

用于自闭症诊断的多尺度融合和跨模态交互图卷积网络

许傲雪, 谢勤岚*

(中南民族大学 生物医学工程学院, 湖北 武汉 430074)

摘要 图神经网络在处理非欧数据方面的优势使得功能网络分析在脑疾病诊断中取得了显著进展,但诊断过程中仍存在多模态信息融合不充分、脑网络多尺度特征捕捉不足的问题.提出了端到端的多尺度融合与跨模态交互(MFCMI)图卷积网络,利用成像数据构建个体图,通过密集-稀疏双分支融合模块,在稀疏分支中利用自注意力图池化筛选关键脑区以捕捉局部精细拓扑,在密集分支中利用节点聚类聚合全脑信息以保留全局关联模式,实现脑网络从局部到全局的多尺度特征提取.同时基于多头交叉注意力和门控融合的跨模态交互模块,使影像特征和非影像特征相互查询检索互补信息,实现双向深度交互,进而在群体图上进行分类.在ABIDE数据集上进行10折交叉验证,结果显示模型准确率达88.40%、AUC为96.35%,优于现有先进方法,且能识别出关键生物标志物脑区,有效提升了ASD诊断的准确性与可解释性,为临床辅助诊断提供了可靠支撑.

关键词 自闭症;图卷积网络;多尺度融合;跨模态交互;脑网络分析

中图分类号 TP391.5 文献标志码 A 文章编号 1672-4321(2026)06-0855-10

doi:10.20056/j.cnki.ZNMDZK.20260614

Multi-scale fusion and cross modal interaction graph convolutional network for Autism Diagnosis

Xu Aoxue, Xie Qinlan*

(College of Biomedical Engineering, South-Central Minzu University, Wuhan 430074, Hubei China)

Abstract The advantages of graph neural networks in processing non Euclidean data have led to significant progress in functional network analysis for brain disease diagnosis. However, there are still problems in the diagnosis process such as insufficient multimodal information fusion and insufficient capture of multi-scale features by brain networks. An end-to-end multi-scale fusion and cross modal interaction (MFCMI) graph convolutional network is proposed, which constructs individual graphs using imaging data. Through a dense sparse dual branch fusion module, self attention graph pooling is used to screen key brain regions in sparse branches to capture local fine topology. In dense branches, node clustering is used to aggregate global brain information to preserve global correlation patterns, achieving multi-scale feature extraction of brain networks from local to global. At the same time, a cross modal interaction module based on multi head cross attention and gating fusion enables image features and non image features to query and retrieve complementary information from each other, achieving bidirectional deep interaction, and then classifying on the group map. A 10 fold cross validation was conducted on the ABIDE dataset, and the results showed that the model had an accuracy of 88.40% and an AUC of 96.35%, which are superior to existing advanced methods. It was also able to identify key biomarkers in brain regions, effectively improving the accuracy and interpretability of ASD diagnosis and providing reliable support for clinical auxiliary diagnosis.

收稿日期 2026-05-10

* 通信作者 谢勤岚(1968-),男,教授,研究方向:医学图像处理,E-mail:xieqinlan@126.com

基金项目 湖北省自然科学基金资助项目(2016CFB489);湖北本科高校省级教学改革研究资助项目(2025171)

Keywords Autism Spectrum Disorder; graph convolutional network; multi-scale fusion; cross modal interaction; brain network analysis

自闭症谱系障碍 (Autism Spectrum Disorder, ASD) 是一种神经发育障碍性疾病, 具有诊断周期长, 行为评估的诊断主观性强, 缺乏客观可靠的生物标志物的特性^[1], 使得对自闭症进行计算机辅助诊断更有必要. 自闭症患者在脑连接方面存在显著异常已经被证明, 为自闭症的诊断提供了新的可能性^[2]. 当前静息态功能磁共振成像 (Resting state functional magnetic resonance imaging, rs-fMRI) 已广泛应用于脑网络分析, 基于 rs-fMRI 数据构建的功能连接网络可有效揭示脑部区域间的协同活动模式. 相比于机器学习方法^[3-5], 深度学习更擅长从大脑网络中提取更具区分性的特征. 图神经网络 (Graph Neural Network, GNN) 作为适用于图结构数据的深度学习方法, 尤其是图卷积网络 (Graph Convolutional Network, GCN), 在非欧几里得数据处理中展现出巨大潜力, 已在脑部疾病诊断领域取得重要进展^[6].

现有基于图神经网络的研究中图构建方法主要分为两大类: 个体图方法和群体图方法. 个体图方法将每个受试者的脑区视为节点, 功能连接视为边, 能够识别与疾病相关的局部脑区和生物标志物. Li 等人开发了可解释的 GNN 用于分析 rs-fMRI 结果和识别生物标志物^[7]. Wen 等人构建了一个基于先验结构知识的多视图 GNN 识别自闭症谱系障碍患者, 并引入了一个图结构学习方案来识别关键功能连接组^[8]. 但使用个体图的诊断方法忽略了受试者之间的潜在关联. 因此有 Parisot 等人首次将图卷积网络应用于群体图模型, 通过整合影像与非影像数据实现对自闭症的诊断^[9]. 群体图方法将每个受试者视为节点, 基于表型数据构建节点间的连接关系, 虽然在分类准确率上表现优异, 却难以解释脑区层面的病理机制, 临床可解释性较差^[10]. Zhang 等人提出的局部到全局图神经网络 (LG-GNN) 尝试结合两种方法的优势, 通过自注意力池化机制保留重要脑区特征^[11]. Zheng 等人利用模态之间的相关性, 提出了一种基于多模态数据的疾病预测的端到端多模态图学习框架 (MMGL)^[12]. 虽然上述方法尝试引入多尺度和多模态信息融合思想并取得进展, 但是现有多尺度方法多采用层级下采样、固定感受野卷积或单一的池化策略, 在放大全局网络表征的同时, 会不可避免地削弱局部脑区连接的细节信息, 且现有方法常使用固定比例的全局池化或无差

别降维, 将局部脑区节点特征向高层簇节点聚合, 这一过程会抹平脑区间的差异化连接模式, 导致难以捕捉局部与全局信息的互补关系^[13]. 此外, 现有方法在多模态信息融合方面多采用简单的特征拼接或加权求和的方法, 忽略了影像特征与人口统计学特征之间的双向引导关系, 且缺乏动态模态权重分配机制, 导致表型信息对诊断的增益受限未能充分挖掘. 这些局限性导致现有方法在小样本、多中心神经影像数据上泛化能力不足, 难以满足临床应用需求.

