2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

MMGT-PD: A Multi-Modal Graph Transformer for Parkinson's Disease Stage Classification Using Clinical Omics and Whole Blood RNA Sequencing Data

Chengjie Ding[1(✉)] , Zeqi Xu[1] and Wei Zhang[1(✉)]

[1] School of Computer Science and Technology, Zhejiang Sci-Tech University, Zhejiang 310018, China
dingchengjie89@gmail.com, zhangweicse@zstu.edu.cn

Abstract. The assessment of both motor and non-motor functions in Parkinson's disease (PD) plays a crucial role in disease diagnosis and early intervention. In recent years, multi-modal deep learning methods have demonstrated excellent performance in identifying disease subtypes. However, previous studies have primarily focused on clinical and transcriptomic data, neglecting the information on gene associations. This paper proposes a multi-modal graph Transformer model named MMGT-PD, which integrates whole blood RNA sequencing data, gene co-expression networks, and Clinical omics data, combining modality-specific and consensus information to significantly enhance the accuracy of Parkinson's disease diagnosis. The model constructs a gene co-expression network using RNA sequencing data and designs an RNA sequencing encoder that combines Graph Attention Network (GAT) and Kolmogorov-Arnold Network (KAN) to extract RNA-specific representations. Additionally, the model introduces the Genegraph-Clinic Fusion (GCFusion) module to enhance the integration of multi-modal data by extracting shared information through inter-modal interactions. This paper conducts extensive comparative experiments on two well-known Parkinson's disease datasets, and the results show that the MMGT-PD method outperforms baseline models.

Keywords: Whole Blood RNA Sequencing Data, Gene Co-expression Networks, Clinical Omics Data, Graph Transformer.

1 Introduction

Parkinson's disease (PD) is a common neurodegenerative disorder that not only significantly impairs the motor functions of the elderly but also has a profound negative impact on their cognitive abilities [1, 2]. Currently, the medical community primarily relies on the MDS-UPDRS scale for the diagnosis and assessment of Parkinson's disease [3, 4]. This scale is the most commonly used clinical assessment tool and comprehensively covers the motor and cognitive conditions of PD patients. However, recent research advancements have brought new hope for the diagnosis and prognosis of Parkinson's disease. Specifically, studies based on blood RNA transcriptomics have

revealed gene expression changes associated with the progression of PD, which hold promise as novel biomarkers to provide robust support for early diagnosis and precise prognosis of the disease [5]. Furthermore, integrating blood RNA Sequencing (RNA-seq) data with MDS-UPDRS scale assessments can complement each other, potentially leading to the construction of more precise bio-AI models.

Previous studies have mainly focused on single-modal models [6, 7]. However, multi-modal learning now enables the integration of blood RNA-seq data with MDS-UPDRS scale assessments for more precise bio-AI models. Research shows that combining these data improves disease diagnosis compared to single-modal approaches [8, 9]. Blood RNA-seq reflects gene expression patterns, capturing microscopic disease mechanisms , while MDS-UPDRS clinical data summarizes macroscopic cognitive and motor symptoms. These two data types exhibit strong complementarity [7].

Previously, the analysis of blood RNA-seq data has predominantly relied on probabilistic statistics, with some applications of machine learning methods[6, 10]. However, these straightforward analyses are only capable of mining information at the level of differential gene expression patterns, lacking insights into the interactions between genes. In recent years, the advent of Graph Neural Network (GNN) has offered a promising approach to capture the interactions between genes [11-13]. By learning the interactions and information flow between nodes, GNN can extract structural features of genes, thereby revealing the underlying principles of biological systems. Nonetheless, these methods have primarily focused on data mining from multiple perspectives within a single information source, without integrating the rich clinical data. Moreover, despite significant progress in multi-modal learning, most current frameworks neglect the complex gene interactions in blood RNA-seq data and their potential links to clinical data when integrating multi-modal data. This oversight may hinder models from fully utilizing the structural information of gene expression networks, limiting the in-depth understanding of disease mechanisms and accurate prediction of disease progression. Based on the aforementioned considerations, we propose a novel MMGT-PD model framework to integrate multi-modal data and explore the rich structural information of patients. Specifically, this model incorporates Graph Attention Networks (GAT) and Kolmogorov–Arnold Networks (KAN) to jointly analyze whole blood RNA-seq data and gene co-expression networks, and then utilizes the GCFusion strategy to extract interactive information between modalities. The main contributions of this work are as follows:

- This work proposes MMGT-PD, a diagnostic solution for Parkinson's disease that integrates whole-blood RNA-seq data, gene co-expression networks, and clinical data. The method integrates modality-specific information and modality-consensus information to improve the diagnostic accuracy of Parkinson's disease (PD).
- A gene co-expression network is constructed using whole-blood RNA-seq data to explore potential associations between genes. An RNA-seq encoder integrating multi-level GAT and KAN is designed to extract RNA-specific representations.
- The GCFusion module is proposed to extract shared information between modalities by leveraging their inter-modal interactions.

2 Method

The workflow of the proposed method is illustrated in **Fig. 1**. Whole-blood RNA-seq data are first used to construct a gene graph via the Graph Construction module and then processed by the proposed RNA-seq Encoder to extract RNA-specific representations. Meanwhile, clinical data are processed by the Clinic Encoder to extract clinic-specific representations. After obtaining the modality-specific representations from both types of data, the GCFusion module is employed to extract modality-consensus representations. Subsequently, a joint representation is obtained by integrating the modality-specific and modality-consensus representations. Finally, a classification head is used to assess motor and non-motor impairments of Parkinson's disease based on the joint representation.

