



学校代码： 10254
学 号： 202330310255
密 级：

上海海事大学

SHANGHAI MARITIME UNIVERSITY

硕 士 学 位 论 文

MASTER'S THESIS

学位类别：	专业学位
论文题目：	基于多分辨率脑图谱分析与异构图学习的神经发育障碍诊断研究
专业名称：	计算机技术
作者姓名：	陈磊
校内导师：	曾卫明
校外导师：	
完成日期：	二〇二五年五月

论文独创性声明

本论文是我个人在导师指导下进行的研究工作及取得的研究成果。论文中除了特别加以标注和致谢的地方外，不包含其他人或者其他机构已经发表或撰写过的研究成果。其他同志对本研究的启发和所做的贡献均已在论文中作了明确的声明并表示了感谢。

作者签名：_____日期：_____

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权上海海事大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于（请在以下相应方框内打“√”）：

☐ 保密，在_____年解密后适用本使用授权书。

☒ 不保密。

作者签名：_____导师签名：_____日期：_____

摘 要

神经发育障碍(NDD)，尤其是自闭症谱系障碍(ASD)和注意力缺陷多动障碍(ADHD)，是复杂的神经发育疾病，特征为大脑功能网络连接异常，导致社交互动、沟通及行为持续性困难。功能性磁共振成像(fMRI)作为无创神经影像技术，为NDD研究提供了重要手段，但基于fMRI的计算机辅助诊断仍面临多尺度特征捕捉、多模态数据整合及可解释性等挑战。本研究针对这些问题，提出了创新性方法框架，提升了NDD诊断的准确性和可解释性。

本研究首先开发了基于多分辨率脑区划分的双通道特征融合模型。该模型采用不同粒度脑区图谱(1000、400和100个区域)进行功能连接分析，实现多尺度特征提取，同时捕捉局部精细连接和全局网络特征。在特征融合方面，模型设计了两个互补通道：图神经网络驱动节点表示学习(GNN-DN)模块和Transformer驱动的功能连接分析(Transformer-FA)模块。GNN-DN通过度归一化策略有效建模局部邻域连接模式，而Transformer-FA利用全连接注意力机制捕捉远距离脑区间依赖关系。

上述方法虽取得了一定成效，但仍存在几个关键局限性，例如非影像数据整合有限、跨模态关系建模不足、理论基础薄弱、可解释性机制有限，以及多分辨率策略导致的高计算复杂度等问题。为解决这些问题，研究创新性地将信息瓶颈理论引入NDD异构图学习，提出了可解释的信息瓶颈异构图神经网络(I^2B -HGNN)框架。该框架包含两个相互配合的核心组件：信息瓶颈图转换器(IBGraphFormer)和信息瓶颈异构图注意力网络(IB-HGAN)。IBGraphFormer整合Transformer的全局依赖建模与GNN的局部结构感知能力，通过信息瓶颈引导的池化机制识别关键生物标志物。IB-HGAN则将信息瓶颈原理扩展至多模态融合领域，通过异构图结构实现fMRI和非影像数据的深度整合，为诊断提供全面依据。

实验结果表明，所提方法在多个数据集上的NDD诊断性能显著优于现有方法。多分辨率脑区划分和双通道特征融合策略使模型在ASD分类任务中达到较高的准确率、敏感性和特异性，验证了多尺度特征提取和平衡局部-全局信息的重要性。同时，基于信息瓶颈的异构图学习框架不仅实现出色诊断性能，还提供了丰富可解释性分析结果，成功识别了与NDD密切相关的关键脑区。这些发现与现有神经科学文献高度

一致，表明这些网络在自我参照处理、社会认知和执行功能等与 NDD 相关的认知过程中扮演重要角色。

综上所述，本研究通过多分辨率脑区划分与双通道特征融合、基于信息瓶颈的异构图学习方法，构建了强大而灵活的 NDD 计算辅助诊断框架。该框架实现了高精度诊断性能和丰富可解释性分析，对促进 NDD 早期诊断以及深化对其神经机制理解具有重要价值。未来研究将扩展这些方法至更广泛的神经精神疾病领域，探索动态时变脑网络特性，并推动相关技术向临床实践转化。

关键词：神经发育障碍，功能性磁共振成像，图神经网络，异构图学习，多模态

Research on Neurodevelopmental Disorder Diagnosis Based on Multi-resolution Brain Atlas Analysis and Heterogeneous Graph Learning

ABSTRACT

Neurodevelopmental disorders (NDDs), particularly Autism Spectrum Disorder (ASD), represent a complex class of neurodevelopmental conditions characterized by aberrant functional brain network connectivity, resulting in persistent difficulties in social interaction, communication, and behavior. While functional magnetic resonance imaging (fMRI) has emerged as a valuable non-invasive neuroimaging technique for NDD research, computer-aided diagnosis based on fMRI data continues to face significant challenges in multi-scale feature extraction, multimodal data integration, and interpretability. This study addresses these challenges by proposing an innovative methodological framework that enhances both the accuracy and interpretability of NDD diagnosis.

The research first develops a dual-channel feature fusion model based on multi-resolution brain parcellation. This model employs brain templates of varying granularity (1000, 400, and 100 regions) for functional connectivity analysis, enabling multi-scale feature extraction that simultaneously captures fine-grained local connections and global network characteristics. For feature fusion, the model implements two complementary channels: a Graph Neural Network-driven Node representation learning (GNN-DN) module and a Transformer-driven Functional connectivity Analysis (Transformer-FA) module. The GNN-DN module effectively models local neighborhood connectivity patterns through degree normalization strategies, while the Transformer-FA module captures long-range dependencies between brain regions using fully-connected attention mechanisms.

While the aforementioned methods have achieved certain efficacy, several critical limitations persist, including limited integration of non-imaging data, insufficient cross-modal relationship modeling, weak theoretical foundations, limited interpretability mechanisms, and high computational complexity resulting from multi-resolution strategies. To address these challenges, this research innovatively incorporates Information Bottleneck theory into NDD heterogeneous graph learning, proposing an Interpretable Information Bottleneck Heterogeneous Graph Neural Network (I²B-HGNN) framework. This framework comprises two core components: Information Bottleneck Graph Transformer (IBGraphFormer) and Information Bottleneck Heterogeneous Graph Attention Network (IB-HGAN). IBGraphFormer integrates Transformer's global dependency modeling with GNN's local structure perception capabilities, identifying key biomarkers through an information bottleneck-guided pooling mechanism. IB-HGAN extends the information bottleneck principle to multimodal fusion, implementing deep integration of fMRI and non-imaging data through heterogeneous graph structures, providing comprehensive diagnostic evidence.

Experimental results demonstrate that the proposed methods significantly outperform existing approaches in NDD diagnosis across multiple datasets. The multi-resolution brain parcellation and dual-channel feature fusion strategies achieve high accuracy, sensitivity, and specificity in ASD classification tasks, validating the importance of multi-scale feature extraction and balanced local-global information processing. Concurrently, the information bottleneck-guided heterogeneous graph learning framework not only delivers excellent diagnostic performance but also provides rich interpretability analysis, successfully identifying key brain regions closely associated with NDD. These findings align with existing neuroscience literature, indicating these networks play crucial roles in self-referential processing, social cognition, and executive functions related to NDD.

In conclusion, this research constructs a robust and flexible computational framework for NDD diagnosis through multi-resolution brain parcellation, dual-

channel feature fusion, and information bottleneck-guided heterogeneous graph learning methods. This framework achieves high-precision diagnostic performance with rich interpretability analysis, offering significant value for promoting early NDD diagnosis and deepening understanding of its neural mechanisms. Future research will extend these methods to broader neuropsychiatric disorders, explore dynamic time-varying brain network characteristics, and advance the translation of these technologies into clinical practice.

Lei CHEN(Computer Science and Technology)

Directed by Professor Weiming ZENG

KEYWORDS: neurodevelopmental disorders, functional magnetic resonance imaging, gnn, heterogeneous graph learning, multimodal

目 录

摘 要 I

ABSTRACT..... III

目 录 VI

图目录 IX

表目录 X

第一章 绪论 1

 1.1 研究背景与意义..... 1

 1.2 国内外研究历史与现状..... 2

 1.2.1 传统方法与早期尝试 2

 1.2.2 深度学习方法的发展 2

 1.2.3 现有方法的局限性 2

 1.3 本文研究工作创新点..... 3

 1.4 文章结构..... 4

第二章 相关理论基础 6

 2.1 fMRI 与 BOLD 信号 6

 2.1.1 fMRI 技术基本原理 6

 2.1.2 BOLD 信号的生理基础 7

 2.2 AAL 图谱与 Schaefer 图谱 8

 2.2.1 图谱基本概念与背景 8

 2.2.2 AAL 图谱..... 9

 2.2.3 Schaefer 图谱..... 10

 2.3 图神经网络及异构图理论..... 12

 2.3.1 图神经网络及其基本原理 12

 2.3.2 异构图理论 14

 2.4 Transformer 模块与图结构结合 16

 2.4.1 Transformer 模块基本概念..... 16

2.4.2 Transformer 与图神经网络的融合	17
2.5 本章小结	19
第三章 基于多分辨率脑区划分与双通道特征融合的自闭症谱系障碍识别 ...	21
3.1 多分辨率脑区域分割	21
3.2 双通道特征融合模块	23
3.2.1 GCN-DN 模块	24
3.2.2 Transformer-FA 模块	25
3.3 基于多分辨率脑区与双通道特征融合的自闭症谱系障碍诊断实验 ...	27
3.3.1 数据预处理	27
3.3.2 实验设置和评估指标	28
3.3.3 分类性能对比实验及消融实验	29
3.3.4 脑网络可解释性	30
3.4 本章小结	32
第四章 基于信息瓶颈引导的异构图学习用于可解释的神经发育障碍诊断 ...	34
4.1 信息论与信息瓶颈原理	34
4.2 信息瓶颈异构图神经网络模型	34
4.2.1 模型结构与流程	34
4.2.2 信息瓶颈图变换器模块	35
4.2.3 信息瓶颈异构图注意力网络模块	39
4.3 基于 I^2B -HGNN 的神经发育障碍诊断实验	43
4.3.1 数据集及数据预处理	43
4.3.2 模型训练与实现	43
4.4 基于 I^2B -HGNN 的神经发育障碍诊断实验结果与分析	44
4.4.1 分类性能对比实验	44
4.4.2 I^2B -HGNN 消融实验	45
4.4.3 解释性分析	46
4.5 本章小结	49
第五章 总结与展望	50

5.1 论文工作总结..... 50

5.2 未来展望..... 51

致 谢 53

参考文献 54

攻读硕士期间发表的论文及获奖情况 61

图目录

图 2-1 fMRI 数据采集的基本流程图.....	7
图 2-2 神经血管耦合模型图.....	8
图 2-3 标准化处理流程图.....	9
图 2-4 AAL 图谱构建过程图	10
图 2-5 Schaefer 图谱构建过程图	11
图 2-6 GNN 基本流程图	13
图 2-7 异构图神经网络流程图.....	15
图 2-8 Transformer 基本架构图	16
图 2-9 融合架构图.....	19
图 3-1 多层脑区划分.....	21
图 3-2 构建功能网络的图结构.....	23
图 3-3 双通道特征融合模块.....	24
图 3-4 ASD 分类的六个最重要的大脑网络.....	31
图 4-1 I ² B-HGNN 框架	35
图 4-2 异构人群图.....	40
图 4-3 ASD 和 ADHD 的前 10 个关键脑区及其归一化互信息分数.....	47
图 4-4 ASD 和 ADHD 的前 10 个关键脑区及其归一化互信息分数.....	48

表目录

表 3-1 在 ABIDE 数据集上的实验结果.....	29
表 3-2 消融实验结果.....	29
表 4-1 不同方法在 ABIDE-I 和 ADHD-200 数据集上的分类性能对比	44
表 4-2 $\text{I}^2\text{B-HGNN}$ 消融实验结果.....	45

第一章 绪论

1.1 研究背景与意义

神经发育障碍(NDD)，尤其是自闭症谱系障碍(ASD)和注意力缺陷多动障碍(ADHD)，是一类对认知和行为发展产生深远影响的疾病。其中，ASD作为一种复杂的神经发育疾病，其特征是长期存在社交互动、沟通和行为困难，对诊断和干预都提出了重大挑战^[1]。据统计，全球范围内 ASD 的患病率约为 1-2%，且呈现出逐年上升的趋势，给患者家庭和社会带来了沉重的负担^[2]。同时，ADHD 作为另一种常见的神经发育障碍，以注意力不集中、多动和冲动为主要表现，影响着全球约 5-7%的儿童和 2-5%的成年人^[3]。传统的 NDD 诊断主要依赖于专业医师基于行为观察和评估量表进行的临床判断，这种方法不可避免地引入了诊断的主观性和延迟。随着神经影像学技术的发展，功能性磁共振成像(fMRI)已成为理解神经发育障碍背后神经机制的重要工具，并为客观计算机辅助诊断提供了坚实基础^[4]。fMRI 能够无创地捕捉大脑活动的空间和时间模式，揭示潜在的神经功能连接异常，这些异常被认为是 NDD 核心症状的神经基础^[5]。

然而，fMRI 数据的高维度、复杂的时空特性以及个体间的异质性，使得从中提取有诊断价值的生物标志物并构建有效的计算模型成为一项挑战。此外，保持模型的可解释性，即能够识别与疾病相关的关键脑区和网络，对于促进临床应用和增进对 NDD 神经机制的理解至关重要^[6]。

另一方面，单一模态的脑影像数据虽然提供了重要信息，但难以全面反映 NDD 的复杂性。临床实践表明，非影像数据（如性别、年龄、惯用手等）也包含重要的诊断信息^[7]。因此，如何有效整合多模态数据，同时保持模型的可解释性，成为计算机辅助 NDD 诊断研究中的一个关键问题。

本研究旨在开发先进的计算方法，通过整合多尺度脑区信息和多模态数据，构建既能提高诊断准确率又能保持可解释性的模型，以支持更加精准和个性化的 NDD 诊断，最终促进相关领域的基础和临床研究进展。