针对上述问题, 本文提出了一种基于多尺度特征融合和跨模态特征交互的图神经网络架构 (Multi-scale Fusion and Cross-modal Interaction Graph Convolutional Network, MFCMI). 模型采用从个体图到群体图层次化的图构建, 设计了密集-稀疏融合 (Dense-Sparse Fusion, DSF) 模块, 通过并行分支架构同时捕获脑网络的全局密集图拓扑与局部稀疏图拓扑特性. 并设计了多头交叉注意力机制的跨模态交互模块 (Cross Modal Interaction, CMI), 使非影像特征与影像特征的双向检索, 实现跨模态的交互, 实现对 ASD 的精准分类.

1 方法

本节首先介绍 MFCMI 的总体结构, 接着详细介绍 MFCMI 的关键模块.

1.1 总体框架

如图 1 所示, 功能磁共振成像通过 Harvard-Oxford (HO) 皮质结构图谱^[14]分割为 111 个脑区为感兴趣区域 (Region of Interest, ROI), 根据脑区的时间序列计算相关性, 构建个体图, 经过 GCN 得到特征矩阵. 接着经过 DSF 模块得到每个受试者的稀疏子图特征和密集图特征融合后的新特征矩阵, 级联后得到高维特征嵌入, 该嵌入与非成像特征经过 CMI 模块进行跨模态的融合后作为群体图的节点特征. 群体图构建过程中使用动态调整头数和维度, 站点嵌入解决采集中心差异, 特征聚合捕获多尺度信息, 最后通过 MLP 分类器完成二分类.

1.2 图构建

在脑网络分析中, 图构建是核心步骤, 图由节点和边构成, 表示为 $G = (V, E)$, 其中 V 表示图中所

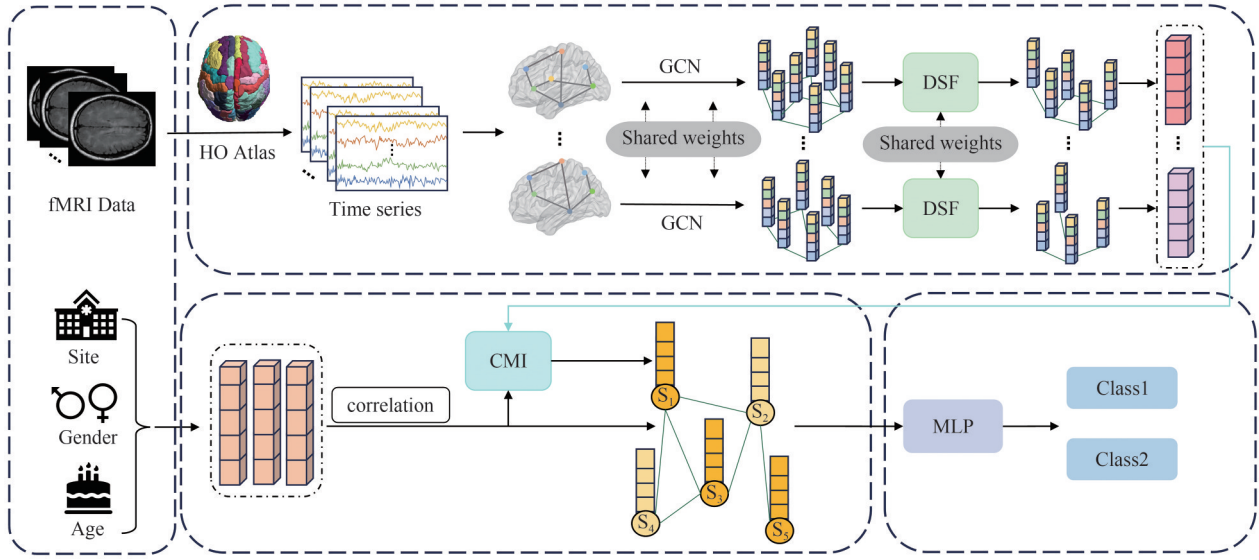


图1 MFCMI整体框架图

Fig. 1 Overall framework of MFCMI

有节点的集合, $V = \{v_1, v_2, v_3, \dots, v_N\}$, E 表示图中所有边的集合. 假设图 G 包含了 N 个二进制的节点, 每个节点与一个二进制的特征相关联. 图神经网络通常利用邻接矩阵 $A \in R^{N \times N}$ 代表图像的结构信息, 而特征矩阵 $X \in R^{N \times C}$ 用来代表节点的功能属性. 在本模型设计了由局部个体图构建到全局群体图构建的双层图结构分析框架, 分别从个体受试的大脑功能连接和跨受试者的群体表型关联两个层面进行特征学习与信息整合. 个体图构建中, 将每个脑区视为节点, 节点特征为该脑区的时间序列. 利用皮尔逊相关系数计算任意两个脑区时间序列之间的功能连接强度, 得到 111×111 的相关矩阵. 为消除噪声和弱连接, 设置阈值 $\alpha=0.65$, 保留相关系数绝对值大于 α 的连接作为边的权重, 最终构建无向图. 在群体图构建中, 将每个受试者视为节点, 利用表型数据计算节点间的相似度. 对于性别, 相同性别计1分, 不同性别计0分; 对于站点, 相同站点计1分, 不同站点计0分. 两项得分相加得到原始相似度矩阵, 再经 Z-score 标准化得到最终边权重, 并且设置阈值 $\beta=1.5$, 仅保留权重高于 β 的边. 特别地, 有研究表明男性和女性之间存在独特的大脑连接差异^[15]. 本研究构建的群体图中根据性别信息将边分为同性别连接和异性连接两类, 捕获不同性别组合间的特异性关联模式.