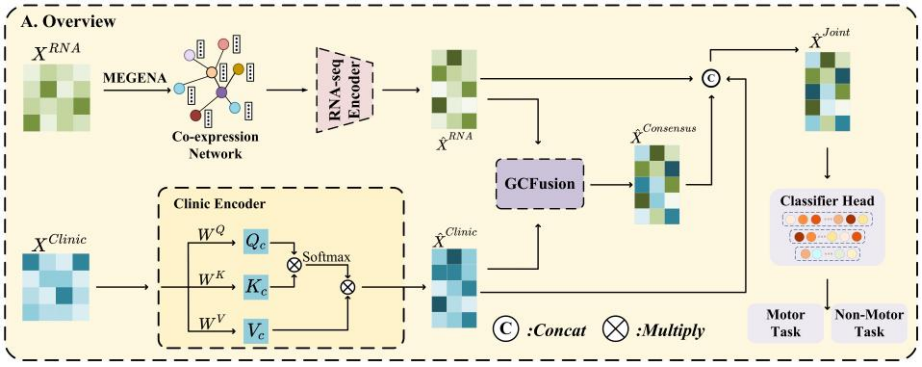


Fig. 1. The framework of the proposed MMGT-PD method.

2.1 Graph Construction

Let $X^{RNA} \in \mathbb{R}^{N\times M}$ denote the input of whole-blood RNA-seq data, where N is the number of samples and M is the number of gene features. Given the large number of genes in RNA-seq data, which contains certain levels of noise and redundancy, we preprocess the data using the scanpy [14] tool and select 500 highly variable genes to retain those with significant expression differences across different states. The resulting data is denoted as $X^{RNA_{500}} \in \mathbb{R}^{N\times 500}$. Subsequently, the associations between genes are analyzed using the MEGENA [15] tool to construct a gene co-expression matrix, thereby generating the gene co-expression network $G =< V, E >$. Where $V = \{v_1, v_2, \dots, v_{500}\}$ is a set containing 500 nodes, and E is the set of edges representing the relationships between nodes. Each edge $(v_i, v_j) \in E$ indicates the weight from node v_i to nodev_j. Furthermore, for the t-th batch B_t with input $X^{RNA_{500}}$ and the gene co-expression network $G =< V, E >$, a gene graph $G^0 = (X^{RNA_{500}}, E)$ can be constructed, where $X^{RNA_{500}} \in \mathbb{R}^{B_t\times 500\times d_0}$ is the feature matrix, $E \in \mathbb{R}^{500\times 500}$ is the edge matrix, and d_0 is the dimension of node features.

2.2 RNA-seq Encoder

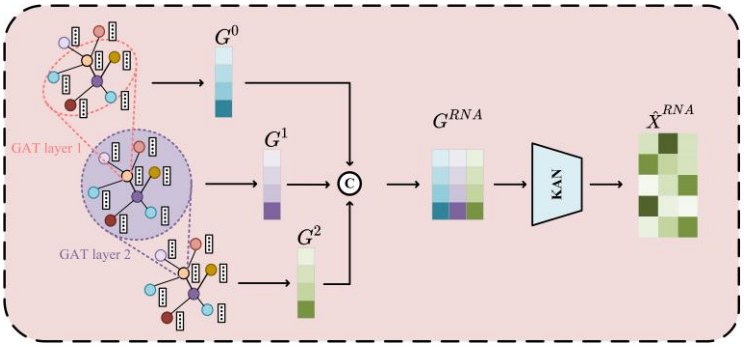


Fig. 2. The details of the RNA-seq encoder.

Previous studies have shown that interactions between genes influence the progression of Parkinson's disease, and graph neural network techniques are effective in capturing such interactions. Given that graph attention mechanisms can adaptively aggregate neighborhood information, we designed an RNA-seq encoder that integrates Graph Attention Network [16] and Kolmogorov–Arnold Networks (KAN) [17] for graph representation learning to explore gene interactions, as illustrated in **Fig. 2**. Taking G^0 as input, a GAT can be built by stacking multi-head attention layers. Each layer is defined as:

$$h'_u = ||_{k=1}^{K} \sigma\left(\sum_{v \in \mathcal{N}_v} \alpha_{uv}^k W^k h_v\right) \tag{1}$$

where, h_v denotes the input features of node v, $\mathcal{N}_v$ represents the first-order neighbors of node u, α_{uv}^k is the k-th normalized attention coefficient, W^k is the weight matrix of the k -th attention head, and $\sigma(\cdot)$ is a nonlinear activation function. The symbol $||$ indicates the concatenation of K attention heads. The attention coefficient α_{uv} is computed via the attention mechanism a:

$$\alpha_{uv} = \frac{exp(h_u, h_v)}{\sum_{v' \in \mathcal{N}_u} exp\left(a(h_u, h_{v'})\right)} \tag{2}$$

$$a(h_u, h_{v'}) = LeakReLU(W_a^T [h_u || h_v]) \tag{3}$$

The attention score between node u and its neighboring node v is computed using a single-layer feedforward neural network parameterized by the weight vector W_a^T, followed by the LeakyReLU activation function.

