gao, J., Hu, Y., & Yin, B. (2023). Hierarchical Spatio-Temporal Graph Convolutional Networks and Transformer Network for Traffic Flow Forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 24(4), 3855-3867. doi:10.1109/TITS.2023.3234512
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*: OTexts.
- Jin, M., Koh, H. Y., Wen, Q., Zambon, D., Alippi, C., Webb, G. I., King, I., Pan, S. (2024). A Survey on Graph Neural Networks for Time Series: Forecasting, Classification, Imputation, and Anomaly Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12), 10466-10485. doi:10.1109/tpami.2024.3443141
- Kim, J., Lee, H., Yu, S., Hwang, U., Jung, W., & Yoon, K. (2023). Hierarchical Joint Graph Learning and Multivariate Time Series Forecasting. *IEEE Access*, 11, 118386-118394. doi:10.1109/ACCESS.2023.3325041
- Kim, J., Oh, S., Kim, H., & Choi, W. (2023). Tutorial on time series prediction using 1D-CNN and BiLSTM: A case example of peak electricity demand and system marginal price prediction. *Engineering Applications of Artificial Intelligence*, 126, 106817.
- Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arxiv preprint arxiv:1609.02907*.
- Kumar, S., Raj, K. K., Cirrincione, M., Cirrincione, G., Franzitta, V., & Kumar, R. R. (2024). A Comprehensive Review of Remaining Useful Life Estimation Approaches for Rotating Machinery. *Energies*, 17(22), 5538.
- Lai, G., Chang, W. C., Yang, Y., & Liu, H. (2018). *Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks*. In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 95-104.
- Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical systems and signal processing*, 104, 799-834.
- Li, S., & Cai, H. (2024). Short-Term Power Load Forecasting Using a VMD-Crossformer Model. *Energies*, 17(11), 2773. doi:10.3390/en17112773
- Li, T., Zhao, Z., Sun, C., Yan, R., & Chen, X. (2021). Hierarchical attention graph convolutional network to fuse multi-sensor signals for remaining useful life prediction. *Reliability Engineering & System Safety*, 215, 107878.
- Liu, Y., Liu, Q., Zhang, J.-W., Feng, H., Wang, Z., Zhou, Z., & Chen, W. (2022). Multivariate Time-Series Forecasting with Temporal Polynomial Graph Neural Networks. *Advances in Neural Information Processing Systems*, 35, 19414-19426.
- Pareja, A., Domeniconi, G., Chen, J., Ma, T., Suzumura, T., Kanezashi, H., Kaler, T., Schardl, T., & Leiserson, C. (2020). EvolveGCN: Evolving Graph Convolutional Networks for Dynamic Graphs. In Proceedings of the 34th AAAI Conference on Artificial Intelligence, 34(04), 5363-5370.
- Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. In Proceedings of the 2008 International Conference on Prognostics and Health Management, 1-9.
- Shih, S. Y., Sun, F. K., & Lee, H. y. (2019). Temporal pattern attention for multivariate time series forecasting. *Machine Learning*, 108(8-9), 1421-1441. doi:10.1007/s10994-019-05815-0
- Su, H., & Lee, J. (2023). Machine learning approaches for diagnostics and prognostics of industrial systems using open source data from PHM data challenges: a review. *arXiv preprint arXiv:2312.16810*.
- Tarmanini, C., Sarma, N., Gezege, C., & Ozgonenel, O. (2023). Short term load forecasting based on ARIMA and ANN approaches. *Energy Reports*, 9, 550-557. doi:10.1016/j.egyr.2023.01.060
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is

- all you need. *Advances in neural information processing systems*, 30.
- Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. *stat, 1050*(20), 10-48550.
- Wang, H., Zhang, W., Yang, D., & Xiang, Y. (2023). Deep-Learning-Enabled Predictive Maintenance in Industrial Internet of Things: Methods, Applications, and Challenges. *IEEE Systems Journal*, 17(2), 2602-2615. doi:10.1109/JSYST.2022.3193200
- Wang, Y., Wu, M., Li, X., Xie, L., & Chen, Z. (2023). Multivariate time-series representation learning via hierarchical correlation pooling boosted graph neural network. *IEEE Transactions on Artificial Intelligence*, 5(1), 321-333.
- Wang, Y., Xu, Y., Yang, J., Chen, Z., Wu, M., Li, X., & Xie, L. (2023). Sensor alignment for multivariate time-series unsupervised domain adaptation. In *Proceedings of the AAAI conference on artificial intelligence*, 37(8), 10253-10261.
- Wang, Y., Xu, Y., Yang, J., Wu, M., Li, X., Xie, L., & Chen, Z. (2024). Fully-connected spatial-temporal graph for multivariate time-series data. In *Proceedings of the AAAI conference on artificial intelligence*, 38(14), 15715-15724.
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4-24. doi:10.1109/TNNLS.2020.2978386
- Wu, Z., Pan, S., Long, G., Jiang, J., & Zhang, C. (2019). Graph WaveNet for Deep Spatial-Temporal Graph Modeling. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 1907-1913.
- Xiao, Y., Yin, H., Zhang, Y., Qi, H., Zhang, Y., & Liu, Z. (2021). A dual - stage attention - based Conv - LSTM network for spatio - temporal correlation and multivariate time series prediction. *International Journal of Intelligent Systems*, 36(5), 2036-2057.
- Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2018). How powerful are graph neural networks?. *arxiv preprint arxiv:1810.00826*.
- Xu, N., Kosma, C., & Vazirgiannis, M. (2023). TimeGNN: Temporal Dynamic Graph Learning for Time Series Forecasting. In *International Conference on Complex Networks and Their Applications*, 87-99.
- Yu, B., Yin, H., & Zhu, Z. (2018). Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 3634-3640.
- Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., & Jin, R. (2022). FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting. In *Proceedings of the International Conference on Machine Learning*, 162, 27268-27286.