1.3 密集-稀疏融合模块 (Dense-Sparse Fusion Module)

DSF模块如图2所示包含两个并行的特征提取分支, 第一个分支为稀疏图特征提取分支, 基于自

注意力图池化实现稀疏子图的构建与特征提取, 该分支学习第 l 层节点的注意力分数 $S \in R^{n \times 1}$ 计算方式如下:

$$S^{(l)} = \text{GCN}(X^{(l)}, A), \quad (1)$$

其中, A 表示邻接矩阵, X 表示特征矩阵. 根据 TOP-K 池化, 设置池化比例 k_l 为 0.9, 即每层保留 90% 的节点, 保留较高比例以避免丢失潜在关键脑区, 筛选出具有高判别性的 k 个关键节点, 每个节点的特征维度为 d , 通过索引操作保留这些节点及其连接关系形成稀疏子图的特征矩阵 $X_s^{(l+1)} \in R^{k \times d}$ 和邻接矩阵 $A_s^{(l+1)} \in R^{(k \times k)}$, 表达如下:

$$\text{idx}^{(l)} = \text{TOP}_k(S^{(l)}), \quad (2)$$

$$X_s^{(l+1)} = X^{(l)} \odot S_{\text{idx}^{(l)}}, \quad (3)$$

$$A_s^{(l+1)} = A^{(l)}_{(\text{idx}^{(l)}, \text{idx}^{(l)})}. \quad (4)$$

第二个分支为密集图特征提取分支, 先使用节点聚类技术生成分配矩阵, 计算过程如下:

$$P = \text{softmax}(\text{GCN}(X, A)), \quad (5)$$

其中, $P \in R^{(m \times k_2)}$, m 为密集图中节点的个数, k_2 为预定义簇的个数, 本研究设置 k_2 为 0.5, 即将节点聚合为约 55 个簇, 该比例经网格搜索 (范围 0.3~0.7, 步长 0.1) 确定, 在保留全局信息与降低计算复杂度间取得平衡. 接着基于密集差分池化实现全节点密集图的特征聚合, 密集图的特征矩阵 X_d 和邻接矩阵 A_d 的计算方式如下:

$$X_d = P^T X, \quad (6)$$

$$A_d = P^T A P. \quad (7)$$

接着将密集图分支的分配矩阵与稀疏图分支

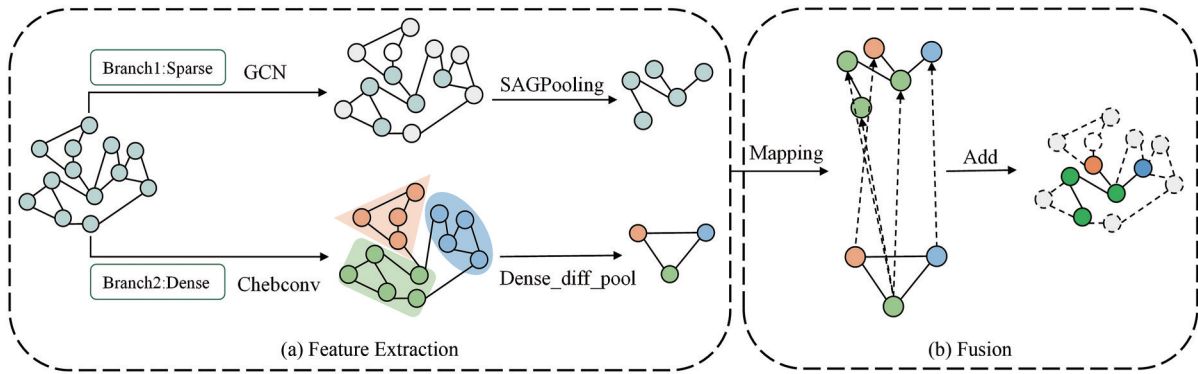


图 2 Dense-Sparse Fusion 模块结构图

Fig. 2 Structure diagram of the Dense-Sparse Fusion module

的关键节点索引进行匹配,实现密集图特征向稀疏图特征空间的映射,确保二者在特征维度上的对齐,最终通过逐元素相加的方式完成稀疏图特征与密集图特征的融合,过程如下:

$$X_c = \sigma([X_s + A_{cro} \cdot X_d] \cdot V), \quad (8)$$

其中 V 是可训练的权重矩阵, A_{cro} 表示稀疏子图和密集图两个分支之间的连通性,和分配矩阵的关系如下:

$$A_{cro}[i] = P[i]. \quad (9)$$

1.4 跨模态交互模块(Cross Modal Interaction Module)

CMI 模块的设计如图 3 所示,构建了一个基于深度注意力与门控机制的协同学习框架,包含特征投影、跨模态多头注意力、门控融合、特征增强和投影输出五阶段的处理流程,旨在通过精细化的特征交互与融合策略,充分挖掘多模态数据间的互补信息.