1.2 国内外研究历史与现状

1.2.1 传统方法与早期尝试

在计算机辅助 NDD 诊断的早期研究中,传统的机器学习方法被广泛应用于提取和分析 fMRI 数据中的特征。支持向量机(SVM)^[8]、岭分类器^[9]和逻辑回归^[10]等方法通过提取大脑功能和结构特征来进行 ASD 和 ADHD 的分类。这些方法通常结合功能连接(FC)特征,如区域间的时间序列相关性,以识别潜在的神经标志物。

早期研究主要集中在单一模态数据上,通过计算不同脑区之间的功能连接矩阵,然后使用特征选择方法筛选出与疾病相关的连接特征^[11]。虽然这些方法取得了一定的成功,但它们在处理高维复杂的 fMRI 数据时面临显著挑战,特别是难以有效捕捉脑区之间的动态交互和层次结构信息^[12]。

1.2.2 深度学习方法的发展

随着深度学习技术的快速发展,研究者开始探索更复杂的网络结构来处理神经影像数据。长短期记忆网络(LSTM)^[13]被用于分析 fMRI 的时间动态特性,而卷积神经网络(CNN)^[14]则被应用于空间特征的提取。这些方法在提高分类性能方面取得了一定进展,但它们往往将脑视为相互独立的体素或区域的集合,忽略了其固有的网络结构。

近年来,图神经网络(GNN)^[15]的兴起为脑连接组分析提供了自然的框架。GNN 将大脑视为一个网络,其中脑区作为节点,功能连接作为边,通过消息传递机制学习节点表示。这种方法能够直接在脑网络上进行学习,保留其拓扑结构信息,因此在 NDD 分类任务中表现出色^[16]。与此同时,基于 Transformer 的模型^[17]因其强大的全局依赖关系建模能力也被引入脑影像分析领域,能够捕捉脑区之间的长距离交互^[18]。

1.2.3 现有方法的局限性

尽管深度学习方法在 NDD 分类性能上有所提高,但当前研究仍然面临几个重要挑战:

- (1) 多尺度脑网络建模的不足:大脑具有复杂的层次组织结构,从微观的神经

元连接到宏观的功能网络。现有方法往往难以同时捕捉不同尺度上的信息，导致特征提取不全面^[19]。

(2) 可解释性缺乏：许多深度学习模型如黑盒运作，难以识别与疾病相关的具体脑区和连接模式，限制了其在临床决策支持中的应用^[20]。

(3) 多模态数据整合困难：虽然非影像数据（如性别、年龄、惯用手等）包含重要的诊断信息，但现有方法在整合这些异构数据时往往效率低下，或者仅简单地将它们映射到边权重上，未能充分利用多源信息^[21]。

(4) 过拟合风险：复杂的深度学习模型在处理相对较小的神经影像数据集时容易过拟合，影响其泛化能力^[22]。

(5) 局部与全局信息平衡：GNN 擅长捕捉局部连接模式但难以处理全局依赖关系，而 Transformer 虽然善于建模全局关系但缺乏对局部结构的感知能力^[23]。

最近的研究试图解决这些挑战，例如时空混合注意力网络^[24]采用基于 Transformer 的注意机制进行时间建模；可解释和可泛化的图神经网络^[25]使用元学习增强模型泛化能力。然而，这些方法仍然难以同时解决上述所有挑战，特别是在保持高分类性能的同时实现良好的可解释性，以及有效整合多模态数据方面存在不足。

1.3 本文研究工作创新点

针对当前研究中存在的局限性，本文提出了一系列创新性方法来改进 NDD 的计算机辅助诊断。主要创新点包括：

(1) 多分辨率脑区划分与双通道特征融合

本研究提出了一种基于多分辨率脑区划分的模型，该模型能够同时捕捉不同尺度上的脑网络结构信息^[26]。具体而言，我们使用多种不同粒度的脑区图谱进行功能连接计算，从而获取从粗粒度到细粒度的全面脑网络特征。这种方法能够更全面地捕捉大脑的层次结构信息，提高模型的代表能力。

同时，我们设计了一种双通道特征融合架构，包含图神经网络驱动节点表示学习 (GNN-DN) 和 Transformer-驱动的功能连接分析 (Transformer-FA) 两个模块。GNN-DN 模块专注于学习局部脑区连接模式，而 Transformer-FA 模块则捕捉全局依赖关系。这种双通道设计能够有效平衡局部和全局信息，克服单一方法的局限性^[27]。

（2）基于信息瓶颈理论的异构图学习框架

本研究引入信息瓶颈 (IB) 原理^[28]来指导 NDD 的异构图学习，提出了一个统一的理论框架，同时优化特征提取和多模态融合过程。该框架通过最少但充分的生物标志物识别和跨模态关系保存，实现了诊断性能和可解释性的平衡。

具体而言，我们设计了信息瓶颈图转换器 (IBGraphFormer)，它通过信息压缩原理从脑功能网络中提取最少的充分生物标志物^[29]，同时保持必要的连接模式。IBGraphFormer 将 Transformer 的全局注意力与 GNN 的局部信息传递机制结合起来，在信息论压缩原则指导下实现可解释的生物标志物提取。

（3）理论上的原则性多模态集成

为了有效整合多模态数据，本研究提出了信息瓶颈异构图注意力网络 (IB-HGAN)，它扩展了压缩原理，通过基于元路径的群体图来指导多模态融合^[30]。IB-HGAN 建立了一种异构图学习的信息论方法，其中基于元路径的群体图和图同构测试确保了神经生物学上有效的特征交互。

此外，我们还引入了图同构语义测试机制来确保结构一致性，使模型能够适应性地规范跨模态交互，以保留来自成像和非成像数据的诊断相关信息。这种方法能够在诊断决策中明确归因于成像和非成像特征，提高模型的可解释性^[31]。

（4）可解释性分析与生物标志物识别

本研究特别关注模型的可解释性，开发了一套完整的可解释性分析方法，能够识别与 NDD 密切相关的脑功能网络 and 关键生物标志物。通过分析模型决策过程中不同脑区和连接的贡献程度^[32]，我们可以将机器学习发现与现有的神经科学知识联系起来，为理解 NDD 的神经机制提供新的视角。

这种可解释性分析不仅增强了模型的临床实用性，还为进一步的神经科学研究提供了有价值的假设和线索。通过明确归因于成像和非成像特征在诊断决策中的作用，我们的方法能够揭示 NDD 的多因素性质^[33]，支持更全面的疾病理解和个性化干预策略制定。

1.4 文章结构

本文的结构安排如下：

第一章 绪论：介绍神经发育障碍的研究背景及其社会意义，详细阐述传统诊断方法的局限性和计算机辅助诊断的必要性。系统回顾国内外研究历史与现状，从传统机器学习方法到深度学习技术的演进过程，并分析当前方法在多尺度特征提取、可解释性、多模态整合等方面存在的不足。明确提出本研究的四大创新点，包括多分辨率脑区划分与双通道特征融合、基于信息瓶颈理论的异构图学习、理论上的原则性多模态集成以及可解释性分析与生物标志物识别，为后续章节奠定基础。

第二章 全面介绍支撑本研究的核心理论与技术框架。首先详细阐述 fMRI 技术原理与 BOLD 信号的生理机制，解释其在脑功能连接研究中的关键作用。系统比较 AAL 和 Schaefer 等不同脑区划分图谱的特点及应用场景，为多分辨率策略提供理论依据。深入剖析图神经网络的基本原理、消息传递机制及典型模型，扩展至异构图理论与元路径概念。介绍 Transformer 模型的自注意力机制及其与图结构结合的优势和策略。这些理论基础相互支持，共同构建了一个多层次的理论框架

第三章 详细阐述多分辨率脑区划分策略的理论基础与实现方法，分析不同粒度脑区图谱(1000、400 和 100 个区域)的互补特性及其捕捉多尺度脑网络特征的优势。深入讲解双通道特征融合架构的设计理念与技术细节，重点说明 GNN-DN 模块如何通过度归一化策略有效建模局部邻域连接模式，以及 Transformer-FA 模块如何利用全连接注意力机制捕捉远距离脑区间依赖关系。分析这种双通道设计如何有效平衡局部结构与全局依赖信息，弥补单一方法的局限性。通过消融实验验证各组件对模型性能的贡献，并通过可视化分析关键脑区的作用机制，增强模型的可解释性。

第四章 系统阐述将信息瓶颈理论引入 NDD 异构图学习的理论基础和技术创新。详细介绍信息瓶颈图转换器 (IBGraphFormer) 的设计原理与实现细节，分析如何通过信息压缩原理从脑功能网络中提取最少但充分的生物标志物。深入讲解信息瓶颈异构图注意力网络 (IB-HGAN) 的工作机制，展示其如何通过基于元路径的群体图指导多模态融合，实现 fMRI 与非影像数据的有效整合。探讨图同构语义测试机制如何确保跨模态交互的结构一致性，使模型能明确归因于不同类型特征。通过理论分析和实验验证，证明该框架在保持高分类性能的同时，提供了丰富的可解释性。

第五章 总结与展望：总结本文的主要贡献和结论，讨论研究的局限性，并提出未来研究方向和改进策略。

第二章 相关理论基础

2.1 fMRI 与 BOLD 信号

2.1.1 fMRI 技术基本原理

功能性磁共振成像是一种非侵入性的神经影像学技术,可用于探测和测量与神经活动相关的脑部功能变化。fMRI 技术基于核磁共振 (Nuclear Magnetic Resonance, NMR) 物理原理^[34],通过检测氢原子核在强磁场中的行为来获取组织结构和功能信息。

在强磁场环境中,氢原子核(主要来自于水分子)会按照拉莫尔频率进行旋转,该频率与磁场强度成正比,表达为数学式 (2-1):

$$\omega_0 = \gamma B_0, \quad (2-1)$$

其中, ω_0 为拉莫尔频率, γ 为旋磁比(对氢原子核而言约为 42.58 MHz/T), B_0 为外部磁场强度。在临床和研究常用的 1.5T 至 7T 磁场强度下,氢原子核的拉莫尔频率约在 63.87MHz 至 298.06MHz 之间。

fMRI 的基本原理涉及射频 (Radio Frequency, RF) 脉冲的应用和梯度磁场的操作。当施加特定频率的 RF 脉冲时,氢原子核会吸收能量并从低能态跃迁至高能态,这个过程称为激发。当 RF 脉冲停止后,氢原子核会释放能量并返回平衡状态,这个过程称为弛豫。弛豫过程中释放的能量可以被接收线圈检测到,并转换为 MR 信号。

弛豫过程主要包含两种类型:纵向弛豫 (T1 弛豫):表征纵向磁化恢复到平衡状态的时间常数,与组织的分子环境相关。横向弛豫 (T2 弛豫):表征横向磁化衰减的时间常数,与局部磁场均匀性相关。

实际测量中,由于磁场不均匀性的影响,我们通常观察到的是 T_2^* 弛豫,其数学表达式为 (2-2):

$$1/T_2^* = 1/T_2 + 1/T_2', \quad (2-2)$$

其中, T_2' 表示由磁场不均匀性导致的额外信号衰减因素。

fMRI 技术正是利用了 T_2^* 加权成像来检测脑部氧合状态的变化,这些变化与神经

活动密切相关，构成了 BOLD (Blood Oxygen Level Dependent) 信号的基础^[35]。

fMRI 数据采集的基本流程如下图 2-1 所示：

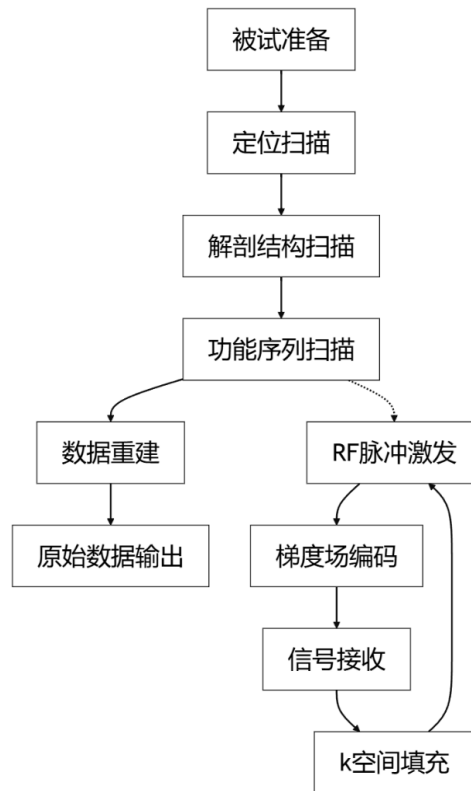


图 2-1 fMRI 数据采集的基本流程图

2.1.2 BOLD 信号的生理基础

BOLD 信号是 fMRI 技术中最广泛应用的对比机制，其生理基础建立在神经活动、能量代谢和血液动力学之间复杂的关系上^[36]。当特定脑区神经元活动增加时，会触发一系列生理反应，最终导致局部血氧水平变化，这一过程被称为神经血管耦合^[37]。

神经血管耦合过程包含以下关键环节：

- (1) 神经元活动增加，导致能量需求上升，主要表现为 ATP 消耗增加。
- (2) 为满足能量需求，局部代谢速率增加，氧气和葡萄糖消耗加速。
- (3) 局部血管扩张，脑血流量 (Cerebral Blood Flow, CBF) 增加。
- (4) 血流增加通常超过实际氧气需求，导致含氧血红蛋白相对于脱氧血红蛋白浓度增加。

这一过程可通过以下神经血管耦合模型图 2-2 表示：

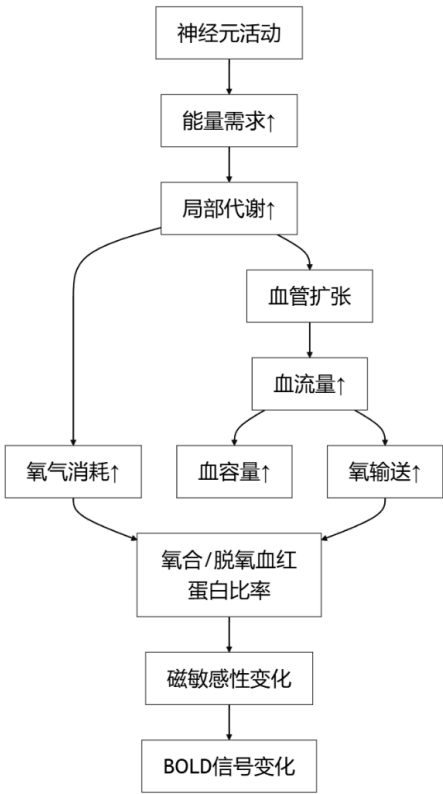


图 2-2 神经血管耦合模型图

BOLD 信号的物理基础在于氧合血红蛋白和脱氧血红蛋白的不同磁性。脱氧血红蛋白是顺磁性的，会导致局部磁场不均匀性增加，加速 T2 弛豫，从而降低 MR 信号强度。因此，当神经活动增加导致局部血流增加，脱氧血红蛋白浓度相对降低时，局部磁场均匀性提高，T2 延长，MR 信号增强，产生正 BOLD 响应。