Furthermore, by applying a multi-head Graph Attention Network layer on G^0, a higher-level graph $G^1 = (X^1, E)$ is generated, where $X^1 \in \mathbb{R}^{B_t \times 500 \times d_1}$. Similarly, $G^2 = (X^2, E)$ is derived from G^1, where $X^2 \in \mathbb{R}^{B_t \times 500 \times d_2}$. To facilitate the concatenation of multi-level features, a fully connected layer is used to generate d_0-dimensional node features for G^1 and G^2. Subsequently, the graph embeddings from the three levels are concatenated, producing more enriched multi-level representations:

$$G^{RNA} = Concat(G^0, G^1, G^2) \tag{4}$$

where, $Concate(\cdot)$ denotes the concatenation operation. Subsequently, the d-dimensional vector $G^{RNA} = (g^1, g^2, .., g^d)$ is fed into a three-layer KAN network for further feature extraction.

$$\hat{X}^{RNA} = \sum_{l=1}^{L}\left(\sum_{j=1}^{J}\phi_{l,j}^{(3)}\left(\sum_{i=1}^{d}\phi_{j,i}^{(2)}\left(\phi_i^{(1)}(g^i)\right)\right)\right) \tag{5}$$

where, J denotes the number of nodes in the second layer, L denotes the number of nodes in the third layer, and ϕ represents the univariate functions in each layer. Through this RNA-seq Encoder, the RNA-seq-specific representation $\hat{X}^{RNA} \in \mathbb{R}^{N\times L}$ is obtained.

2.3 Clinic Encoder

Let $X^{Clinic} \in \mathbb{R}^{N\times C}$ denote the input of the clinical data, where N is the number of samples and C is the number of clinical features. The proposed Clinic Encoder, powered by a self-attention mechanism [18], aims to adaptively capture the correlations between clinical data features and assign more weight to important information. Specifically, three weight matrices are defined: the query weight matrix $W_Q \in \mathbb{R}^{C\times C}$, the key weight matrix $W_K \in \mathbb{R}^{C\times C}$, and the value weight matrix $W_V \in \mathbb{R}^{C\times C}$. The query vector Q_c, key vector K_c, and value vector V_c of the modality features are computed and can be defined as follows:

$$Q_c = W_Q X^{Clinic} \tag{6}$$

$$K_c = W_K X^{Clinic} \tag{7}$$

$$V_c = W_V X^{Clinic} \tag{8}$$

Therefore, the self-attention (SA) mechanism is employed to compute the intra-modal associations, and the enhanced feature $\hat{X}^{Clinic} \in \mathbb{R}^{N\times L}$ for the clinical data is obtained through a linear transformation. The computation of $\hat{X}^{Clinic}$ is as follows:

$$\hat{X}^{Clinic} = Linear\left(Softmax\left(\frac{Q_c K_c^T}{\sqrt{d_k}}\right)V_c\right) \tag{9}$$

where d_k is a hyperparameter.

2.4 Genegraph-Clinic Fusion(GCFusion)

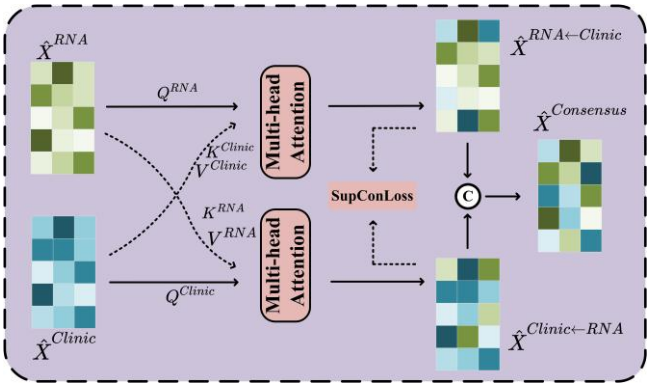


Fig. 3. The details of the GCFusion encoder.

A fusion strategy based on cross-attention mechanisms, GCFusion, is proposed to integrate RNA-specific representations and clinic-specific representations, as illustrated in **Fig. 3**. This strategy computes inter-modal relationships to generate a consensus representation, which is further combined with single-modal information to obtain a joint representation. By computing attention weights among Query, Key, and Value, two cross-modal representations are obtained:

$$\hat{X}^{RNA\leftarrow Clinic} = Softmax\left(\frac{Q_{RNA}K_{Clinic}^{T}}{\sqrt{d_k}}\right)V_{Clinic} \tag{10}$$

$$\hat{X}^{Clinic\leftarrow RNA} = Softmax\left(\frac{Q_{Clinic}K_{RNA}^{T}}{\sqrt{d_k}}\right)V_{RNA} \tag{11}$$

where, Q_{RNA}, K_{RNA}, and V_{RNA} represent the query matrix, key matrix, and value matrix of the RNA-seq modality, respectively; Q_{Clinic}, K_{Clinic}, and V_{Clinic} represent the query matrix, key matrix, and value matrix of the clinical modality, respectively. d_k denotes the dimensionality of the key vectors, used to scale the dot-product attention scores. $Softmax(\cdot)$ is the normalization function applied to compute the attention weights.