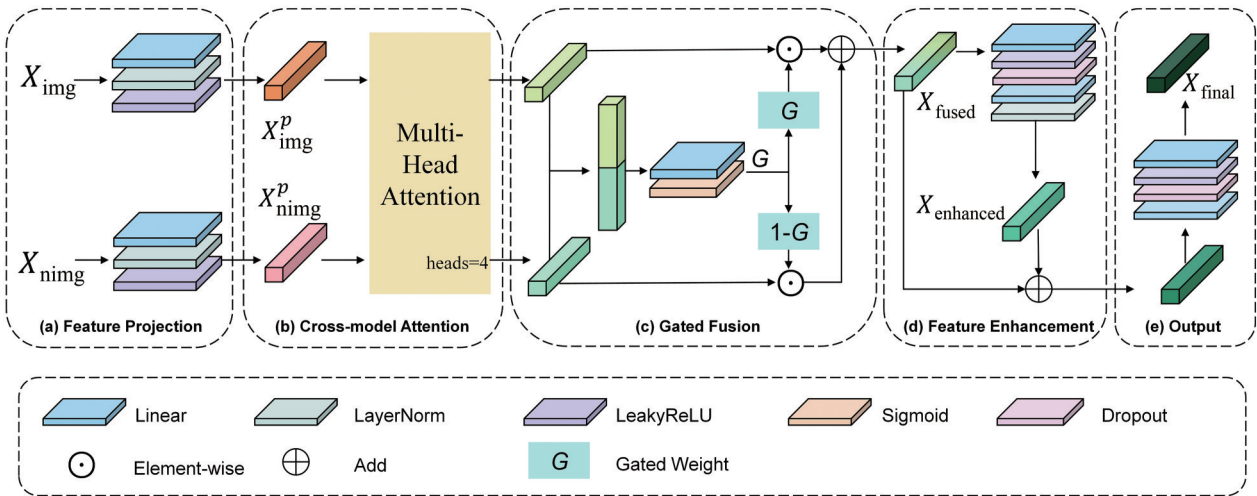


图 3 Cross Modal Interaction 模块结构图

Fig. 3 Structure diagram of the Cross Modal Interaction module

CMI 模块首先对两种模态特征分别进行投影变换如图 3(a)所示,经过 DSF 模块得到的影像特征和原始非影像特征共同映射到统一的 256 维度潜在空间,并通过层归一化和 LeakyReLU 激活函数增强特征表达能力与训练稳定性,得到 X_{img}^p 和 X_{nimg}^p . 接着添加维度为 1 的序列长度被扩展为三维的序列形式 $X_{img}^{p'}$ 和 $X_{nimg}^{p'}$,由 $W_i^Q \in R^{256 \times 256}$, $W_i^K \in R^{256 \times 256}$, $W_i^V \in R^{256 \times 256}$ 可学习参数矩阵将其投影到跨模态注意力单元中的查询(Query)空间、键(Key)空间和值(Value)空间中,并且成像特征的 QKV 矩阵的计算

如下:

$$Q_{img} = X_{img}^{p'} W_i^Q, \quad (10)$$

$$K_{img} = X_{img}^{p'} W_i^K, \quad (11)$$

$$V_{img} = X_{img}^{p'} W_i^V. \quad (12)$$

同理,非成像特征的 QKV 矩阵的计算如下:

$$Q_{nimg} = X_{nimg}^{p'} W_i^Q, \quad (13)$$

$$K_{nimg} = X_{nimg}^{p'} W_i^K, \quad (14)$$

$$V_{nimg} = X_{nimg}^{p'} W_i^V. \quad (15)$$

跨模态注意力单元如图 3(b)所示采用多头交

叉注意力机制,并且本文中注意力头个数为4.在计算注意力增强后的影像特征时,每个注意力头的输出都是用影像特征为查询去检索非影像特征中的上下文信息,拼接所有注意力头的输出,得到增强后影像特征,过程如下:

$$\text{Head}_{\text{img}}^i = \text{Softmax}\left(\frac{Q_{\text{img}}^i K_{\text{nimg}}^i}{\sqrt{D_k}}\right) V_{\text{nimg}}^i, \quad (16)$$

$$X_{\text{img}}^{\text{attn}} = \text{Concat}(\text{Head}_{\text{img}}^1, \text{Head}_{\text{img}}^2, \dots, \text{Head}_{\text{img}}^i). \quad (17)$$

同样的,在计算注意力增强后的非影像特征时,每个注意力头的输出都是用非影像特征为查询去探寻影像特征中的相关模式,拼接所有注意力头的输出,过程如下:

$$\text{Head}_{\text{nimg}}^i = \text{Softmax}\left(\frac{Q_{\text{nimg}}^i K_{\text{img}}^i}{\sqrt{D_k}}\right) V_{\text{img}}^i, \quad (18)$$

$$X_{\text{nimg}}^{\text{attn}} = \text{Concat}(\text{Head}_{\text{nimg}}^1, \text{Head}_{\text{nimg}}^2, \dots, \text{Head}_{\text{nimg}}^i). \quad (19)$$

如图3(c)模块引入了一个门控融合网络负责动态平衡两种模态特征的贡献权重,该阶段首先将拼接后的影像特征与非影像特征通过由线性层(输入512维)和Sigmoid激活函数组成的子模块生成256维的门控权重 G .最终,融合特征通过“ $G \times$ 影像特征 + $(1-G) \times$ 非影像特征”的加权方式生成.

融合后的特征将经过一个包含非线性变换与残差连接的特征增强子网络如图3(d)所示,由两层线性层(中间维度为512)、LeakyReLU激活函数和dropout层(比例0.3)组成,对门控融合后的特征进行深度编码;并将编码结果与原始融合特征残差连接,缓解深层网络中的梯度消失问题.最终,输出投影阶段如图3(e)所示将增强后的256维特征通过两层线性变换映射至2000维,以匹配后续群体图构建模块的输入维度要求.该投影过程通过非线性变换保留了融合特征的语义信息,确保输入至群体图模型的特征既包含多模态融合的丰富信息,又能与群体图的特征空间兼容.