神经元活动与 BOLD 信号之间的关系通常表现为血液动力学响应函数（Hemodynamic Response Function, HRF），其典型时间曲线包括，初始负相：活动开始后约 1-2 秒，由于氧气消耗暂时超过血供应而导致信号轻微下降。正峰值：活动后约 4-6 秒达到峰值，反映血流超量补偿。欠冲：活动结束后约 10-30 秒，信号低于基线，可能反映血容量和氧代谢的不同恢复速率。

2.2 AAL 图谱与 Schaefer 图谱

2.2.1 图谱基本概念与背景

在神经影像学研究中，脑图谱是理解大脑结构与功能的基础工具^[38]。脑图谱提供

了一个标准化的参考框架,使研究者能够将不同被试的脑影像数据归一化到同一空间中进行对比和分析。这些图谱通常基于多个健康被试的脑影像数据构建,通过复杂的配准算法将个体间的解剖差异最小化,从而形成一个代表性的平均脑。脑区分割是在脑图谱基础上,将大脑皮层和皮下结构划分为多个具有解剖学或功能意义的区域的过程^[39]。这一过程可基于不同的标准进行,如细胞构筑学特征、功能连接模式或多模态信息的整合。

标准化处理流程通常遵循以下步骤图 2-3:

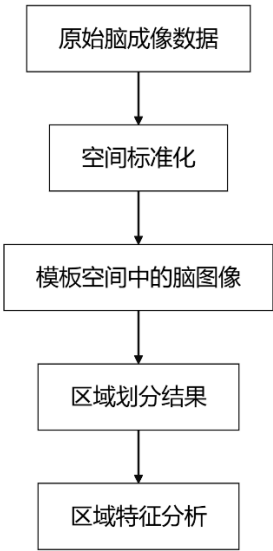


图 2-3 标准化处理流程图

在神经影像学研究背景下,脑图谱和脑区分割方案的选择对研究结果有显著影响。不同图谱间的解剖学差异可能导致 15-30%的体积测量变异^[40],这在神经发育障碍研究中尤为重要。

随着神经影像技术的发展,多种脑图谱和分割方案被提出,其中 AAL 图谱和 Schaefer 图谱是当前神经影像学研究中广泛应用的两种不同类型的分割方案,它们代表了解剖学驱动和功能驱动两种不同分割思路。

2.2.2 AAL 图谱

AAL (Automated Anatomical Labeling) 图谱^[41]基于单个年轻健康男性志愿者的 MRI 数据开发。该图谱将大脑皮层和皮下灰质结构分割为 116 个解剖区域 (包括 90 个大脑皮层区和 26 个小脑区域),主要基于解剖学标志进行划分。

AAL 图谱的构建始于对单个志愿者的高分辨率 T1 加权 MRI 扫描，通过手动描绘大脑沟回，将大脑划分为解剖学区域，随后为每个体素分配区域标签，最后转换到 MNI 标准空间形成最终图谱^[42]。这一流程如下图 2-4 所示：

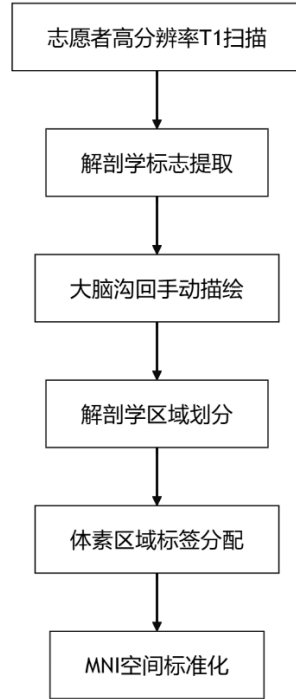


图 2-4 AAL 图谱构建过程图

AAL 图谱基于 MNI 空间（Montreal Neurological Institute）标准化坐标系，这使其能够与大多数功能和结构神经影像学研究兼容。AAL 图谱的解剖学标签通过将三维空间中的每个体素分配给特定的脑区来实现，数学式为（2-3）：

$$AAL(x, y, z) = R_i, \text{ for } (x, y, z) \in R^3, i \in 1, 2, \dots, 116, \quad (2-3)$$

其中， (x, y, z) 是 MNI 空间中的坐标， R_i 表示 116 个预定义区域中的一个。

AAL 图谱的主要优势在于其严格遵循经典神经解剖学定义的脑区边界，尤其是大脑沟回结构。基于解剖学标志的分割相对稳定，不受功能状态变化影响，这使其在多种神经影像研究中得到广泛验证和应用。

2.2.3 Schaefer 图谱

与 AAL 图谱不同，Schaefer 图谱^[43]代表了一种基于功能连接的脑区分割方案。该图谱利用大规模功能磁共振数据（来自 1489 名被试），通过梯度加权马尔可夫随机

场算法，将大脑皮层划分为具有相似功能连接模式的区域。Schaefer 图谱的构建过程如下图 2-5 所示：

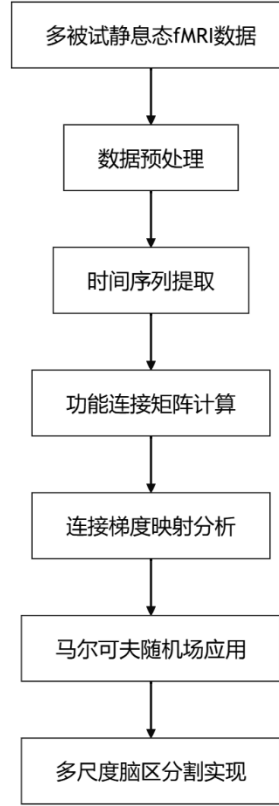


图 2-5 Schaefer 图谱构建过程图

Schaefer 图谱的关键创新在于提供了多尺度分割方案，包括 100、200、300、400、500、600、800 和 1000 个区域的版本，每个版本都保持与 Yeo 等人^[44]定义的 7 个功能网络的对应关系。这种多尺度特性使研究者能够根据具体研究问题选择适当的分割粒度^[45]。Schaefer 图谱构建的数学表达可概括为式 (2-4)：

$$P_{\text{Schaefer}} = \arg \min_P \sum_{i,j \in \Omega} w_{ij} \cdot \delta(P_i \neq P_j) + \sum_{i \in \Omega} g_i \cdot \delta(P_i \neq P_i^{\text{prior}}), \quad (2-4)$$

其中， P 表示分割方案， w_{ij} 代表体素 i 和 j 之间的功能相似性， g_i 是梯度信息， P_i^{prior} 是基于先验知识的分配， δ 是指示函数。

Schaefer 图谱的优势在于其基于功能连接模式而非解剖学边界构建，提供多种分辨率的分割方案，基于大样本数据构建具有良好群体代表性，并且与经典功能网络分类保持一致性^[46]。这些特性使其在功能连接组学、脑网络分析、个体化脑功能映射以及发育和精神疾病研究领域得到广泛应用^[47]。

然而, Schaefer 图谱也存在一些限制。功能连接模式可能受到功能状态(如静息态与任务态)的影响而变化^[48]; 功能边界在个体间存在显著变异; 功能连接模式可能随时间变化导致不稳定性; 且该图谱主要关注皮层, 对皮下结构的描述较为有限。

2.3 图神经网络及异构图理论

2.3.1 图神经网络及其基本原理

图数据在现实世界中无处不在, 如社交网络、化学分子结构、知识图谱等^[49]。然而, 传统的深度学习模型(如 CNN、RNN)在处理此类数据时面临显著挑战:

(1) 不规则性: 图数据缺乏规则的欧几里德结构, 节点的邻居数量不固定, 无法直接应用标准的卷积操作。

(2) 拓扑敏感性: 传统深度学习难以有效捕捉图的拓扑结构信息, 而这些信息对理解图数据至关重要。

(3) 表示困难: 图数据难以表示为固定大小的张量, 这使得传统深度学习模型无法直接处理。

(4) 关系建模: 传统模型难以同时建模节点特征和节点间的关系。

图神经网络正是为解决这些问题而设计的, 它能够直接在图结构上进行计算, 保留拓扑信息, 同时处理节点特征和结构信息, 通过消息传递机制学习节点间的依赖关系, 处理变长和不规则的输入数据^[50]。

图结构是一种描述实体之间关系的数据结构, 通常表示为式 (2-5):

$$G = (V, E), \quad (2-5)$$

其中 V 表示节点(顶点)集合, $E \subseteq V \times V$ 表示边集合。

在实际应用中, 图上的节点和边通常具有属性特征, 因此可以扩展表示为数学式 (2-6):

$$G = (V, E, X^V, X^E), \quad (2-6)$$

其中 X^V 和 X^E 分别表示节点特征矩阵和边特征矩阵。

图的邻接矩阵 $A \in \mathbb{R}^{n \times n}$ 是描述图拓扑结构的重要表示方式, 其中 $A_{ij} = 1$ 表示节点 i 和节点 j 之间存在边, 否则 $A_{ij} = 0$ 。对于带权图, A_{ij} 可以是任意实数, 表示边的

权重^[51]。

度矩阵 D 是一个对角矩阵，其中 $D_{ii} = \sum_j A_{ij}$ 表示节点 i 的度（连接的边数）。拉普拉斯矩阵定义为 $L = D - A$ ，归一化拉普拉斯矩阵为式（2-7）：

$$L_{norm} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}, \quad (2-7)$$

其中 I 为单位矩阵。这些矩阵在图信号处理和图神经网络中具有重要作用。

图神经网络是为处理图结构数据而设计的深度学习模型。不同于传统的神经网络处理欧几里德空间中的规则数据，GNN 专门设计用来处理具有复杂拓扑结构的非欧几里德空间数据^[52]。GNN 的核心思想是通过消息传递机制学习节点的表示，同时保留图的结构信息。GNN 的基本架构可通过以下流程图 2-6 表示：

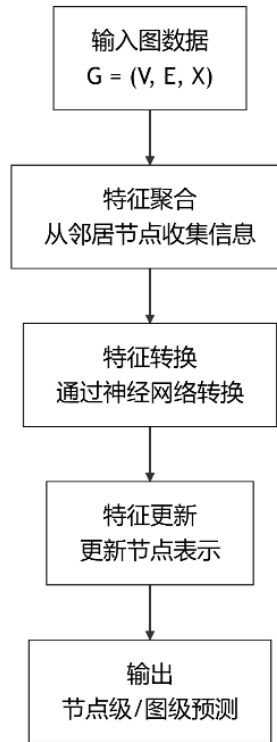


图 2-6 GNN 基本流程图

消息传递机制是 GNN 的核心操作，可概括为两个步骤：消息生成与聚合^[53]。在第 l 层，节点 v 的表示更新过程可表示为数学式（2-8）：

$$h_v^{(l+1)} = \text{UPDATE}^{(l)} \left(h_v^{(l)}, \text{AGGREGATE}^{(l)} \left(\{h_u^{(l)} : u \in \mathcal{N}(v)\} \right) \right), \quad (2-8)$$

其中 $h_v^{(l)}$ 是节点 v 在第 l 层的表示， $\mathcal{N}(v)$ 是节点 v 的邻居集合，AGGREGATE 函数聚合

来自邻居的信息，UPDATE 函数结合节点自身信息和聚合的邻居信息更新节点表示。

这一框架具有普遍性，不同的 GNN 变体主要区别在于它们如何实现 AGGREGATE 和 UPDATE 函数^[54]。消息传递实际上模拟了信息在图中的扩散过程，使得每个节点能够感知到其局部邻域的结构特征和属性信息。经过多层传递后，节点的表示能够包含越来越远的邻域信息，从而获得更全面的图结构感知能力。

2.3.2 异构图理论

现实世界中的大多数复杂系统不仅包含单一类型的实体和关系，而是由多种类型的实体和关系组成的异构系统^[55]。同构图虽然能够描述统一类型的实体间关系，但难以表达这种复杂性。因此，异构图作为图模型的重要扩展应运而生。

与同构图只包含单一类型的节点和边不同，异构图（Heterogeneous Graph）包含多类型的节点和边。异构图形式化定义如下式（2-9）：

$$G = (V, E, \mathcal{T}_V, \mathcal{T}_E, \phi, \psi), \quad (2-9)$$

其中， V 和 E 分别是节点集合和边集合， \mathcal{T}_V 和 \mathcal{T}_E 分别是节点类型集合和边类型集合， $\phi: V \rightarrow \mathcal{T}_V$ 是将节点映射到其类型的函数， $\psi: E \rightarrow \mathcal{T}_E$ 是将边映射到其类型的函数^[56]。

异构图的主要特性包括：多类型实体：包含不同类型的节点，如学术网络中的作者、论文、会议等；复杂关系：节点间的关系多样且含义不同，如作者发表论文、论文引用论文等；语义丰富：节点和边类型包含丰富的语义信息；结构复杂性：拓扑结构更为复杂，需要考虑不同类型节点间的交互。异构图的特殊性质使得处理它们需要特别设计的算法和模型。

异构图神经网络（Heterogeneous Graph Neural Networks, HGNNs）是为处理异构图而设计的 GNN 扩展^[57]。不同于同构图神经网络，HGNNs 必须考虑节点和边的类型差异，因此设计了专门的消息传递和聚合机制。对于异构图中的节点更新，一个基础的公式可表示为式（2-10）：

$$h_i^{(l+1)} = \sigma \left(\sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_i^r} \frac{1}{|\mathcal{N}_i^r|} W_r^{(l)} h_j^{(l)} + b_r^{(l)} \right), \quad (2-10)$$

其中， \mathcal{R} 是边类型集合， \mathcal{N}_i^r 是节点 i 在关系 r 下的邻居集合， $W_r^{(l)}$ 是关系 r 对应的权重矩阵， $b_r^{(l)}$ 是关系 r 对应的偏置项。