To fuse the consensus information from the dual modalities, the two cross-modal representations are concatenated to obtain the consensus representation.

$$X^{Consensus} = Concat\left(\hat{X}^{RNA\leftarrow Clinic}, \hat{X}^{Clinic\leftarrow RNA}\right) \tag{12}$$

The consensus representation $X^{Consensus}$ incorporates bidirectional interaction information, encompassing both RNA-seq to clinical data and clinical data to RNA-seq interactions. To further enrich the feature information, this consensus representation is integrated with single-modal information to obtain a joint representation.

$$X^{Joint} = Concat\left(X^{Consensus}, \hat{X}^{RNA}, \hat{X}^{Clinic}\right) \tag{13}$$

To further evaluate whether the modality interaction features possess consensus properties across modalities, we leverage label information for each data sample and employ a supervised contrastive loss to provide effective guidance. Specifically, if data samples from different modalities share the same label, we expect their modality consensus features to be as close as possible in the embedding space. For a set of N labeled data pairs $\{(X_j^{RNA}, X_j^{Clinic}, y_j)\}_{j=1,2,\ldots,N}$, we first map the features of each modality through their respective feature extractors and an GCFusion module into a shared embedding space, obtaining the feature representations f^{RNA} and f^{Clinic}. Subsequently, we compute the similarity between these features and optimize them using the supervised contrastive loss function.

$$\mathcal{L}_{con} = -\frac{1}{N}\sum_{j=1}^{N} \log\left(\frac{\sum_{k\in P(j)} exp\left(sim\left(f_j^{RNA}, f_k^{Clinic}\right)\right)/\tau}{\sum_{k\neq j} exp\left(sim\left(f_j^{RNA}, f_k^{Clinic}\right)\right)/\tau}\right) \tag{14}$$

where, $P(j)$ denotes the set of indices of all positive samples in the batch that share the same label as sample j. τ is a scalar temperature parameter that controls the range of the similarity scores. $sim\left(f_j^{RNA}, f_k^{Clinic}\right)$ represents the similarity measure between the features of sample j and sample k, which is typically computed using the normalized dot product.

$$sim\left(f_j^{RNA}, f_k^{Clinic}\right) = \frac{f_j^{RNA} \cdot f_k^{Clinic}}{\left\|f_j^{RNA}\right\|\left\|f_k^{Clinic}\right\|} \tag{15}$$

The final loss is obtained as the weighted sum of the previously defined losses:

$$\mathcal{L} = \mathcal{L}_{cls} + \alpha_{con}\mathcal{L}_{con} \tag{16}$$

where $\mathcal{L}_{cls}$ is the cross-entropy loss for classification, and where α_{con} are hyperparameters that control the relative importance of the contrastive learning.

Experiments

2.5 Dataset description and task settings

The two datasets were both downloaded from the Accelerating Medicines Partnership Parkinson's Disease (AMP-PD). The Parkinson's Progression Marker Initiative (PPMI) dataset contains 4,397 samples, while the Parkinson's Disease Biomarker Program (PDBP) dataset contains 3,345 samples. The clinical data used are from the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) Part I, II, and III, where Part I includes information on cognitive assessments, and Parts II and III provide evaluations of motor functions. Based on this, two tasks were established: the motor disability assessment task and the non-motor disability assessment task. The labels for the non-motor disability assessment task are derived from the cognitive impairment score, which includes five stages; the labels for the motor disability

assessment task are derived from the Hoehn and Yahr staging criteria, which includes six stages.

(a) PPMI (b) PDBP

Fig. 4. The gene networks constructed from the whole-blood RNA-seq data of the PPMI dataset and the PDBP dataset are demonstrated.

To further illustrate the distribution of the two datasets, we visualized the constructed gene graph, as shown in **Fig. 4**. It can be observed that there are diverse connections among genes, which may potentially offer fundamental explanations for the pathogenesis of Parkinson's disease.

2.6 Implementation Details

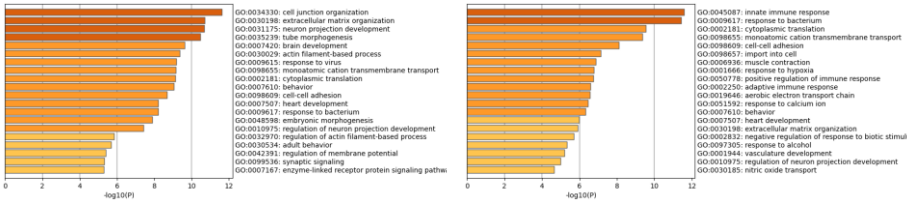
In the work of staging diagnosis for Parkinson's disease, we conducted an in-depth analysis of motor and non-motor assessments based on whole blood RNA-seq data and clinical data. The model training process was carried out on a single Nvidia GeForce RTX 4090 GPU, utilizing the Adam optimizer for optimization. Throughout the training process, the batch size was fixed at 32.

Four statistical metrics are employed to evaluate model performance: Accuracy, F1 Score, Recall, and Precision. All experiments were performed in quintuplicate, and the results were averaged. Comparative analysis of different thresholds for highly variable gene selection (200, 500, 800, and 1000 genes) demonstrated that the 500-gene threshold yielded the best performance.

2.7 Key Biological Processes

Functional enrichment analysis of the significantly associated GO terms revealed that multiple biological processes related to neural development, immune response, cardiovascular regulation, and metabolic function were consistently enriched across both

datasets, as shown in **Fig. 5**, indicating the pivotal roles of the identified key genes within multi-scale biological networks.