2 实验和结果

2.1 实验设置

所有实验均在配备NVIDIA A100 GPU的服务器上,使用PyTorch1.13.0框架进行.实验预先测试了各模型在验证集上的收敛曲线,300轮后均趋于收敛,后续轮次波动 $<0.5\%$,因此统一设置训练400个epoch保证结果可重复性.模型选用AdamW优化

器,初始学习率为0.01,权重衰减为 $5e-4$.为了进行公正的比较,实验中本模型与可复现的对比模型均采用相同的数据集和预处理流程,均采用10折交叉验证,每折将数据集按8:1:1的比例划分为训练集、验证集与测试集,且保证各类别比例与总体一致.

2.2 数据集及预处理

实验是在自闭症脑成像数据交换(Autism Brain Imaging Data Exchange, ABIDE)^[16]上进行的,该数据库从全球17个站点收集了1112名受试者的数据.本文使用rs-fMRI和非成像数据,并使用由预处理连接体项目倡议提供的公开的预处理数据集版本,由连接体分析的可配置管道^[17](Configurable Pipeline for the Analysis of Connectomes, CPAC)进行预处理,具体流程如下:在BOLD信号校正阶段,采用复合型去噪策略如下:(1)伪影与噪声去除:通过ICA-based CompCor方法精准分离头部运动伪影和生理噪声成分(平均保留有效信号成分率 $>95\%$),避免单一去噪方法的局限性;(2)频率滤波:实施0.01~0.1 Hz带通滤波器,有效抑制低频呼吸运动、心跳等生理干扰及高频电子噪声;(3)信号强度校正:采用GAMMA校正模型校正全局信号强度波动,确保校正后信号拟合优度 $R^2>0.92$;(4)异常帧剔除与信号重建:预处理后的时序数据经清洗技术剔除运动超标帧(帧位移 $FD>0.5$ mm或 $DVARS>0.001$),再通过线性回归模型完成残差信号重建,保障时序数据的连续性与可靠性.本研究最终纳入了871名受试者的数据^[18],其中包括403名ASD患者和468名典型对照,受试者信息如表1所示.

表1 ABIDE数据集中受试者的表型信息

Tab. 1 Phenotypic information of subjects in the ABIDE dataset

Total Subjects	871
HC/Patients	468/403
Female/Male	144/727
Age range	6-58
Age mean	16.94

2.3 评估指标

对于分类性能的数值评估,采用主流分类评估指标,即准确率(Accuracy, ACC)、特异度(Specificity, SPE)、灵敏度(Sensitivity, SEN)、F1值和受试者工作特征曲线下面积(Area Under the Receiver Operating Characteristic Curve, AUC),定义公式如下:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (20)$$

$$\text{SPE} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (21)$$

$$SEN = \frac{TP}{TP + FN},$$

(22)

$$Pre = \frac{TP}{TP + FP},$$

(22)

$$F1 = \frac{2 \times Pre \times SEN}{Pre + SEN},$$

(23)

其中,TP(真阳性)为 ASD 患者被正确诊断的样本数. TN(真阴性)是为正常对照组被正确诊断的样本数. FP(假阳性)为正常样本被误判为 ASD 的样本数. FN(假阴性)为 ASD 患者被误判为正常的样本数.

2.4 实验结果

为了验证 MFCMI 模型的优越性能,同时确保结果的可比性,本研究将 MFCMI 与其他模型进行了比较. 所有对比的模型分为单模态方法和多模态方法的两类模型.

表2展示了对不同图神经网络模型在 ASD 诊断中的性能进行对比分析. 从整体性能分布来看,模

型性能呈现明显的多模态优于单模态的趋势. 单模态模型(仅依赖 rs-fMRI 数据)的表现普遍滞后,其中 GCN 模型性能最差,反映出仅利用脑网络拓扑特征难以捕捉 ASD 的复杂病理机制. 即使经过 ROI-aware 卷积优化的 BrainGNN,其性能指标的低表现说明缺乏表型数据的补充时,模型对患者群体的识别能力有限. ContrastPool 虽引入对比双注意力机制筛选判别性脑区,但性能依旧低于多模态的模型,进一步验证了单一模态数据的表征瓶颈. 多模态模型(融合 rs-fMRI 与表型数据)的性能实现显著跃升. 早期多模态模型 Population-GCN 凭借表型相似度构建群体图,ACC 达 70.54%,较基础的 GCN 提升 3.74%,证明群体关联信息的价值. FC-HGNN 通过半球内信息递进到半球间信息提取以及性别差异建模方法,ACC 提升至 87.29%,展现出精细图结构设计的优势.

表2 本方法与其他方法在 ASD 分类中的性能比较(平均值)

Tab. 2 Performance comparison of the proposed method with other methods in ASD classification (mean)

Methods	ACC/%	SEN/%	SPE/%	F1/%	AUC/%	模态
GCN[19]	66.80	70.20	71.20	70.28	67.49	单
BrainGNN[7]	67.40	67.30	70.61	65.32	67.42	单
MVS-GCN[8]	67.51	66.93	69.75	67.39	68.71	单
MAHGCN[20]	70.87	70.92	72.43	71.21	70.62	单
MGCA-RAFFNet[21]	72.30	71.06	70.25	70.13	76.22	单
MGATBC[22]	76.61	77.93	79.62	79.33	82.43	单
ContrastPool[23]	65.12	65.10	66.50	64.92	68.74	单
Population-GCN[9]	70.54	76.21	68.02	73.29	71.07	多
EV-GCN[18]	80.26	80.77	82.35	80.60	83.06	多
MCGC-SPFFN[13]	68.77	76.68	70.10	72.59	68.08	多
LG-GNN[11]	71.53	76.52	65.67	74.32	71.55	多
ACA-GCN[24]	86.91	89.04	87.39	87.77	88.51	多
WL_DeepGCN[25]	71.21	71.81	69.21	69.76	73.36	多
FC-HGNN[26]	87.29	89.12	85.54	88.30	94.83	多
MFCMI	88.40	90.82	85.60	89.37	96.35	多

注:加粗字体表示最优结果.