这种关系特定的参数设计使得模型能够区分不同类型关系的语义,适应异构图的复杂性^[58]。消息传递过程中,不同类型的边扮演着不同的角色,传递着不同的信息。例如,在学术网络中,“作者-撰写-论文”和“论文-引用-论文”这两种关系虽然都是连接节点的边,但它们传递的信息内涵完全不同。

异构图神经网络的一般架构如下图 2-7 所示:

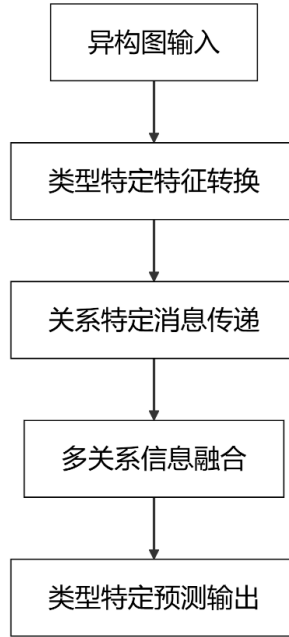


图 2-7 异构图神经网络流程图

异构图注意力网络 (Heterogeneous Graph Attention Network, HAN) 是异构图神经网络的主要模型之一^[59], HAN 利用节点级和语义级的双重注意力机制处理异构图。首先,针对每个元路径,使用节点级注意力聚合邻居信息,如下数学式 (2-11):

$$z_i^\Phi = \sigma \left(\sum_{j \in \mathcal{N}_i^\Phi} \alpha_{ij}^\Phi W^\Phi h_j \right), \quad (2-11)$$

其中 Φ 表示元路径, \mathcal{N}_i^Φ 是基于元路径 Φ 的节点 i 的邻居, α_{ij}^Φ 是注意力系数,然后,使用语义级注意力融合不同元路径的结果,如下数学式 (2-12):

$$z_i = \sum_{\Phi \in P} \beta_\Phi \cdot z_i^\Phi, \quad (2-12)$$

其中 β_Φ 是元路径 Φ 的重要性权重,通过注意力机制学习得到。

2.4 Transformer 模块与图结构结合

2.4.1 Transformer 模块基本概念

Transformer 模型^[60]改变了自然语言处理的架构设计范式。在 Transformer 之前，循环神经网络（RNN）及其变体如长短期记忆网络（LSTM）和门控循环单元（GRU）是序列数据处理的主流方法。然而，这些模型存在三个关键问题：序列计算限制：无法并行处理，导致训练效率低下；长序列依赖问题：难以捕捉远距离依赖关系；信息损失：长序列信息在传递过程中可能会逐渐衰减。

Transformer 通过全新的注意力机制和并行计算范式解决了上述问题，使得模型能够高效地处理长序列数据，并有效捕捉全局依赖关系^[61]。Transformer 的核心架构由编码器（Encoder）和解码器（Decoder）组成，采用了完全基于注意力机制的设计，摒弃了循环结构。其基本架构如下图 2-8 所示：

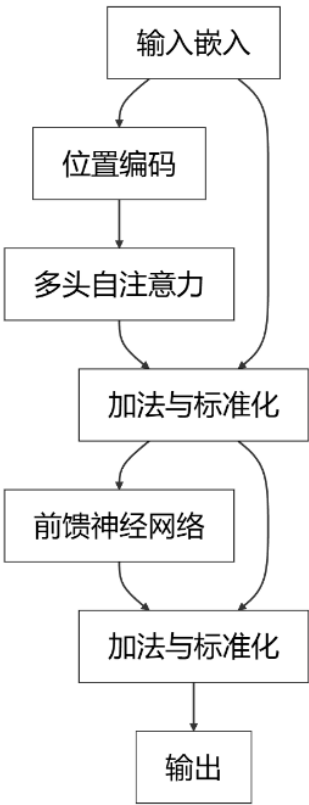


图 2-8 Transformer 基本架构图

自注意力机制是 Transformer 最关键的创新，它允许模型在处理序列中任意位置

时考虑整个序列的信息^[62]。其基本思想是计算序列中每个元素与所有元素(包括自身)的相关性,然后基于这种相关性进行加权求和。给定一个序列的输入嵌入矩阵 $X \in \mathbb{R}^{n \times d}$,其中 n 是序列长度, d 是嵌入维度,自注意力机制通过以下步骤计算:首先,通过三个不同的权重矩阵 $W^Q, W^K, W^V \in \mathbb{R}^{d \times d_k}$ 将输入转换为查询(Query)、键(Key)和值(Value),见数学式(2-13):

$$Q = XW^Q, K = XW^K, V = XW^V. \quad (2-13)$$

接着,计算注意力权重,即查询和键之间的相似性,如下数学式(2-14):

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right), \quad (2-14)$$

其中 $\sqrt{d_k}$ 是缩放因子,用于稳定梯度。再利用注意力权重对值进行加权求和,得到自注意力输出,而多头注意力机制是通过并行计算多组不同的自注意力,然后将结果拼接,见数学式(2-15):

$$\text{MultiHead}(X) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W^O, \quad (2-15)$$

其中每个头(head)的计算见数学式(2-16):

$$\text{head}_i = \text{Attention}(XW_i^Q, XW_i^K, XW_i^V), \quad (2-16)$$

其中 W^O 是输出投影矩阵, h 是头的数量。

自注意力机制在捕捉全局依赖关系方面具有显著优势,直接长距离依赖:每个位置可以直接与任何其他位置交互,计算路径长度为1;并行计算:所有位置的计算可同时进行,大幅提高训练效率;自适应关注:通过学习注意力权重,模型能够自适应地关注序列中相关的部分;丰富的特征表示:多头机制允许关注不同子空间的信息,增强表达能力^[63]。

2.4.2 Transformer 与图神经网络的融合

传统图神经网络在表示学习方面表现出一系列内在的结构性限制,这些限制严重影响了其在复杂图数据处理中的应用效能。首要的挑战源于其固有的局部感知机制,即每个计算层仅能聚合与处理直接邻接节点的信息,这种设计导致模型需要通过堆叠多层网络才能实现远距离节点间关系的有效捕捉^[64]。与此同时,随着网络深度的增加,

GNN 不可避免地面临表示过平滑的问题，即节点表征在迭代消息传递过程中逐渐趋同，最终导致表示特征的区分度严重下降。此外，传统 GNN 在表达能力上也存在明显局限，难以有效建模和表达复杂的高阶关系以及全局依赖结构。更为根本的是，这类模型在设计哲学上过度依赖于显式图拓扑结构，这种严格的结构约束使其在处理隐含关系或动态演化的图结构时表现不佳^[65]。

鉴于上述局限性，将 Transformer 架构引入图神经网络领域成为一种颇具前景的研究方向。Transformer 的核心优势在于其自注意力机制能够突破拓扑结构的限制，直接建立节点间的全局依赖关系，无需依赖多层迭代即可实现远距离信息交互^[66]。这种全局注意力机制同时有效缓解了过平滑问题，即使在深层网络中仍能保持节点表示的独特性与区分性。多头注意力的设计进一步增强了模型的表达能力，使其能够从多个语义空间同时学习节点关系，超越了传统图结构所能提供的表达限制。Transformer 在图领域的另一显著优势体现在其自适应性上，注意力权重可以根据输入数据动态调整，从而更好地适应图结构的时序变化和演化特性。最后，这种结构设计还赋予了模型隐式关系建模的能力，能够有效识别和表征那些在原始图拓扑中并未显式表达但客观存在的潜在节点关系，这对于复杂系统的建模与理解具有重要意义。

在神经发育障碍诊断的特定场景下，这种基于 Transformer 的图学习范式展现出独特优势。传统图神经网络在处理异构医疗数据时往往依赖于预定义的图结构，难以充分捕捉不同模态数据间的复杂交互。而全局注意力机制通过学习数据驱动的关联模式，能够自动发现临床相关的跨模态特征组合，提高诊断准确性。实验结果表明，在多层注意力结构中，第一层通常倾向于学习解剖学相邻区域间的局部关联，而更深层则逐渐发展出功能连接模式，这与神经科学中的层级信息处理理论高度一致。

值得注意的是，这种注意力机制不仅提升了模型性能，还增强了结果的可解释性。通过分析注意力权重矩阵，研究人员可以识别出对诊断决策具有关键影响的脑区组合，为临床解释提供依据。这种解释即设计的理念将可解释性作为模型架构的内在特性，而非事后分析的附加功能，从而在保证高精度诊断的同时满足医疗应用对透明度的严格要求。

结合 GNN 和 Transformer 的混合架构可以设计为多种形式，当前用的是分层融合架构，具体见图 2-9 所示：

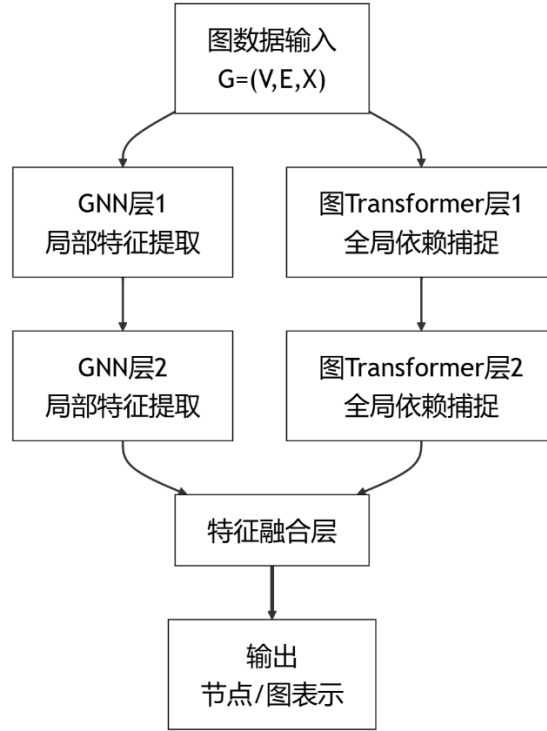


图 2-9 融合架构图

在这种架构中，GNN 分支负责捕捉局部结构特征，Transformer 分支负责建模全局依赖关系，最后通过融合层将两种特征整合起来。融合可以通过简单的拼接、加权和或注意力机制实现，见数学式（2-17）：

$$H_{\text{fusion}} = \text{Fusion}(H_{\text{GNN}}, H_{\text{Transformer}}), \quad (2-17)$$

在本研究中，*Fusion*用的是线性融合，如数学式（2-18）所示：

$$H_{\text{fusion}} = W_1 H_{\text{GNN}} + W_2 H_{\text{Transformer}}. \quad (2-18)$$

Transformer 与图神经网络的结合为图结构数据的处理提供了新的范式，克服了传统 GNN 难以捕捉全局依赖的局限性，同时保留了对局部图结构的感知能力。这种结合在理论上具有更强的表达能力，能够同时建模局部和全局关系，特别适合处理复杂的网络数据。

2.5 本章小结

本章系统阐述了支撑本研究的核心理论与技术基础，为后续研究方法的开发与实现奠定了坚实的理论框架。首先，对功能性磁共振成像技术及其产生的血氧水平依赖信号进行了详细介绍。作为一种无创的脑功能成像技术，fMRI 通过检测神经活动引

起的局部血流动力学变化，提供了研究大脑功能连接的重要手段。BOLD 信号的生理机制解释了神经元活动与血氧水平变化之间的关系，为理解功能连接数据的生物学意义提供了基础。

随后，本章阐释了脑区划分图谱在神经影像分析中的关键作用。AAL 图谱和 Schaefer 图谱代表了两种不同的脑区划分方法，前者基于解剖学划分，后者基于功能连接特征进行分区。两种图谱各具特点，为本研究中采用的多分辨率脑区划分策略提供了理论依据，使我们能够从不同尺度和视角捕捉脑网络的组织结构。

图神经网络及异构图理论部分详细介绍了图数据的数学表示、GNN 的基本架构与消息传递机制，以及典型的 GNN 模型如 GCN、GAT 等。更进一步，异构图理论的阐述为处理包含多类型节点和边的复杂网络提供了理论框架，包括元路径、元图等重要概念，为整合多模态神经影像和非影像数据奠定了基础。

最后，Transformer 模块与图结构结合部分分析了 Transformer 的自注意力机制及其在捕捉全局依赖关系方面的优势，讨论了将 Transformer 与 GNN 结合的动机和策略。这种结合能够克服传统 GNN 在建模长距离依赖关系方面的局限性，同时保留对局部图结构的感知能力，为本研究中提出的双通道特征融合架构提供了理论支撑。

综上所述，这些理论基础相互关联、相互支持，共同构建了一个多层次的理论框架，为本研究中开发的多分辨率脑区划分与双通道特征融合模型以及基于信息瓶颈的异构图学习框架提供了充分的理论依据。

第三章 基于多分辨率脑区划分与双通道特征融合的自闭症谱系障碍识别

首先提出了一种基于 GNN-DN(Graph Neural Network with Degree-Normalized) 和 Transformer-FA(Full Attention Transformer) 模块的多分辨率脑区划分模型，用于识别自闭症。通过利用 GNN-DN 模块的度归一化和 Transformer-FA 模块的全连接注意力机制，该模型有效地整合了脑功能网络中的局部和全局特征。此外，该模型成功识别了与自闭症相关的关键脑网络，与现有的医学研究保持一致并支持其生物学可解释性。

3.1 多分辨率脑区域分割

本研究采用 Schaefer2018 图谱将大脑分割为 1000、400 和 100 个感兴趣区域 (ROI)，分别对应三个分辨率级别：微观、中观和宏观。该过程如图 3-1 所示。