(a) PPMI (b) PDBP

Fig. 5. GO biological processes from whole-blood RNA-seq data in the PPMI and PDBP datasets.

Specifically, several neural system-related pathways—such as regulation of neuron projection development, synaptic signaling, and behavioral processes—were significantly activated, suggesting that these genes may contribute to neuronal morphogenesis, synaptic plasticity, and maintenance of neural function, with potential implications in the pathogenesis of neurological disorders.

Concurrently, enrichment of immune-related pathways, including both innate and adaptive immune responses and the regulation of inflammatory signals, supports the involvement of neuroinflammation as a key driver in disease progression. Furthermore, the enrichment of pathways associated with extracellular matrix organization, cell-cell adhesion, and cardiovascular system development points to a potential link between these genes and blood-brain barrier integrity, tissue microenvironment homeostasis, and cerebral blood flow regulation.

Notably, metabolic pathways related to mitochondrial function, oxidative stress, and hypoxia response were also enriched, implying that the identified genes may influence neuronal energy metabolism and cellular viability.

Collectively, these findings highlight the coordinated involvement of the selected genes in neural, immune, and metabolic systems, providing mechanistic insights and a solid foundation for exploring their clinical relevance in neurodegenerative diseases.

2.8 Comparison Methods

We compare our model with the following six baselines. Transformer [19] and Acmix [20] are two single-modality methods that utilize the Transformer architecture and a combination of convolution and attention mechanisms, respectively, to process RNA or clinical data. MLA-GNN [13] and GREMI [12] are graph neural network-based methods for the joint analysis of RNA and gene co-expression networks. Additionally, MADDI [21] and SimMMDG [22] integrate blood RNA-seq and clinical data from two modalities, with MADDI capturing inter-modal interactions through cross-modal attention mechanisms and SimMMDG separating modality-specific and shared

features via contrastive learning. Ultimately, the proposed MMTG-PD method further integrates RNA, gene co-expression networks, and clinical data to achieve a more comprehensive multi-modal analysis.

2.9 Comparison and Result Analysis

Table 1. Performance Comparison for Motor Tasks.

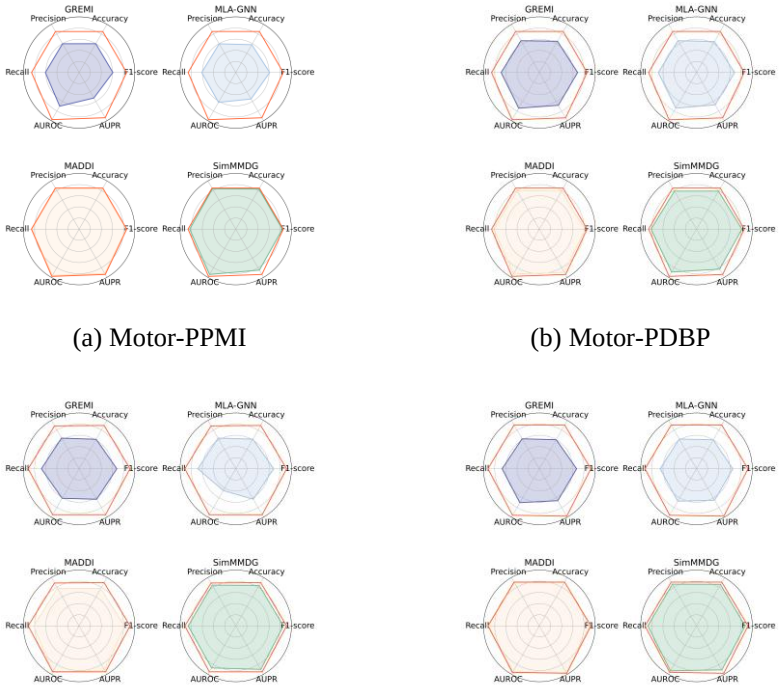
Model	Motor-PPMI				Motor-PDBP			
	Acc	**F1-score**	**Recall**	**Precision**	**Acc**	**F1-score**	**Recall**	**Precision**
Transformer -Clinic[19]	0.836 (±0.013)	0.830 (±0.018)	0.836 (±0.013)	0.830 (±0.018)	0.808 (±0.017)	0.788 (±0.021)	0.801 (±0.022)	0.778 (±0.012)
Transformer-RNA	0.603 (±0.003)	0.591 (±0.007)	0.613 (±0.003)	0.589 (±0.002)	0.641 (±0.018)	0.590 (±0.019)	0.643 (±0.017)	0.579 (±0.022)
ACmix-Clinic[20]	0.815 (±0.006)	0.795 (±0.011)	0.815 (±0.006)	0.795 (±0.013)	0.800 (±0.012)	0.774 (±0.023)	0.742 (±0.006)	0.770 (±0.033)
ACmix-RNA	0.574 (±0.003)	0.556 (±0.004)	0.575 (±0.003)	0.546 (±0.003)	0.646 (±0.004)	0.636 (±0.001)	0.646 (±0.004)	0.625 (±0.001)
GREMI[12]	0.608 (±0.007)	0.598 (±0.004)	0.608 (±0.007)	0.599 (±0.006)	0.675 (±0.009)	0.660 (±0.006)	0.685 (±0.009)	0.648 (±0.008)
MLA-GNN[13]	0.611 (±0.011)	0.596 (±0.006)	0.612 (±0.008)	0.578 (±0.003)	0.684 (±0.008)	0.659 (±0.006)	0.684 (±0.008)	0.643 (±0.008)
MADDI[21]	0.846 (±0.007)	0.841 (±0.009)	0.846 (±0.007)	0.841 (±0.009)	0.816 (±0.011)	0.798 (±0.003)	0.817 (±0.011)	0.789 (±0.009)
SimMMDG[22]	0.838 (±0.009)	0.837 (±0.009)	0.827 (±0.008)	0.832 (±0.011)	0.825 (±0.008)	0.789 (±0.013)	0.816 (±0.008)	0.789 (±0.019)
MMGT-PD (Ours)	**0.855** (±0.012)	**0.852** (±0.013)	**0.857** (±0.012)	**0.852** (±0.013)	**0.842** (±0.006)	**0.809** (±0.019)	**0.836** (±0.006)	**0.790** (±0.021)