本文提出的 MFCMI 模型在所有评估指标中均表现最优,ACC 达 88.40%,AUC 96.35%,较对比的最优模型 FC-HGNN 的 ACC 提升 1.11%,AUC 提升 1.52%. 从指标对比来看,MFCMI 的 SEN 显著高于其他模型,说明模型对 ASD 患者的识别能力更强;AUC 突破 96%,表明模型的分类边界更清晰,对难分样本的判别能力更优. 十折交叉验证中每一折的 ROC 曲线如图 4 所示,展示了 MFCMI 的稳定性. 对比单模态中最优性能的 MGATBC 与多模态方法中性能最优的 FC-HGNN,三种方法的性能指标差异在

图 5 中得到很好的表示.

为评估模型的实际部署潜力,本研究统计了 MFCMI 的可训练参数量及推理效率. 模型总参数量约为 8.32M. 在 NVIDIA A100 GPU 上,单个受试者完成完整前向传播的平均耗时约为 13.6ms,对整个数据集(871 例)的推理总耗时约为 11.8s. 上述指标表明,MFCMI 具有适中的参数量和较快的推理速度,能够满足临床辅助诊断对计算效率的基本要求.

2.5 消融实验

为验证 MFCMI 模型中 DSF 模块和 CMI 模块的

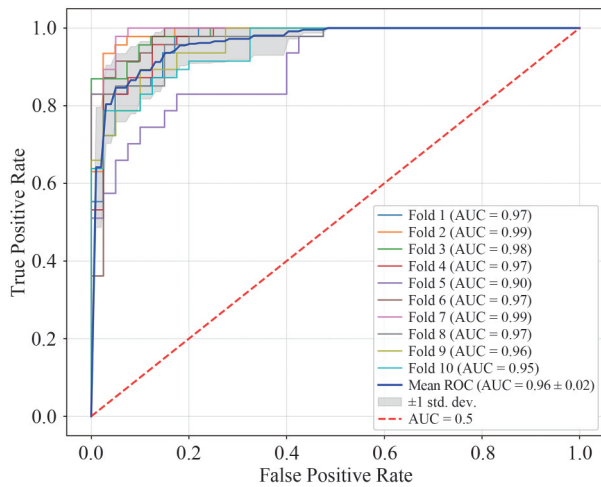


图4 十折交叉验证中每一折的ROC曲线图

Fig. 4 ROC curves of each fold in the 10-fold cross-validation

必要性和优越性,本文设计了针对两个模块的定向消融实验.基线模型采用完整的MFCMI架构.

2.5.1 针对DSF模块的消融实验

表3展示了针对DSF模块的三个消融实验以及完整的MFCMI模型在ABIDE数据集上的平均性

表3 Dense-Sparse Fusion模块的消融实验结果

Tab. 3 Ablation experiment results of the Dense-Sparse Fusion module

Ablations	Features		Metrics				
	Dense	Sparse	ACC/%	SEN/%	SPE/%	F1/%	AUC/%
w/o Sparse	√	×	85.29	90.62	79.13	86.91	93.53
w/o Dense	×	√	87.02	89.33	84.34	88.12	95.42
w/o DSF	×	×	86.45	88.23	84.32	87.53	95.29
MFCMI	√	√	88.40	90.82	85.60	89.37	96.35

注:加粗字体表示最优结果.

由表3可知,仅使用密集分支的模型各项性能显著下降,原因在于丢失了Sparse分支保留的局部关键脑区连接特征,导致特征的细粒度区分能力下降.仅使用稀疏分支的模型性能仍低于基线模型,ACC和F1较基线模型分别下降1.38和1.25个百分点,说明仅依赖局部拓扑特征无法完整捕捉脑连接组的全局关联模式.完全移除DSF模块的模型ACC和F1较基线模型分别下降2.25和1.84个百分点.这一结果证明,直接扁平化GCN节点特征的方式无法有效整合脑连接组的拓扑信息,DSF模块缺失会导致特征表达能力的本质性下降.

2.5.2 针对CMI模块的消融实验

表4展示了针对CMI模块的两个消融实验及完整的MFCMI模型在ABIDE数据集上的平均性能.第一组实验移除CMI模块及非图像特征的融合过程,仅保留个体图提取的图像特征直接输入群体图构建并进行分类.第二组实验采用简单拼接和线性

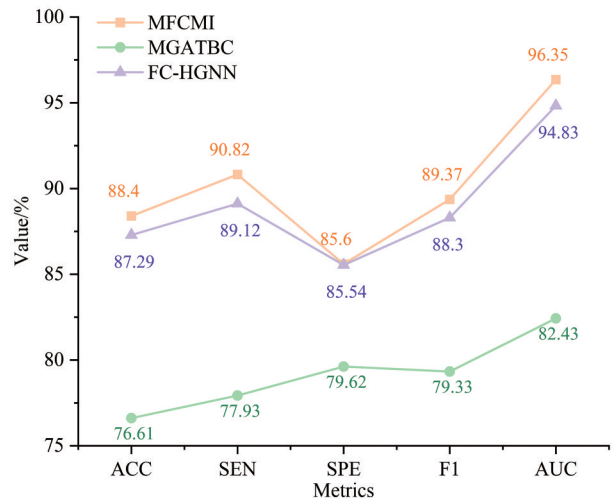


图5 模型与MGATBC和FC-HGNN的性能指标差异对比图

Fig. 5 Comparison of performance metrics between the model and MGATBC and FC-HGNN

能.第一组实验移除DSF模块的Sparse分支.第二组实验移除DSF模块的Dense分支.第三组实验完全移除DSF模块,无任何密集与稀疏的融合过程.

投影替代CMI模块.

可知MFCMI在所有评估指标上均表现最优,仅使用成像数据的消融模型的各项性能下降,这一结果表明人口统计学特征与ASD发病机制的关联、站点差异的校正作用的补充具有重要价值.值得注意的是,使用简单融合的消融模型性能甚至次于仅使用成像数据的消融模型,且仍低于基线模型,其平均ACC为84.26%,F1分数为85.05%,较基线模型分别下降4.14和4.32个百分点.这说明简单的特征拼接可能实现模态信息的冗余叠加,甚至可能不如单独使用影像数据的诊断效果好.