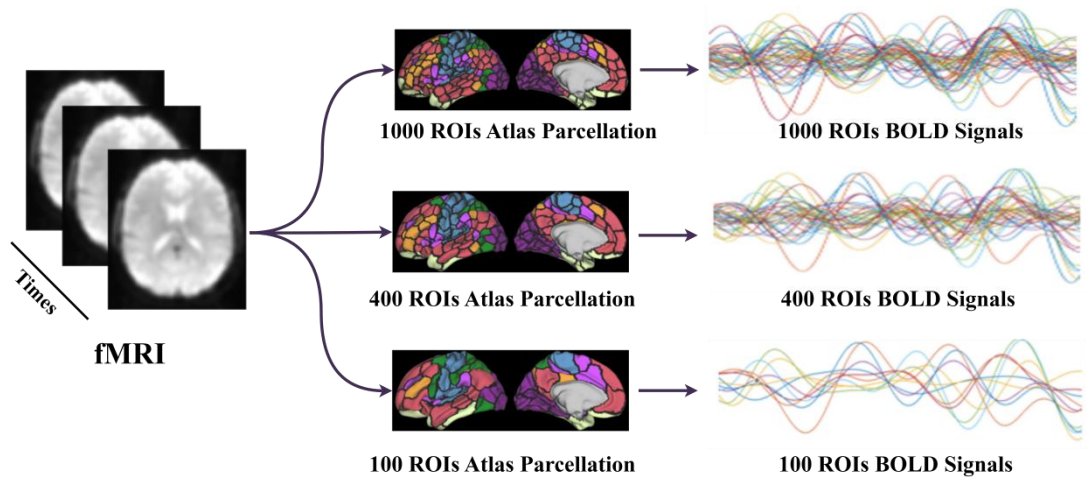


图 3-1 多层脑区划分

在多尺度层面上对大脑进行划分有助于从不同分辨率提取大脑功能信息：

(1) 微观层面（1000 个 ROI）：侧重捕捉细粒度的脑区活动与连接模式，能够呈现更精细的功能差异。然而，过高的分辨率也意味着更大的数据维度和更高的计算开销。这一层面的分析能够检测到功能连接的局部异常和细微变化，特别适用于早期神经发育障碍的诊断，其精细的空间分辨率有助于定位特定亚型的神经回路改变。

(2) 中观层面 (400 个 ROI): 相比微观层面, 中观层面在分辨率与计算效率之间达到了较好的折中, 既能保留一定的细节信息, 又不会在计算量方面过度消耗。中观尺度特别适合检测功能网络内部的组织结构变化和模块性调整, 既能保留足够的生物学信息, 又能有效控制特征维度带来的过拟合风险。

(3) 宏观层面 (100 个 ROI): 提供了较粗粒度的全局视角, 更有利于快速洞察脑网络整体布局和大规模群体分析, 能够凸显出主要功能网络之间的宏观交互模式。宏观尺度分析优势在于其稳健性和较低的噪声敏感性, 特别适合分析发育过程中的长期轨迹变化和大样本流行病学研究

通过在微观、中观和宏观三个层面分别进行分析并互相对比, 可以更全面地理解大脑功能活动的多尺度特征: 微观层面可能揭示精细的局部功能差异, 中观层面更适合在保留适度细节的同时兼顾计算可行性, 宏观层面则能够体现大脑整体功能网络的全局布局。

多尺度分析框架遵循了神经科学中广泛接受的层级整合原理, 即大脑信息处理通过不同空间尺度的神经元集合体进行逐级整合。值得注意的是, 不同尺度的信息并非简单冗余, 而是相互补充的。某些神经发育障碍在不同尺度上表现出不同的异常模式: 微观尺度可能显示局部连接增强, 而宏观尺度则表现为网络间通信障碍。

在获得不同分辨率的 ROI 时间序列后, 接下来构建功能连接图, 以揭示各脑功能网络之间的相互作用。首先, 基于广泛应用的 Yeo 17 网络划分, 从每个功能网络内提取所有 ROI 的时间序列, 并将这些时间序列按垂直方向连接成矩阵, 从而构成了各功能网络整体活动模式的表达。该方法不仅保留了不同脑区在各自功能网络内的局部动态信息, 也为后续跨网络相似性分析提供了基础数据。

随后, 为了量化不同功能网络之间的相似程度, 我们采用了 RV 系数作为衡量指标。RV 系数通过比较两个矩阵的协方差结构, 能够有效反映两个功能网络在时间序列活动模式上的一致性, 数学式 (3-1) 为:

$$RV(A, B) = \frac{tr((A^T A)(B^T B))}{\sqrt{tr((A^T A)^2) \cdot tr((B^T B)^2)}}, \quad (3-1)$$

其中分子部分表示矩阵A和B的协方差矩阵间的相似性, 分母部分是归一化因子, 确保RV系数的值在[0,1]范围内。通过这种标准化方法, 不同网络间的相似性可以在相同尺

度上进行直接比较，从而更准确地揭示各功能网络之间的关联强度。

最终，通过对所有功能网络两两计算 RV 系数，构建出一个 17×17 的相关性矩阵，其中每个矩阵元素均代表了相应两个功能网络之间的相似性强度。基于该相关性矩阵，进一步构建功能网络图结构，将每个功能网络视为图中的一个节点，并通过预先设定的相似性阈值，仅保留高于该阈值的边，从而仅展示显著的功能连接。如下，图 3-2 则直观地展示了这一过程的最终输出结果，为后续的脑网络拓扑分析及其与临床特征关联研究提供了坚实的数据基础和理论支持。

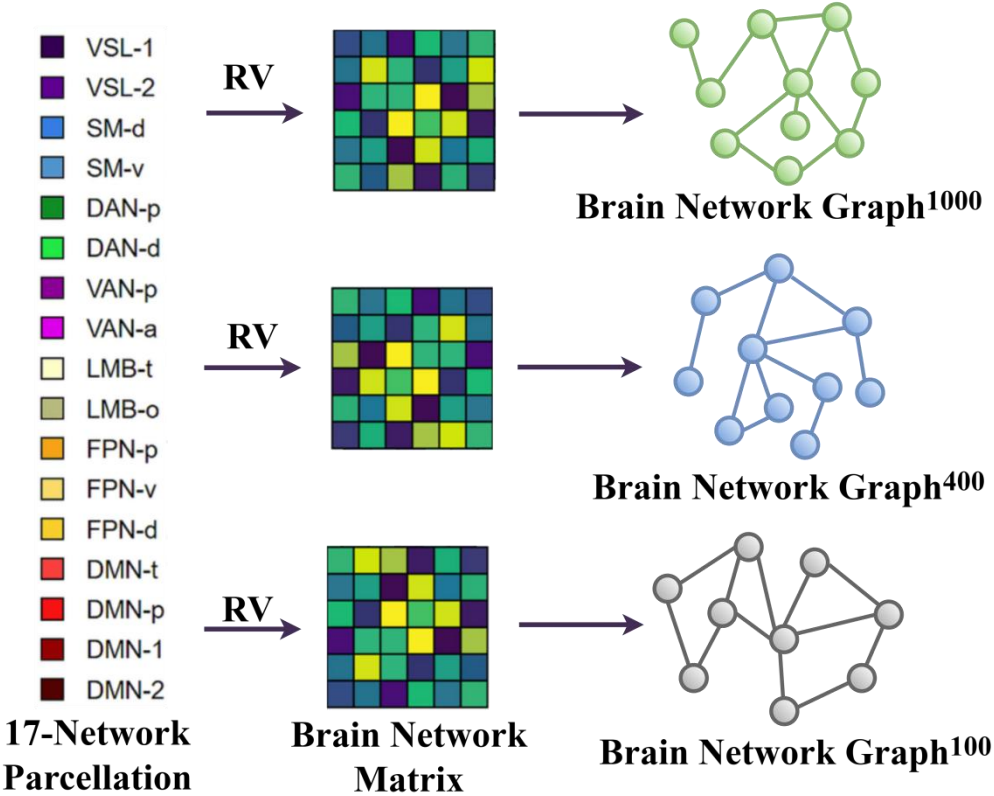


图 3-2 构建功能网络的图结构

3.2 双通道特征融合模块

为了充分挖掘大脑功能网络中的局部与全局信息，设计了一个双通道特征融合模块。该模块由两部分组成：一方面使用 GCN-DN 模块 以强化局部节点特征的提取，另一方面利用 Transformer-FA 模块 捕捉跨脑区的全局关联。整体流程如图 3-3 所示。

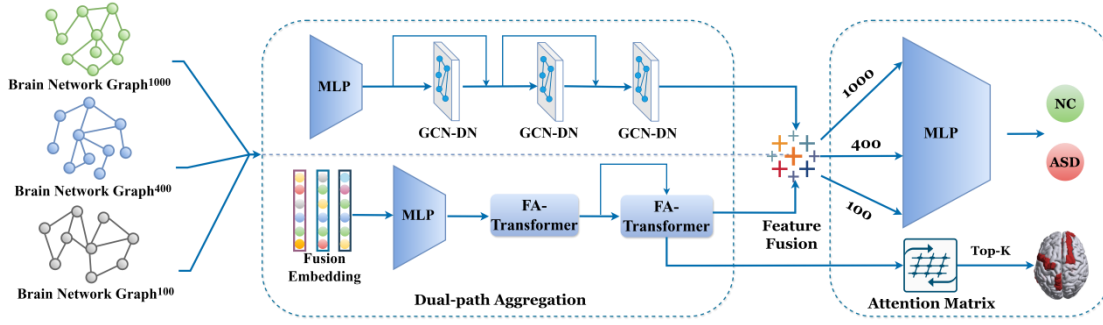


图 3-3 双通道特征融合模块

构建完成的功能网络的图结构，分别进入两个通道进行特征提取，最后在特征融合层中将局部和全局信息有效结合，用于后续分类任务。以下将对两个模块的原理和实现进行详细介绍。

3.2.1 GCN-DN 模块

GNN-DN 模块是本研究设计的用于处理大脑功能网络的核心组件之一。该模块在传统图神经网络的基础上，引入了度归一化和自适应残差连接，使模型在处理脑网络中的异质性节点时能够更加有效地进行特征传递。具体而言，GNN-DN 模块的创新点主要体现在以下几个方面：

（1）度归一化的加权机制

大脑功能网络中的节点（脑区）度分布存在显著差异。传统图卷积方法多采用拉普拉斯正则化，但未能充分考虑节点度的异质性，可能导致高低度节点之间的特征传播不均。GNN-DN 模块在特征传递中引入基于节点度的自适应加权机制，确保特征传播的平衡性。具体而言，对于节点*i*和*j*之间的边，度归一化权重 W_{ij} 定义如下（3-2）：

$$W_{ij} = \frac{1}{\sqrt{d_i d_j}}, \quad (3-2)$$

其中， d_i 和 d_j 分别表示节点*i*和节点*j*的度。基于此加权机制，特征传递的数学公式如下（3-3）：

$$H^{(l+1)} = \sigma(D^{-1/2} A D^{-1/2} H^{(l)} W^{(l)}), \quad (3-3)$$

其中， A 是邻接矩阵， D 是度矩阵，其对角元素 $D_i = d_i$ ， H 表示第*l*层的节点特征， $W^{(l)}$

是当前层的权重矩阵， σ 是非线性激活函数。该度归一化机制确保了特征在高低度节点之间的传播一致性。

(2) 自适应残差连接

为进一步优化特征传递效果并提高模型训练的稳定性，GNN-DN 模块引入了自适应残差连接机制。该机制通过在每一层卷积操作后，将初始输入特征 $H^{(0)}$ 直接叠加到当前层的输出中，从而有效缓解了多层网络中梯度消失的问题。具体而言，在每一层的特征更新过程中，模型不仅依赖于本层卷积获得的新特征，还保留了原始输入特征的信息，使得模型在深层结构中依然能够保持较高的信息传递效率。特征更新公式为 (3-4)：

$$H^{(l+1)} = H^{(l+1)} + H^{(0)}, \quad (3-4)$$

其中， $H(0)$ 是输入层的初始特征。

通过这种自适应残差连接方式，模型在反向传播过程中能够更有效地传递梯度，防止深层网络中梯度逐渐衰减，进而提升整体训练效果和收敛速度。此外，该机制还增强了特征表达的鲁棒性，使得网络能够更好地捕捉原始输入中的关键信息，从而提高了对复杂数据模式的建模能力。这一设计不仅借鉴了 ResNet 等经典深度学习模型中的残差思想，同时结合了图神经网络的特点，为 GNN-DN 模块在处理高维、多尺度特征时提供了坚实的理论支持和实践基础。

3.2.2 Transformer-FA 模块

Transformer-FA 模块用于捕捉大脑功能网络中跨区域的全局特征关联。该模块结合了全连接注意力机制和基于注意力的脑网络选择方法。

(1) 全连接注意力卷积

大脑功能网络的节点之间存在跨区域依赖性，远距离脑区之间的互动在分类中具有重要价值。Transformer-FA 模块通过引入全连接注意力机制，使每个节点与所有其他节点进行全局信息交互，全面捕捉跨区域的特征。

给定查询、键和值的输入特征向量，分别记为 $Q \in \mathbb{R}^{N \times H \times M}$ 、 $K \in \mathbb{R}^{L \times H \times M}$ 和 $V \in \mathbb{R}^{L \times H \times D}$ ，其中： N 为查询节点的数量， L 为源节点的数量， H 为注意力头的数量， M 和 D 分别为特征维度。

首先对查询 Q 和键 K 进行L2归一化，数学式(3-5)如下，以确保特征的尺度一致：

$$Q = \frac{Q}{\|Q\|_2}, K = \frac{K}{\|K\|_2}. \quad (3-5)$$

接下来，通过键和值的乘积计算源节点的加权和，得到一个中间聚合张量 KV_{agg} ，其计算公式(3-6)为：

$$KV_{agg} = \sum_{l=1}^L K_l \otimes V_l, \quad (3-6)$$

其中，操作符 \otimes 表示张量按维度 l 的内积运算。此步骤生成的 KV_{agg} 捕捉了源节点的加权特征信息。

然后，将查询向量 Q 与中间聚合张量 KV_{agg} 相乘，得到注意力加权的聚合结果 A_{num} ，数学式(3-7)：

$$A_{num} = \sum_{h=1}^H Q_n \cdot KV_{agg}. \quad (3-7)$$

该结果 A_{num} 表示每个查询节点与源节点之间的相似性权重，用于进一步的特征聚合。

为了对注意力权重进行归一化，定义一个归一化因子 A_{denom} 。首先，计算键向量的和 K_{sum} ，然后将查询向量与该和相乘，得到归一化因子(3-8)：

$$A_{denom} = \sum_{h=1}^H Q_n \cdot K_{sum}. \quad (3-8)$$

将注意力数值 A_{num} 除以归一化因子 A_{denom} 后，得到加权的特征聚合输出 A_{out} 。表示每个查询节点在不同头上加权后的特征表示。这个输出将与查询节点相关性最大的源节点特征聚合在一起，强化了与查询节点相似性最高的脑区。