Table 2. Performance Comparison for Non-Motor Tasks.

Model	Non-Motor-PPMI				Non-Motor-PDBP			
	Acc	**F1-score**	**Recall**	**Precision**	**Acc**	**F1-score**	**Recall**	**Precision**
Transformer -Clinic[19]	0.728 (±0.007)	0.665 (±0.015)	0.729 (±0.007)	0.661 (±0.028)	0.815 (±0.025)	0.794 (±0.029)	0.816 (±0.027)	0.787 (±0.003)
Transformer-RNA	0.660 (±0.025)	0.630 (±0.030)	0.663 (±0.021)	0.615 (±0.033)	0.654 (±0.001)	0.583 (±0.006)	0.655 (±0.001)	0.552 (±0.012)
ACmix-Clinic[20]	0.740 (±0.002)	0.677 (±0.004)	0.740 (±0.002)	0.666 (±0.004)	0.869 (±0.008)	0.710 (±0.002)	0.869 (±0.008)	0.774 (±0.004)
ACmix-RNA	0.677 (±0.012)	0.628 (±0.009)	0.679 (±0.012)	0.617 (±0.001)	0.626 (±0.004)	0.600 (±0.002)	0.625 (±0.002)	0.584 (±0.002)
GREMI[12]	0.678 (±0.006)	0.634 (±0.007)	0.678 (±0.006)	0.620 (±0.009)	0.667 (±0.009)	0.620 (±0.012)	0.667 (±0.009)	0.602 (±0.016)
MLA-GNN[13]	0.682 (±0.002)	0.635 (±0.008)	0.682 (±0.002)	0.613 (±0.006)	0.658 (±0.003)	0.616 (±0.009)	0.659 (±0.003)	0.599 (±0.012)
MADDI[21]	0.814 (±0.017)	0.779 (±0.026)	0.814 (±0.017)	0.774 (±0.024)	0.908 (±0.021)	0.898 (±0.003)	0.908 (±0.021)	0.903 (±0.037)
SimMMDG[22]	0.855 (±0.024)	0.843 (±0.025)	0.854 (±0.025)	0.839 (±0.024)	0.871 (±0.009)	0.863 (±0.010)	0.871 (±0.009)	0.860 (±0.011)
MMGT-PD (Ours)	**0.926** (±0.037)	**0.911** (±0.042)	**0.928** (±0.037)	**0.914** (±0.041)	**0.923** (±0.025)	**0.906** (±0.028)	**0.921** (±0.021)	**0.905** (±0.024)

As shown in **Table 1**, **Table 2** and **Fig. 6**, in the classification and diagnostic tasks of motor and non-motor dysfunctions, the MMGT-PD method significantly outperforms other fusion methods across various evaluation metrics. To validate the effectiveness of the multimodal learning approach, baseline tests were conducted on RNA-seq data and clinical diagnostic data using two single-modal models. The experimental results indicate that while single-modal data alone can achieve basic classification of

motor and non-motor dysfunctions in Parkinson's disease, the joint representations constructed through multimodal fusion schemes (MADDI and SimMMDG) significantly enhance classification accuracy. This finding further confirms the substantial complementarity between whole-blood transcriptomic data and clinical data. Additionally, it was observed that graph neural network models combining RNA-seq data with gene co-expression networks (GREMI and MLA-GNN) capture deeper-level information more effectively compared to deep learning models using only RNA-seq data, highlighting the significant value of integrating gene association information with RNA-seq data.



(a) Motor-PPMI (b) Motor-PDBP

(c) Non-motor-PPMI (d) Non-motor-PDBP

Fig. 6. In (a) and (b), the PPMI dataset and the PDBP dataset are respectively used to evaluate the baseline methods for assessing patient motor disability, where the red solid lines in each radar subplot represent the MMGT-PD method. In (c) and (d), the PPMI dataset and the PDBP dataset are respectively used to compare the baseline methods for assessing non-motor disabilities in patients.

The MMGT-PD model, through its design of multi-layer graph learning and the fusion of modality-specific representations with modality-consensus representations, significantly improves diagnostic accuracy and reliability. Specifically, compared to the two existing graph representation learning methods, the MMGT-PD method effectively

captures inter-gene relationships using graph attention mechanisms and the KAN model, resulting in more efficient RNA-seq representations. Furthermore, in comparison to the two existing multimodal deep learning methods, the fusion strategy designed in the MMGT-PD method captures more comprehensive and effective joint representations, further enhancing the model's performance.