3 讨论

3.1 可解释性分析

本研究最重要的目标之一是识别疾病分类的显著生物标志物区域,为此,DSF模块中的分支1被

专门设计用于获得每个节点的分类重要性得分,使用 BrainNetViewer^[27] 工具绘制贡献度排行前 10 的 ROI 位置分布如图 6 显示.

表 4 Cross Modal Interaction 模块消融实验结果
Tab. 4 Ablation experiment results of the Cross Modal Interaction module

Ablations	Features		Metrics				
	Img	Nonimg	ACC/%	SEN/%	SPE/%	F1/%	AUC/%
w/ img	√	×	86.33	88.02	84.36	87.36	94.56
w/ simple fusion	√	√	84.26	83.35	85.36	85.05	94.00
MFCMI	√	√	88.40	90.82	85.60	89.37	96.35

注:加粗字体表示最优结果.

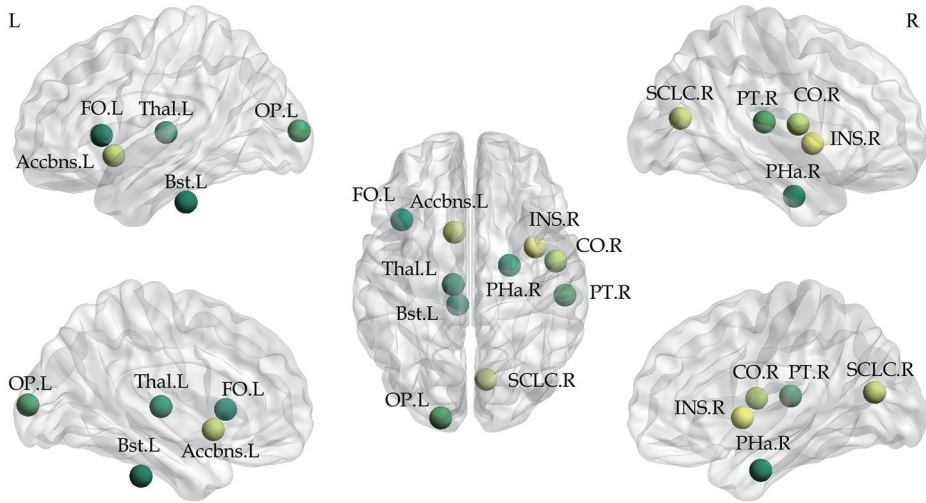


图 6 贡献度排行前 10 的 ROI 位置分布图
Fig. 6 Distribution of the top 10 ROIs with the highest contribution

模型通过节点得分排序识别出对自闭症诊断贡献度最高的十个脑区,分别为右侧岛叶(INS.R)、左侧额下回(FO.L)、右侧角回(PT.R)、左侧伏隔核(Accbns.L)、右侧上丘(SCLC.R)、左侧枕叶(OP.L)、右侧楔叶(CO.R)、左侧终纹床核(Bst.L)、右侧海马旁回(PHa.R)及左侧丘脑(Thal.L),该结果与自闭症神经影像学领域的已有研究结论高度契合^[28-29],为模型诊断的可靠性提供了神经解剖学层面的解释.例如,已有研究证实 ASD 患者岛叶与默认模式网络、突显网络的功能连接存在显著异常,其结构与功能的改变会直接影响个体对社交信号的识别与情绪共情能力,这与本研究中岛叶成为最高贡献脑区的结果一致^[30].其余的大脑区域也都与社交,感知和执行的神经功能有关^[31-33].

3.2 局限性及未来发展

本研究虽在 ASD 诊断中取得了较好的性能,但仍存在若干局限性.在数据层面,研究仅基于 rs-fMRI 数据和有限的人口统计学特征展开,未整合结构磁共振成像、扩散张量成像等其他模态神经影像数据,后续可以整合多模态的神经影像数据以避免

错失与 ASD 病理机制相关的关键互补信息;同时,脑网络分析技术和可解释性方面,研究虽识别了关键生物标志物脑区,但对脑区之间动态功能连接的时序变化、模态间交互的生理意义解释不够深入,后续可结合脑网络动态分析技术、可视化模态交互过程深入解析关键脑区的功能协同机制,为临床干预提供更明确的生物标志物.

4 结语

本文提出了一种基于多尺度融合与跨模态交互的图卷积网络(MFCMI)进行自闭症谱系障碍的诊断,旨在改善多模态信息融合不充分、脑网络多尺度特征捕捉不足及群体关联建模单一的问题.该模型构建了个体图到群体图的层次化结构,通过 DSF 模块有效整合了脑网络的局部核心特征与全局拓扑关联. CMI 模块采用双向多头注意力与门控融合机制,实现了脑影像特征与人口统计学特征的深度交互.在 ABIDE 数据集上进行 10 折交叉验证,结果表明模型各项性能指标均优于现有先进方法,且