(2) 基于注意力的脑网络选择

在 Transformer-FA 模块中，注意力机制使每个查询节点能够与所有源节点进行信息交互，但并非所有交互都对最终的全局特征聚合有重要贡献。为此，采用了 Top-k 策略，仅保留每个查询节点与其最相关的 k 个源节点对应的注意力权重，从而突出对最终特征聚合贡献最大的连接。

输入特征经过线性变换后分别得到查询矩阵 $Q \in \mathbb{R}^{N \times d}$ 和键矩阵 $K \in \mathbb{R}^{N \times d}$ ，其中 N

为节点数, d 为特征维度。注意力矩阵 A 的计算公式为 (3-9):

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right), \quad (3-9)$$

其中, softmax 操作是逐行计算的, 确保每个查询节点对应的权重和为 1。

接下来, 对于每个查询节点 i , 我们在其对应的注意力向量 A_i 中选择出 Top-k 的最大值, 并构造一个二值掩码矩阵 $M \in \{0,1\}^{N \times N}$, 然后, 通过将注意力矩阵与掩码矩阵逐元素相乘, 得到筛选后的注意力矩阵 (3-10):

$$A^{(\text{top-k})} = M \odot A, \quad (3-10)$$

其中 \odot 表示逐元素乘积操作。

为了进一步量化每个源节点在全局特征聚合中的重要性, 我们对 $A^{(\text{top-k})}$ 的每一列求和, 得到每个节点的全局重要性评分 I , 具体如数学式 (3-11):

$$I_j = \sum_{i=1}^N A_{ij}^{(\text{top-k})}. \quad (3-11)$$

得分 I_j 越高, 表明该源节点与大部分查询节点之间存在较强的相似性, 其在后续特征聚合中的贡献也越突出。

这种基于注意力和 Top-k 策略的筛选方法, 不仅能够有效地减少低相关性节点的干扰, 还能确保模型聚焦于关键连接, 从而提升全局信息提取的效率和诊断相关脑区特征的准确性。

3.3 基于多分辨率脑区与双通道特征融合的自闭症谱系障碍诊断实验

3.3.1 数据预处理

本章研究使用了 Autism Brain Imaging Data Exchange (ABIDE) 数据集的 fMRI 数据, 包含 ASD 患者和健康对照组 (HC)。由于数据来自不同机构, 部分数据存在图像伪影和头动干扰, 需在使用前进行筛选。本研究排除了严重运动伪影的样本, 最终保留了 871 名被试数据。

对于这 871 个受试者, 应用了标准化的预处理流程, 以确保数据的一致性和可比性, 并最大程度地减少伪影和噪声的影响。该预处理流程包括了以下几个关键步骤:

时间切片校正、头部运动校正、空间标准化和时间平滑。

(1) 时间切片校正：由于 fMRI 数据是通过连续切片采集的，不同切片的采集时间可能存在微小差异，这可能会导致时序数据的不一致性，进而影响后续分析。为了解决这个问题，采用插值方法对不同切片的采集时间进行对齐。通过这种方式，所有切片的采集时间被标准化，从而消除了时间上的差异。这一校正过程确保了数据中每个时间点的每个体素都具有相同的时间戳，为后续的分析提供了更加精确的时序数据。

(2) 头部运动校正：在进行 fMRI 扫描时，受试者的头部运动可能会导致图像的空间偏差，进而影响图像质量。为了减少运动伪影的影响，在预处理过程中，计算每个时间点的平移和旋转信息，采用逆变换方法对每一帧图像进行校正。这一步骤能够有效地减少由于头部运动引起的空间偏差，避免了运动伪影对后续分析的干扰。

(3) 空间标准化：为了确保不同受试者之间的数据具有可比性，需要将每个受试者的脑部影像数据转换到统一的空间坐标系。为此，所有图像数据都被注册到标准的 MNI152 空间，这是一种常用的标准大脑图谱。空间标准化不仅保证了不同受试者的大脑结构对齐，而且为不同数据集之间的比较提供了基础。

(4) 时间平滑：fMRI 数据通常包含大量的噪声，例如设备噪声和生理噪声。为了减少这些噪声的影响并提高信号的质量，应用了时间平滑技术。在这一过程中，使用高斯核对每个体素的时间序列进行平滑处理，从而降低短期波动对信号的影响。这不仅可以增强信号的稳定性，还能提高信号与噪声之间的比率，为后续的神经科学分析提供更加可靠的数据。

这些预处理步骤共同作用，确保了数据的一致性、准确性和质量。通过这些处理，原始 fMRI 数据中的伪影和噪声得到了有效减少，脑区之间的功能连接可以更加真实地反映出来，从而为后续的分析奠定了坚实的基础。

3.3.2 实验设置和评估指标

在实验设置中，我们将 871 名被试的数据随机划分为训练集、验证集和测试集。具体而言，训练集、验证集和测试集的划分比例分别为 70%、15% 和 15%。

在模型训练与评估过程中，我们选用了分类任务中常用的多项性能评估指标。Acc 反映了整体分类的准确性，Prec 和 Rec 可以衡量模型在正类上的精确性和识别能

力，而 AUC 则提供了一个整体的区分能力度量。结合这些指标可以更全面地评估模型在 ASD 分类任务中的有效性。

3.3.3 分类性能对比实验及消融实验

（1）模块性能对比

为验证 GNN-DN 和 Transformer-FA 模块的有效性，我们分别进行了多组实验，使用的是 ABIDE 数据集和 Schaefer2018 图谱。实验结果如表 3-1 所示。可以看出，本研究提出的组合模型在各项指标上均优于其他基线模型，特别是在 Rec 和 AUC 上表现突出，表明该模块在 ASD 患者识别方面具有更强的判别能力。

表 3-1 在 ABIDE 数据集上的实验结果

Method	Acc	Prec	Rec	AUC
SVM	65.72±3.11	54.68±6.88	66.38±6.79	71.33±4.56
LSTM	66.45±3.25	65.35±8.92	66.71±7.52	70.15±3.55
SwinT	68.45±4.64	60.32±7.58	68.36±6.78	70.26±4.26
BrainGNN	73.86±0.57	68.30±0.92	73.89±1.37	69.94±0.31
Ours	78.96±0.13	75.46±0.14	78.34±0.51	74.56±0.23

（2）消融实验

为进一步评估各模块和分层脑区划分的独立贡献，进行了消融实验，分别去掉 GNN-DN 模块、Transformer-FA 模块以及分层脑区划分。结果如表 3-2 所示。

表 3-2 消融实验结果

A	B	C	Acc	Prec	Rec	AUC
√	×	√	74.72±1.13	68.48±1.85	73.18±2.58	69.23±1.56
√	√	×	75.45±2.25	70.35±1.52	74.71±1.02	73.65±0.55
×	√	√	72.45±3.54	65.32±5.68	71.53±3.28	70.86±2.86

表中标记为 A、B 和 C 的列分别代表以下模型组件：A：多层次划分；B：GNN-DN 模块；C：Transformer-FA 模块。

从消融实验的结果可以看出，各模块对整体模型性能均有显著贡献，多层次脑区划分（A）：去除该策略后，各项指标明显下降，表明多尺度脑区划分能够更全面地捕

捉大脑功能信息，并显著提升模型对 ASD 相关异常的敏感性。

GNN-DN 模块 (B): 比较第一行与其他组合，缺失 GNN-DN 模块会使准确率和召回率有所下降。该模块通过自适应残差连接与图卷积操作，有效增强了特征传递与信息整合能力，从而改善了对复杂脑网络数据的表征。GNN-DN 的优势主要来源于其创新的深度归一化机制，该机制有效缓解了图神经网络中的过平滑问题。

Transformer-FA 模块 (C): 从第二行数据可见，剔除 Transformer-FA 模块后模型的 AUC 明显降低，这充分证明了其在捕捉全局注意力特征和实现跨脑区信息交互中的关键作用，进一步提升了模型的判别力。

总消融实验验证了各模块及多层次划分策略在构建高性能 ASD 识别模型中的互补作用，并为后续模型优化和神经机制探讨提供了坚实的理论依据。多层次划分策略能够适应不同成像设备和采集参数下的数据，增强了模型在多中心研究中的适用性；GNN-DN 模块提高了模型对低质量数据的鲁棒性，解决了儿童患者成像质量不稳定的实际挑战；Transformer-FA 模块通过提供可视化的注意力权重分布，增强了诊断结果的可解释性

3.3.4 脑网络可解释性

在本研究中，通过 Transformer-FA 模块中的注意力矩阵来分析大脑功能网络间的相关性，并进一步探讨模型的可解释性。

具体而言，模型在对所有脑区节点进行注意力权重计算后，能够识别出与 ASD 识别最密切相关的网络连接。基于全连接注意力机制所生成的注意力矩阵，我们在所有网络节点中筛选出了注意力权重最高的前六个脑网络，并对这些网络节点之间的交互进行了可视化，以揭示它们在 ASD 识别任务中的重要性，具体可视化如下图 3-4 所示。

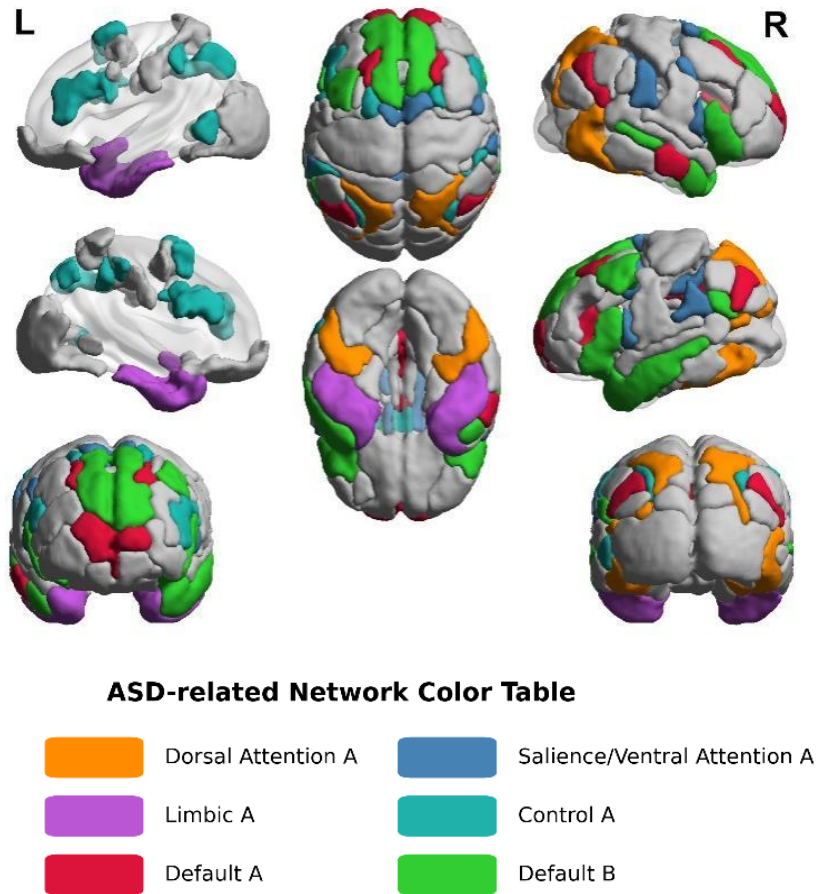


图 3-4 ASD 分类的六个最重要的大脑网络

在所选的六大网络中，识别了与注意、情感调节、社会认知和自我参照加工相关的脑区或亚网络，这些网络在 ASD 识别中均显示出较高的注意力权重，具体体现在以下几个方面：

（1）背侧注意网络（Dorsal Attention Network）

涉及顶叶和额叶关键区域，主要负责目标刺激的定向和选择性注意^[67]。ASD 患者在注意力保持和转换上常表现异常，其高权重可能反映了对外部信息处理中的神经机制失调。

（2）显著性/腹侧注意网络（Salience/Ventral Attention Network）

主要检测内外部的突发或显著性事件，并调控注意力的切换^[68]。该网络异常可能与 ASD 患者在情绪加工和社交线索识别方面的困难有关，为情绪反应和环境适应提供了神经标志。

（3）边缘网络（Limbic Network）

包括杏仁核、海马等结构, 关键在于情绪调节、记忆和动机^[69]。高权重提示其在 ASD 患者的情绪识别与社会情感交互中可能存在功能失调, 这与临床观察一致。

(4) 控制网络 (Control Network)

涉及额叶和顶叶区域, 关联执行功能和认知控制^[70]。该网络在 ASD 患者中可能反映为执行功能障碍或行为刻板性, 其高注意力权重支持了模型在识别认知调控缺陷方面的有效性。

(5) 默认模式网络 (Default Mode Network, Default A 与 Default B)

与自我参照、内省和社会认知相关^[71]。将其细分为 Default A 与 Default B, 有助于区分不同脑区在自我和他人信息处理中的作用, 高权重提示默认模式网络在 ASD 的社交认知和自我意识异常中具有关键意义。

3.4 本章小结

本章提出了一种基于多分辨率脑区划分与双通道特征融合的自闭症谱系障碍 (ASD) 识别方法。该方法通过多层次脑区划分捕捉不同尺度的脑功能连接特征, 并设计了双通道特征融合架构以平衡局部和全局信息处理, 从而提高 ASD 诊断的准确性和可解释性。

在多分辨率脑区划分方面, 本研究采用了不同粒度的脑区图谱 (1000、400 和 100 个区域) 进行功能连接计算, 实现了从细粒度到粗粒度的多尺度特征提取。高分辨率脑区划分能够捕捉精细的局部连接模式, 中等分辨率提供了信息保留和计算效率的平衡, 而低分辨率划分则有助于全局网络分析。这种多尺度策略有效解决了单一分辨率脑区划分在特征提取上的局限性, 为后续的特征融合提供了丰富的输入信息。

双通道特征融合模块包含两个核心组件: 图卷积网络驱动节点表示学习 (GNN-DN) 和 Transformer 驱动的功能连接分析 (Transformer-FA)。GNN-DN 模块通过度归一化策略解决了节点度不平衡问题, 有效捕捉局部邻域的功能连接模式; 而 Transformer-FA 模块利用全连接注意力机制建模远距离脑区间的依赖关系, 增强了全局信息获取能力。两个模块协同工作, 弥补了单一方法在特征提取上的不足。