2.10 Ablation study

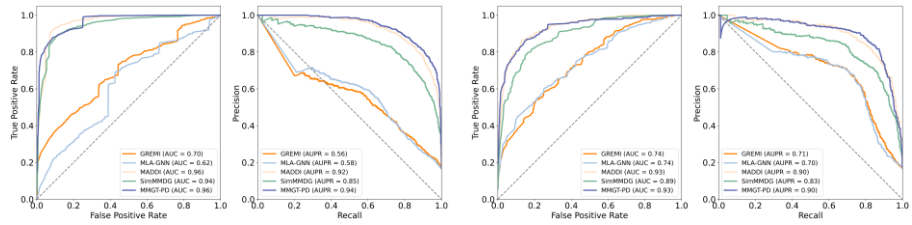
Ablation experiments on two datasets across four subtasks validated the effectiveness of the proposed modules. Specifically, we individually removed the RNA-seq Encoder (GAT-KAN) module and the GCFusion module, with the results shown in **Table 3** and **Fig. 7**. The results indicate that each module, when used alone, can enhance the performance of the MMGT-PD baseline model to some extent. However, the optimal performance is achieved only when both modules are integrated into the model.

Table 3. Ablation study of the RNA-seq Encoder and GCFusion module in the MMGT-PD method.

		GAT-KAN	GCFusion	Acc	F1-score	Recall	Precision
Motor Task	PPMI			0.848	0.838	0.847	0.839
		✓		0.848	0.845	0.851	0.848
			✓	0.831	0.825	0.838	0.826
		✓	✓	**0.855**	**0.852**	**0.857**	**0.852**
	PDBP			0.819	0.782	0.815	0.753
		✓		0.832	0.800	0.835	**0.832**
			✓	0.819	0.776	0.822	0.738
		✓	✓	**0.842**	**0.809**	**0.836**	0.790
Non-motor Task	PPMI			0.900	0.928	0.902	0.901
		✓		0.904	0.900	0.903	0.881
			✓	0.918	0.910	0.917	0.904
		✓	✓	**0.926**	**0.911**	**0.928**	**0.914**
	PDBP			0.896	0.859	0.887	0.835
		✓		0.910	0.911	0.915	0.899
			✓	0.911	0.883	0.923	0.894
		✓	✓	**0.923**	**0.906**	**0.921**	**0.905**

Effectiveness of GAT-KAN. In the experimental design of this work, we focused on the characteristics of whole-blood RNA-seq data, particularly the interactions between genes and their impact on the progression of Parkinson's disease. To verify the effectiveness of the gene interactions we extracted, we designed an ablation experiment for the RNA-seq Encoder (GAT-KAN) module. As shown in **Table 3**, the results indicate that compared with the traditional combination of graph neural networks and multilayer perceptrons, the GAT-KAN module can more effectively capture the associations between genes, thereby significantly improving the classification performance of Parkinson's disease. Specifically, the multi-head attention mechanism in GAT dynamically learns the regulatory weights between genes, effectively addressing the signal

dilution issue commonly caused by traditional GNN pooling operations. At the same time, the B-spline basis functions in KAN offer strong nonlinear adaptability, with piecewise polynomial fitting capabilities that accurately capture high-order interactions among complex biomarkers. More importantly, the integration of GAT and KAN establishes an attention-guided strategy for representation refinement: the topology-aware features produced by GAT are further transformed by KAN's flexible basis functions to enhance pathway-specific feature representations. A gated feature fusion module is then used to jointly preserve local neighborhood structures and global functional associations. This architecture demonstrates strong effectiveness in capturing multi-scale biological network features across four tasks on the PPMI and PDBP datasets.



(a) Motor-PPMI (b) Motor-PDBP

(c) Non-motor-PPMI (d) Non-motor-PDBP

Fig. 7. In (a) and (b), the PPMI dataset and the PDBP dataset are respectively used to evaluate the baseline methods for assessing patient motor disability. In (c) and (d), the PPMI dataset and the PDBP dataset are respectively used to compare the baseline methods for assessing non-motor disabilities in patients.

Effectiveness of GCFusion. In the implementation of multi-modal approaches, designing the interaction between two types of modality information is crucial. This is because simple fusion strategies may struggle to effectively balance two types of information: one is the modality-specific information that is closely related to the disease but only exists in a single modality, and the other is the consensus information shared between modalities. As shown in **Table 3**, experimental results indicate that compared with simple concatenation-based fusion strategies, the GCFusion module we designed can more fully integrate modality-specific and consensus information, thereby significantly enhancing the classification performance of the model. Specifically, GCFusion leverages a cross-modal attention mechanism to establish explicit interaction pathways

between different modalities, enabling the model to dynamically perceive and select the most discriminative modality-specific information during the fusion process. This mechanism not only enhances the collaborative representation capability across modalities but also strengthens the focus on key features, effectively addressing the insufficient cross-modal interaction problem inherent in simple concatenation strategies.