成功识别出与自闭症密切相关的关键脑区生物标志物,为模型诊断提供了神经解剖学层面的解释。

参 考 文 献

- [1] Liu M, Li B, Hu D. Autism spectrum disorder studies using fMRI data and machine learning: A review [J]. *Frontiers in Neuroscience*, 2021, 15: 697870.
- [2] Korgaonkar M S, Fornito A, Williams L M, et al. Abnormal structural networks characterize major depressive disorder: A connectome analysis [J]. *Biological Psychiatry*, 2014, 76(7): 567-574.
- [3] Guo X, Dominick K C, Minai A A, et al. Diagnosing autism spectrum disorder from brain resting-state functional connectivity patterns using a deep neural network with a novel feature selection method [J]. *Frontiers in Neuroscience*, 2017, 11: 460.
- [4] Cao J, Yang L, Sarriani P G, et al. Dementia classification using a graph neural network on imaging of effective brain connectivity [J]. *Computers in Biology and Medicine*, 2024, 168: 107701.
- [5] Soussia M, Rekik I. High-order connectomic manifold learning for autistic brain state identification [C]// *Connectomics in NeuroImaging*. Cham: Springer, 2017: 51-59.
- [6] Wu Z, Pan S, Chen F, et al. A comprehensive survey on graph neural networks [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(1): 4-24.
- [7] Li X, Zhou Y, Dvornek N, et al. BrainGNN: Interpretable brain graph neural network for fMRI analysis [J]. *Medical Image Analysis*, 2021, 74: 102233.
- [8] Wen G, Cao P, Bao H, et al. MVS-GCN: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis [J]. *Computers in Biology and Medicine*, 2022, 142: 105239.
- [9] Parisot S, Ktena S I, Ferrante E, et al. Disease prediction using graph convolutional networks: Application to Autism Spectrum Disorder and Alzheimer's disease [J]. *Medical Image Analysis*, 2018, 48: 117-130.
- [10] Mohammadi H, Karwowski W. Graph neural networks in brain connectivity studies: Methods, challenges, and future directions [J]. *Brain Sciences*, 2024, 15(1): 17.
- [11] Zhang H, Song R, Wang L, et al. Classification of brain disorders in rs-fMRI via local-to-global graph neural networks [J]. *IEEE Transactions on Medical Imaging*, 2023, 42(2): 444-455.
- [12] Zheng S, Zhu Z, Liu Z, et al. Multi-modal graph learning for disease prediction [J]. *IEEE Transactions on Medical Imaging*, 2022, 41(9): 2207-2216.
- [13] Jiang X, Ding X, Xia Z, et al. Self-perceptive feature fusion network with multi-channel graph convolution for brain disorder diagnosis [J]. *Expert Systems with Applications*, 2025, 284: 127984.
- [14] Desikan R S, Ségonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest [J]. *NeuroImage*, 2006, 31(3): 968-980.
- [15] Ingallhalikar M, Smith A, Parker D, et al. Sex differences in the structural connectome of the human brain [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2014, 111(2): 823-828.
- [16] Di Martino A, Yan C G, Li Q, et al. The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism [J]. *Molecular Psychiatry*, 2014, 19(6): 659-667.
- [17] Craddock C C, Sikka S, Cheung B, et al. Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (C-PAC) [C]// *5th INCF Congress of Neuroinformatics*. Munich: INCF, 2013: 1-2.
- [18] Huang Y, Chung A C S. Edge-variational graph convolutional networks for uncertainty-aware disease prediction [C]// *Medical Image Computing and Computer Assisted Intervention - MICCAI 2020*. Cham: Springer, 2020: 562-572.
- [19] Jiang B, Zhang Z, Lin D, et al. Semi-supervised learning with graph learning-convolutional networks [C]// *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach: IEEE, 2019: 11305-11312.
- [20] Liu M, Zhang H, Shi F, et al. Hierarchical graph convolutional network built by multiscale atlases for brain disorder diagnosis using functional connectivity [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(11): 15182-15194.
- [21] Ma Y, Cui W, Liu J, et al. A multi-graph cross-attention-based region-aware feature fusion network using multi-template for brain disorder diagnosis [J]. *IEEE Transactions on Medical Imaging*, 2024, 43(3): 1045-1059.
- [22] Yu R, Pan C, Fei X, et al. Multi-graph attention networks with bilinear convolution for diagnosis of schizophrenia [J]. *IEEE Journal of Biomedical and Health Informatics*, 2023, 27(3): 1443-1454.
- [23] Xu J, Bian Q, Li X, et al. Contrastive graph pooling for explainable classification of brain networks [J]. *IEEE*

- Transactions on Medical Imaging, 2024, 43(9): 3292-3305.
- [24] Han L, Wang M, Zhang X, et al. Adaptive composing augmentations on multi-modal graph convolutional network for disease prediction [J]. Computers and Electrical Engineering, 2025, 123: 110277.
- [25] Wang M, Guo J, Wang Y, et al. Multimodal autism spectrum disorder diagnosis method based on DeepGCN [J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023, 31: 3664-3674.
- [26] Gu Y, Peng S, Li Y, et al. FC-HGNN: A heterogeneous graph neural network based on brain functional connectivity for mental disorder identification [J]. Information Fusion, 2025, 113: 102619.
- [27] Xia M, Wang J, He Y. BrainNet Viewer: A network visualization tool for human brain connectomics [J]. PLoS One, 2013, 8(7): e68910.
- [28] Schuetze M, Park M T M, Cho I Y, et al. Morphological alterations in the thalamus, striatum, and pallidum in autism spectrum disorder [J]. Neuropsychopharmacology, 2016, 41(11): 2627-2637.
- [29] Monk C S, Peltier S J, Wiggins J L, et al. Abnormalities of intrinsic functional connectivity in autism spectrum disorders, [J]. NeuroImage, 2009, 47(2): 764-772.
- [30] Amaral D G, Schumann C M, Nordahl C W. Neuroanatomy of autism [J]. Trends in Neurosciences, 2008, 31(3): 137-145.
- [31] Parellada M, Penzol M J, Pina L, et al. The neurobiology of autism spectrum disorders [J]. European Psychiatry, 2014, 29(1): 11-19.
- [32] Li D, Karnath H O, Xu X. Candidate biomarkers in children with autism spectrum disorder: A review of MRI studies [J]. Neuroscience Bulletin, 2017, 33(2): 219-237.
- [33] Press C, Weiskopf N, Kilner J M. Dissociable roles of human inferior frontal gyrus during action execution and observation [J]. NeuroImage, 2012, 60(3): 1671-1677.

(责编&校对 雷建云)