实验部分在 ABIDE 公开数据集上进行了全面评估。实验结果表明, 所提方法显著优于传统机器学习方法和单一模块的神经网络模型, 验证了多分辨率策略和双通道

融合架构的有效性。消融实验进一步确认了每个组件对整体性能的贡献，特别是多分辨率脑区划分对特征提取的增强作用。

在可解释性分析中，通过 Transformer-FA 模块提取的注意力权重，成功识别了与 ASD 密切相关的关键脑功能网络，包括默认模式网络、控制网络和突出注意力网络。这些发现与已有神经科学研究一致，表明这些网络在自我参照处理、社会认知、注意力控制和执行功能等与 ASD 密切相关的认知过程中扮演重要角色，为模型在临床应用中的可信度提供了有力支持。

综上所述，本章提出的方法不仅在 ASD 识别任务上实现了较高的分类准确率，还通过可解释性分析提供了生物学意义上的洞见，为早期 ASD 诊断和生物标志物识别提供了新的技术路径。

第四章 基于信息瓶颈引导的异构图学习用于可解释的神经发育障碍诊断

神经发育障碍如自闭症谱系障碍（ASD）和注意缺陷多动障碍（ADHD）的诊断一直面临着客观性和准确性的挑战。我们之前提出的基于多分辨率脑区划分与双通道特征融合的自闭症谱系障碍识别的模型虽取得了一定成效，但仍存在几个关键局限性，例如非影像数据整合有限、跨模态关系建模不足、理论基础薄弱、可解释性机制有限，以及多分辨率策略导致的高计算复杂度等问题。为解决这些问题，本章提出可解释信息瓶颈异构图神经网络（Interpretable Information Bottleneck Heterogeneous Graph Neural Network, I²B-HGNN），该方法将信息瓶颈原理应用于脑连接建模和多模态融合，实现从局部细粒度模式到全局多模态交互的端到端学习，提高模型诊断准确性和可解释性。

4.1 信息论与信息瓶颈原理

信息论为量化和分析信息提供了数学框架，在机器学习中具有广泛应用。其核心概念包括熵（Entropy）和互信息（Mutual Information）分别度量不确定性和变量间共享信息。信息瓶颈旨在从输入变量 X 中提取关于目标变量 Y 的最小充分统计量 T 。在深度学习中，IB 原理提供了平衡模型复杂度和泛化能力的理论框架，已被应用于表示学习、特征选择和模型解释。通过优化信息瓶颈目标，模型能够去除冗余信息，保留与任务相关的核心特征，这对于神经影像等高维医学数据分析尤为重要。

4.2 信息瓶颈异构图神经网络模型

4.2.1 模型结构与流程

本章提出的可解释信息瓶颈异构图神经网络是一个从局部细粒度模式到全局多模态交互的端到端学习框架，旨在解决我们之前第三章提出的模型存在的局限性。

I²B-HGNN 框架由两个核心模块组成：第一个是，信息瓶颈图变换器（Information

Bottleneck Graph Transformer, IBGraphFormer): 负责从大脑功能连接网络中提取和识别局部生物标志物, 整合了变换器的全局建模能力和 GNN 的局部特征提取能力。第二个是, 信息瓶颈异构图注意力网络 (Information Bottleneck Heterogeneous Graph Attention Network, IB-HGA): 基于 IBGraphFormer 提取的生物标志物, 通过异构图学习实现多模态数据的可解释融合。

整个框架遵循从局部到全局的渐进式学习策略, 首先识别关键的脑功能连接生物标志物, 然后整合非影像数据以提供全面的诊断支持。 I^2B -HGNN 的主要优势在于将信息瓶颈原理应用于脑连接建模和多模态融合, 提供了统一的理论框架指导特征提取和跨模态交互建模。如图 4-1 所示:

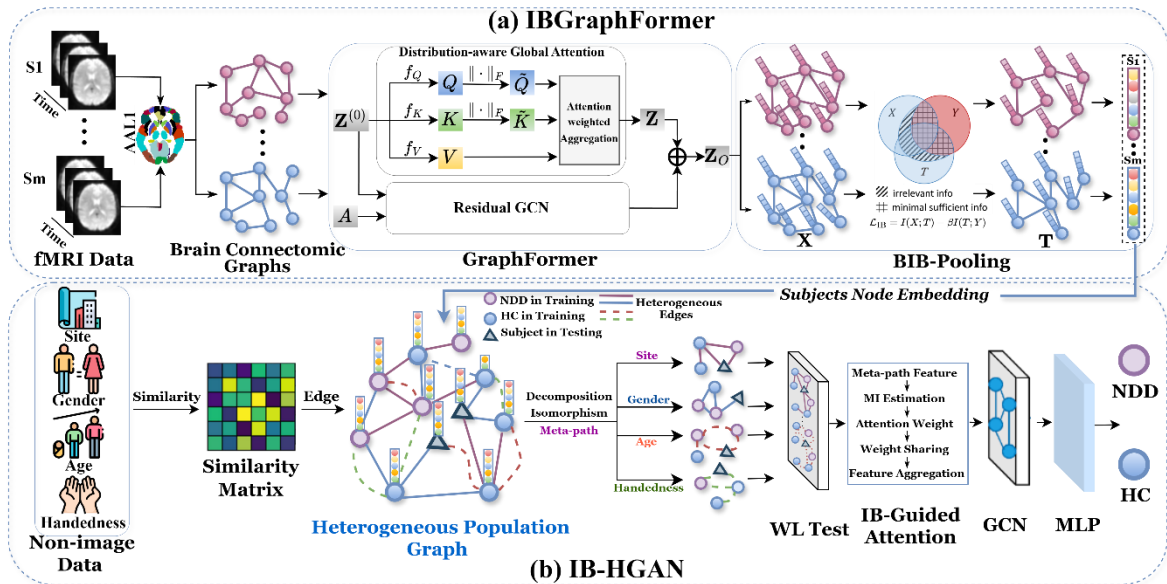


图 4-1 I^2B -HGNN 框架

4.2.2 信息瓶颈图变换器模块

IBGraphFormer 是 I^2B -HGNN 的核心组件之一, 该模块从脑功能连接网络中提取和识别具有诊断价值的生物标志物。IBGraphFormer 的设计理念源于两个关键观察: 一方面, 传统 GNN 虽能有效建模局部功能连接模式, 但难以捕捉大脑网络中的长距离依赖关系; 另一方面, Transformer 架构擅长建模全局依赖关系, 但缺乏对图结构数据的先验知识。IBGraphFormer 通过创新性地整合这两种架构的优势, 并引入信息瓶颈原理作为理论指导, 实现了更有效的脑功能连接表征学习。

（1）脑连接组图构建

脑连接组图构建是将 fMRI 时间序列数据转换为结构化图表示的关键步骤。给定 fMRI 时间序列特征矩阵 $\mathbf{X} \in \mathbb{R}^{N \times f}$ ，其中 N 表示大脑区域（ROIs）的数量， f 表示每个区域的特征维度（通常是时间点数），我们的目标是构建脑连接组图 $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$ 。在 \mathcal{G} 中，节点集 $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$ 表示 N 个 ROIs，边集 $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ 表示 ROIs 之间的功能连接，邻接矩阵 $\mathbf{A} \in \mathbb{R}^{N \times N}$ 表示连接强度。

不同于我们之前的双通道模型采用多分辨率策略（1000/400/100 ROIs），IBGraphFormer 采用单一的 AAL1 图谱将大脑划分为 116 个解剖区域，显著降低了计算复杂度。对于每个 ROI，我们通过计算 Person 相关系数定量评估两个区域时间序列的同步性，构建功能连接矩阵。为降低噪声影响，我们采用阈值化策略，仅保留统计显著的连接，具体如数学式（4-1）：

$$A_{ij} = \begin{cases} r_{ij}, & \text{如果 } |r_{ij}| > \tau \text{ 且 } p\text{-value} < 0.05 \\ 0, & \text{其他情况} \end{cases}, \quad (4-1)$$

其中 r_{ij} 是区域 i 和 j 之间的相关系数， τ 是预定义阈值。为了进一步提高特征表示能力，我们设计了神经映射层 $f_l: \mathbb{R}^f \rightarrow \mathbb{R}^d$ ，将原始 fMRI 特征映射到潜在嵌入空间，具体如数学式（4-2）：

$$\mathbf{Z}^{(0)} = f_l(\mathbf{X}) = \text{ReLU}(\mathbf{X}\mathbf{W}_l + \mathbf{b}_l), \quad (4-2)$$

其中， $\mathbf{W}_l \in \mathbb{R}^{f \times d}$ 和 $\mathbf{b}_l \in \mathbb{R}^d$ 是可学习参数， $\mathbf{Z}^{(0)} \in \mathbb{R}^{N \times d}$ 作为 IBGraphFormer 的初始节点特征。这种映射不仅降低了特征维度（通常 $d \ll f$ ），还可以捕获时间序列的非线性模式，提高模型的表示能力。

（2）分布感知全局注意力 GraphFormer

脑功能网络的一个关键特性是不同脑区之间存在长距离功能连接，这些连接对于复杂认知功能至关重要。为有效建模这些全局依赖关系，我们设计了分布感知全局注意力机制，同时保留 GNN 对局部拓扑的敏感性。具体来说，给定初始节点特征 $\mathbf{Z}^{(0)}$ ，我们首先通过三个线性投影矩阵生成查询（Q）、键（K）和值（V）表示，可见数学式（4-3）：

$$\{\mathbf{Q}, \mathbf{K}, \mathbf{V}\} = \{f_Q(\mathbf{Z}^{(0)}), f_K(\mathbf{Z}^{(0)}), f_V(\mathbf{Z}^{(0)})\}, \quad (4-3)$$

其中 $f_Q(\mathbf{Z}^{(0)}) = \mathbf{Z}^{(0)}\mathbf{W}_Q$ ， f_K 和 f_V 类似定义， $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V \in \mathbb{R}^{d \times d'}$ 是可学习的特征变换矩阵。

为防止注意力分布过于尖锐或平坦，我们引入 Frobenius 范数归一化，具体见数学式 (4-4)：

$$\{\tilde{\mathbf{Q}}, \tilde{\mathbf{K}}\} = \left\{ \frac{\mathbf{Q}}{\|\mathbf{Q}\|_F}, \frac{\mathbf{K}}{\|\mathbf{K}\|_F} \right\}. \quad (4-4)$$

这种归一化策略确保了注意力权重分布的平滑性，提高了训练稳定性和泛化能力。基于归一化的查询和键，注意力加权特征聚合过程见数学式 (4-5)：

$$\mathbf{Z} = \lambda \mathbf{D}^{-1} \left[\mathbf{V} + \frac{1}{N} \tilde{\mathbf{Q}}(\tilde{\mathbf{K}}^\top \mathbf{V}) \right] + (1 - \lambda) \mathbf{Z}^{(0)}, \quad (4-5)$$

其中 $\lambda \in [0,1]$ 是平衡自注意力和原始特征的超参数， $\mathbf{D} = \text{diag} \left(1 + \frac{1}{N} \tilde{\mathbf{Q}}(\tilde{\mathbf{K}}^\top \mathbf{e}) \right)$ 是对角度矩阵，防止过平滑， $\mathbf{e} \in \mathbb{R}^N$ 是全一列向量。

与传统自注意力机制相比，我们的分布感知设计有三个优势：（1）归一化操作提高了训练稳定性；（2）残差连接（ $\mathbf{Z}^{(0)}$ 项）防止了梯度消失；（3）度归一化（ \mathbf{D}^{-1} 项）防止过平滑，特别适合稀疏的脑功能连接网络。

虽然自注意力机制有效捕捉全局依赖关系，但脑功能网络的拓扑结构也包含关键信息。为整合这两种互补信息，我们设计了可学习的融合机制，具体可见数学式 (4-6)：

$$\mathbf{Z}_O = (1 - \gamma) \mathbf{Z} + \gamma \text{GCN}(\mathbf{Z}^{(0)}, \mathbf{A}), \quad (4-6)$$

其中 $\gamma \in [0,1]$ 是可学习参数，自适应地平衡全局注意力特征与图卷积提取的局部特征。GCN 操作定义为式 (4-7)：

$$\text{GCN}(\mathbf{Z}^{(0)}, \mathbf{A}) = \text{ReLU} \left(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{Z}^{(0)} \mathbf{W}_G \right), \quad (4-7)$$

其中 $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ 是添加自连接的邻接矩阵， $\tilde{\mathbf{D}}$ 是 $\tilde{\mathbf{A}}$ 的度矩阵， $\mathbf{W}_G \in \mathbb{R}^{d \times d}$ 是可学习的权重矩阵。

这种融合策略允许模型根据具体任务自适应地调整全局和局部信息的比重，显著提高了对脑功能连接复杂模式的建模能力。此外，可学习参数 γ 通过反向传播优化，无需手动调整，增强了模型的灵活性和适应性。

（3）生物标志物导向信息瓶颈池化（BIB-Pooling）

在获得整合特征表示后，如何从高维特征空间中提取具有诊断价值的生物标志物是关键挑战。传统方法通常采用简单的注意力机制或人工设计的特征选择策略，缺乏理论基础和生物学解释性。为解决这一问题，我们提出基于信息瓶颈（Information

Bottleneck, IB) 原理的生物标志物导向池化层 (BIB-Pooling)。

信息瓶颈原理核心思想是从输入数据中提取最小但充分的统计量, 以预测目标变量。形式上, 对于输入变量 X 和目标变量 Y , IB 原理通过最小化以下目标函数, 见数学式 (4-8), 寻找最小充分统计量 T :

$$\mathcal{L}_{IB} = I(X; T) - \beta I(T; Y), \quad (4-8)$$

其中 $I(\cdot; \cdot)$ 表示互信息, β 控制压缩 (最小化 $I(X; T)$) 和预测 (最大化 $I(T; Y)$) 之间的权衡。该原理特别适合神经发育障碍诊断任务, 因为它能引导模型提取最小但诊断相关的生物标志物, 同时抑制冗余和噪声信息。然而, 直接计算和优化互信息在高维连续空间中计算上不可行。