3 Conclusion and Future Work

In this work, we used multi-modal deep learning to integrate blood RNA-seq data and clinical assessments. The results highlight the importance of gene interactions and clinical data for understanding disease mechanisms and improving diagnosis. The MMTG-PD framework, integrating RNA, gene co-expression networks, and clinical data, underscores the benefits of a comprehensive approach. This strategy provides deeper disease insights and more effective diagnostic tools, leading to more accurate predictions and improved patient outcomes. However, our work has limitations. The complexity of biological systems and the heterogeneity of clinical data pose challenges in fully capturing the nuances of disease mechanisms. Future research should address these challenges by incorporating more data types such as imaging and proteomics data, and by developing more complex models.

References

1. Fiorini, M.R., Dilliott, A.A., Thomas, R.A., Farhan, S.M.: Transcriptomics of human brain tissue in parkinson's disease: a comparison of bulk and single-cell rna sequencing. Molecular Neurobiology 61(11), 8996–9015 (2024)
2. Hindle, J.V.: Ageing, neurodegeneration and parkinson's disease. Age and ageing 39(2), 156–161 (2010)
3. Postuma, R.B., Berg, D., Stern, M., Poewe, W., Olanow, C.W., Oertel, W., Obeso, J., Marek, K., Litvan, I., Lang, A.E., et al.: Mds clinical diagnostic criteria for parkinson's disease. Movement disorders 30(12), 1591–1601 (2015)
4. Martínez-Martín, P., Rodríguez-Blázquez, C., Alvarez, M., Arakaki, T., Arillo, V.C., Chaná, P., Fernández, W., Garretto, N., Martínez-Castrillo, J.C., Rodríguez-Violante, M., et al.: Parkinson's disease severity levels and mds-unified parkinson's disease rating scale. Parkinsonism & related disorders 21(1), 50–54 (2015)
5. Craig, D.W., Hutchins, E., Violich, I., Alsop, E., Gibbs, J.R., Levy, S., Robison, M., Prasad, N., Foroud, T., Crawford, K.L., et al.: Rna sequencing of whole blood reveals early alterations in immune cells and gene expression in parkinson's disease. Nature Aging 1(8), 734–747 (2021)
6. Su, C., Tong, J., Wang, F.: Mining genetic and transcriptomic data using machine learning approaches in parkinson's disease. npj Parkinson's Disease 6(1), 24 (2020)
7. Regnault, A., Boroojerdi, B., Meunier, J., Bani, M., Morel, T., Cano, S.: Does the mds-updrs provide the precision to assess progression in early parkinson's disease? learnings from the parkinson's progression marker initiative cohort. Journal of neurology 266, 1927–1936 (2019)

8. Wekesa, J.S., Kimwele, M.: A review of multi-omics data integration through deep learning approaches for disease diagnosis, prognosis, and treatment. Frontiers in genetics 14, 1199087 (2023)
9. Irmady, K., Hale, C.R., Qadri, R., Fak, J., Simelane, S., Carroll, T., Przedborski, S., Darnell, R.B.: Blood transcriptomic signatures associated with molecular changes in the brain and clinical outcomes in parkinson's disease. Nature Communications 14(1), 3956 (2023)
10. Pantaleo, E., Monaco, A., Amoroso, N., Lombardi, A., Bellantuono, L., Urso, D., Lo Giudice, C., Picardi, E., Tafuri, B., Nigro, S., et al.: A machine learning approach to parkinson's disease blood transcriptomics. Genes 13(5), 727 (2022)
11. Zhang, X.M., Liang, L., Liu, L., Tang, M.J.: Graph neural networks and their current applications in bioinformatics. Frontiers in genetics 12, 690049 (2021)
12. Liang, H., Luo, H., Sang, Z., Jia, M., Jiang, X., Wang, Z., Cong, S., Yao, X.: Gremi: an explainable multi-omics integration framework for enhanced disease prediction and module identification. IEEE Journal of Biomedical and Health Informatics (2024)
13. Lu, C.Y., Liu, Z., Arif, M., Alam, T., Qiu, W.R.: Integration of gene expression and dna methylation data using mla-gnn for liver cancer biomarker mining. Frontiers in Genetics 15, 1513938 (2024)
14. Wolf, F.A., Angerer, P., Theis, F.J.: Scanpy: large-scale single-cell gene expression data analysis. Genome biology 19, 1–5 (2018)
15. Song, W.M., Zhang, B.: Multiscale embedded gene co-expression network analysis. PLoS computational biology 11(11), e1004574 (2015)
16. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
17. Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T.Y., Tegmark, M.: Kan: Kolmogorov-arnold networks. arXiv preprint arXiv:2404.19756 (2024)
18. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems 30 (2017)
19. Han, K., Xiao, A., Wu, E., Guo, J., Xu, C., Wang, Y.: Transformer in transformer. Advances in neural information processing systems 34, 15908–15919 (2021)
20. Pan, X., Ge, C., Lu, R., Song, S., Chen, G., Huang, Z., Huang, G.: On the integration of self-attention and convolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 815–825 (2022)
21. Golovanevsky, M., Eickhoff, C., Singh, R.: Multimodal attention-based deep learning for alzheimer's disease diagnosis. Journal of the American Medical Informatics Association 29(12), 2014–2022 (2022)
22. Dong, H., Nejjar, I., Sun, H., Chatzi, E., Fink, O.: Simmmdg: A simple and effective framework for multi-modal domain generalization. Advances in Neural Information Processing Systems 36, 78674–78695 (2023)