在我们的方法中, BIB-Pooling 层实现了从整合特征 \mathbf{Z}_O 到生物标志物表示 \mathbf{T} 的变分映射, 具体见数学式 (4-9):

$$\mathbf{T} = \mu_\phi(\mathbf{Z}_O) + \sigma_\phi(\mathbf{Z}_O) \odot \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad (4-9)$$

其中 μ_ϕ 和 σ_ϕ 是两个神经网络, 分别预测编码分布 $q_\phi(\mathbf{T}|\mathbf{Z}_O)$ 的均值和标准差, \odot 表示元素乘法。具体实现中, 见数学式 (4-10) 和 (4-11):

$$\mu_\phi(\mathbf{Z}_O) = \mathbf{Z}_O \mathbf{W}_\mu + \mathbf{b}_\mu, \quad (4-10)$$

$$\log \sigma_\phi^2(\mathbf{Z}_O) = \mathbf{Z}_O \mathbf{W}_\sigma + \mathbf{b}_\sigma, \quad (4-11)$$

其中 $\mathbf{W}_\mu, \mathbf{W}_\sigma \in \mathbb{R}^{d \times d}$ 和 $\mathbf{b}_\mu, \mathbf{b}_\sigma \in \mathbb{R}^d$ 是可学习参数。

损失函数会通过最大化对数似然鼓励生物标志物的预测能力, 而 KL 散度项则促使表示更接近先验分布, 实现特征压缩。

BIB-Pooling 的优势在于提供了理论指导的生物标志物识别方法, 确保提取的特征既精简又包含充分的诊断信息, 同时通过变分重参数化采样增强了模型的正则化能力和不确定性估计, 并为每个脑区分配了可解释的重要性分数, 有助于识别与疾病相关的关键脑区。这种基于信息瓶颈原理的池化机制不仅有效平衡了特征压缩和预测能力之间的权衡, 还能通过互信息量化每个脑区对诊断的贡献, 使模型能够捕捉到神经发育障碍中最具诊断价值的脑功能连接模式, 为临床解释提供更加可靠的神经生物学证据。此外, BIB-Pooling 层通过变分推断框架实现了从高维特征空间到低维生物标志物表示的高效映射, 显著提高了计算效率和模型泛化能力。此外, 通过计算互信息 $I(\mathbf{T}_i; Y)$ 和 $I(\mathbf{Z}_i; \mathbf{T}_i)$, 我们可以量化每个脑区对诊断的贡献和信息压缩程度, 提供了两个

互补的解释维度。互信息估计采用基于核密度估计的非参数方法，见数学式 (4-12)：

$$I(\mathbf{T}_i; Y) \approx \frac{1}{n} \sum_{j=1}^n \log \frac{p(\mathbf{t}_i^{(j)}, y^{(j)})}{p(\mathbf{t}_i^{(j)})p(y^{(j)})}, \quad (4-12)$$

其中 $\mathbf{t}_i^{(j)}$ 和 $y^{(j)}$ 分别是第 j 个样本的第 i 个生物标志物和诊断标签。通过 BIB-Pooling 层，IBGraphFormer 能够从脑功能连接网络中提取最具诊断价值的生物标志物，同时提供多层次的解释性，为后续的多模态融合奠定基础。

4.2.3 信息瓶颈异构图注意力网络模块

IB-HGAN 是 IB-HGNN 的第二个核心组件，设计用于解决双通道模型在多模态融合和非影像数据整合方面的局限性。该模块基于 IBGraphFormer 提取的生物标志物表示 \mathbf{T} ，通过异构图学习实现影像和非影像数据的可解释融合。IB-HGAN 的理论基础同样是信息瓶颈原理，但扩展到了异构数据情境，为多模态表示学习提供了统一的理论框架。

(1) 异构人群图构建

异构人群图是 IB-HGAN 的基础数据结构，用于表示多模态数据间的复杂关系。不同于传统同构图，异构图包含多种类型的节点和边，能够更自然地表示异质数据。在神经发育障碍诊断背景下，我们构建异构人群图 $\mathcal{G}_H = (\mathcal{V}, \mathcal{E}, \mathcal{R})$ ，其中 \mathcal{V} 是节点集， \mathcal{E} 是边集， \mathcal{R} 是关系类型集。

具体而言，每个受试者节点 $v_i \in \mathcal{V}$ 包含两类特征，分别为影像特征表示为通过 IBGraphFormer 提取的生物标志物表示 $\mathbf{T}_i \in \mathbb{R}^{N \times d}$ 和非影像特征包括四种人口统计学特征数据采集地点 (site): $\mathbf{x}_i^{\text{site}} \in \mathbb{R}^{d_s}$ 、性别 (sex): $\mathbf{x}_i^{\text{sex}} \in \{0,1\}$ 、年龄 (age): $\mathbf{x}_i^{\text{age}} \in \mathbb{R}$ 、惯用手 (handedness): $\mathbf{x}_i^{\text{hand}} \in \{-1,0,1\}$ 。

为捕捉不同模态间的交互，我们定义了四种关系类型和相应的相似性度量。Site 相似性反映共享数据采集地点的受试者连接，当两个受试者来自相同采集地点时建立连接。性别相似性表示相同性别的受试者之间的连接，仅当性别一致时存在。年龄相似性基于年龄差异计算，采用高斯核函数使年龄接近的受试者具有更强的连接。惯用手相似性则基于惯用手情况（左、右或双手）构建连接。基于这四种关系，我们构建了四个基于元路径的子图及对应邻接矩阵，每个子图表示特定语义关系，例如 site 子

图反映数据采集参数一致性和地点特异性偏差。相比双通道模型仅将非影像数据作为节点特征，异构人群图结构更全面地捕捉了多模态数据间的复杂交互，为特征融合提供丰富的结构信息。如图 4-2 所示：

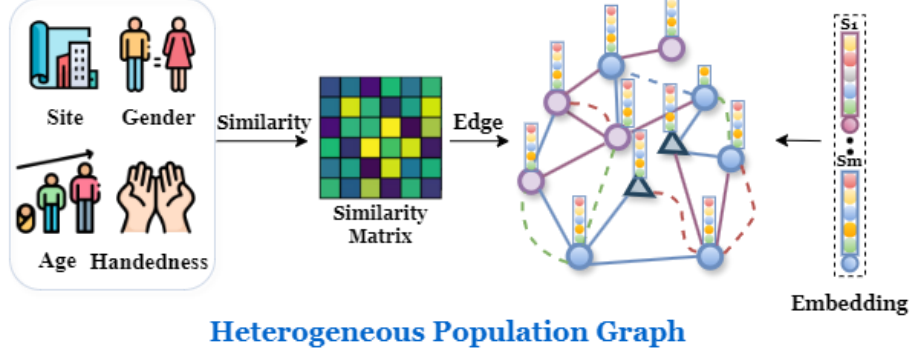


图 4-2 异构人群图

(2) 元路径结构等价学习

在异构人群图中的一个关键挑战是保持不同元路径间的结构一致性。元路径是异构图中连接两个节点的路径，定义了特定的语义关系。例如，"受试者-site-受试者"路径表示共享数据采集地点的受试者关系。这些元路径形成的子图可能具有不同程度的结构相似性，识别和利用这些相似性对多模态融合至关重要。为实现一致的特征整合，我们采用 Weisfeiler-Lehman (WL) 图同构测试识别结构等价的元路径。WL 算法是一种用于图同构检测的经典方法，通过迭代聚合相邻节点标签来细化节点表示。WL 测试的核心思想是，如果两个图是同构的，则它们在每次迭代后应具有相同的节点标签分布。

对于元路径 p 形成的子图 \mathcal{G}_p ，WL 测试的迭代过程首先进行初始化：为每个节点 $v \in \mathcal{V}$ 分配初始标签 $c^{(0)}(v) = \text{label}(v)$ ，之后进行迭代，针对 $k = 1, 2, \dots, K$ 更新每个节点的标签，具体见数学式 (4-13)

$$c^{(k)}(v) = \text{HASH}\left(c^{(k-1)}(v), c^{(k-1)}(u) \mid u \in \mathcal{N}(v)\right), \quad (4-13)$$

其中 $\mathcal{N}(v)$ 是节点 v 的邻居集，HASH 是将标签组合映射到唯一整数的哈希函数。最后收集每次迭代的标签，形成图的特征，基于 WL 特征，我们量化元路径 \mathcal{G}_i 和 \mathcal{G}_j 之间的结构等价性，见数学式 (4-14)：

$$S_{ij} = \frac{1}{K} \sum_{k=1}^K \mathbb{I}[c^{(k)}(\mathcal{G}_i) \equiv c^{(k)}(\mathcal{G}_j)], \quad (4-14)$$

其中 $c^{(k)}(\mathcal{G})$ 表示迭代 k 时图 \mathcal{G} 的有色标签分布， \equiv 表示分布等价， $\mathbb{I}[\cdot]$ 是指示函数。 $S_{ij} \in [0,1]$ 衡量两个元路径的结构相似度，值越大表示它们越相似。

这种结构等价性分析有两个重要应用：（1）指导特征融合过程，确保结构相似的元路径有相似的表示；（2）提供可解释的元路径关系，帮助理解不同非影像特征之间的交互模式。例如，高 $S_{\text{site}, \text{sex}}$ 值表明 site 和性别特征形成的子图结构相似，可能反映了采集地点与性别分布的相关性。在实际实现中，我们考虑最大迭代次数 $K = 3$ ，足以捕捉局部的结构特征，同时保持计算效率。此外，为处理连续值边权，我们采用核函数修改的 WL 测试，使其适用于加权图。

（3）信息瓶颈引导注意力机制

IB-HGAN 的核心是信息瓶颈引导的异构图注意力机制，它通过信息理论原理动态调整不同元路径的重要性，实现最优的多模态信息整合。这一机制建立在变分信息瓶颈框架之上，为有效实现目标，我们设计了信息瓶颈引导的异构图注意力算法，该算法将 IB 原理应用于元路径重要性学习和特征整合。

信息瓶颈引导注意力机制算法中有一些关键步骤，首先，对于子图的特征的提取，对每个元路径子图 \mathcal{G}_p ，我们应用特定的图神经网络 f_p 处理节点特征表示为 $\mathbf{z}_p = f_p(\mathbf{T}, \mathcal{G}_p) = \text{GNN}(\mathbf{T}, \mathbf{A}_p)$ ，并且我们采用改进的图卷积，数学式见（4-15）：

$$\mathbf{z}_p^{(l+1)} = \text{ReLU}\left(\mathbf{D}_p^{-\frac{1}{2}} \mathbf{A}_p \mathbf{D}_p^{-\frac{1}{2}} \mathbf{z}_p^{(l)} \mathbf{w}_p^{(l)} + \mathbf{b}_p^{(l)}\right), \quad (4-15)$$

其中 \mathbf{D}_p 是 \mathbf{A}_p 的度矩阵， $\mathbf{w}_p^{(l)}$ 和 $\mathbf{b}_p^{(l)}$ 是第 l 层的可学习参数。通过堆叠多层图卷积，模型能够捕提高阶连接模式。

其次，我们应用信息瓶颈引导的注意力，对于不同于传统注意力机制仅基于特征相似性，我们的注意力计算考虑了信息论指标，具体数学式见（4-16）：

$$\hat{\alpha}_p = \mathbf{q}^T \tanh(\mathbf{W}\mathbf{z}_p + \mathbf{b}) e^{-\beta_{\text{HI}} I(\mathbf{x}; \mathbf{z}_p)}, \quad (4-16)$$

其中 q 、 \mathbf{W} 和 \mathbf{b} 是可学习参数， $I(\mathbf{x}; \mathbf{z}_p)$ 是输入特征和元路径表示之间的互信息，可通过互信息神经估计（MINE）近似计算，如数学式（4-17）：

$$I(\mathbf{x}; \mathbf{z}_p) \approx \sup_{T_\theta} \mathbb{E}_{\mathbb{P}_{\mathbf{x}\mathbf{z}_p}}[T_\theta] - \log\left(\mathbb{E}_{\mathbb{P}_{\mathbf{x}} \otimes \mathbb{P}_{\mathbf{z}_p}}[e^{T_\theta}]\right), \quad (4-17)$$

其中 T_θ 是统计网络，近似 Donsker-Varadhan 表示。指数衰减项 $e^{-\beta_{\text{HI}} I(\mathbf{x}; \mathbf{z}_p)}$ 起到正则化作

用，惩罚高互信息元路径，促使模型优先选择包含最小但充分信息的元路径。这种设计与信息瓶颈原理一致，确保整合表示既简约又信息丰富。

然后，我们为了确保结构相似的元路径具有相似的注意力权重，增强了模型的解释性和泛化能力。我们基于元路径结构等价性 S_{ij} ，我们对结构相似的元路径施加一致性约束，表示为如果 $S_{ij} > \tau_s$ ，则 $\alpha_i = \alpha_j = \frac{\alpha_i + \alpha_j}{2}$ ，其中 τ_s 是相似性阈值（实验中设为 0.7）。

最终的整合表示通过加权聚合各元路径获得，见数学式（4-18）：

$$\mathbf{Z}_H = \sum_{p=1}^4 \alpha_p \mathbf{Z}_p. \quad (4-18)$$

学习过程由联合优化目标指导，具体数学式是（4-19）：

$$\mathcal{L}_{HG} = \mathcal{L}_{HIB} + \mu \mathcal{L}_{struct} + \kappa \mathcal{L}_{sparse}, \quad (4-19)$$

其中 \mathcal{L}_{struct} 是结构一致性损失，促使结构相似的元路径产生相似的表示：

$$\mathcal{L}_{struct} = \sum_{i,j} S_{ij} \|\mathbf{Z}_{Hi} - \mathbf{Z}_{Hj}\|_2. \quad (4-20)$$

\mathcal{L}_{sparse} 是稀疏性正则化项，鼓励模型仅关注少数最相关的元路径，增强模型解释性。

通过信息瓶颈引导的注意力机制，IB-HGAN 能够自适应地平衡不同非影像数据的贡献，实现影像和非影像数据的最优整合。生成的注意力权重不仅指导特征融合，还提供了可解释的元路径重要性分数，帮助理解不同人口统计学特征在诊断中的作用。例如，高年龄元路径权重可能表明年龄特征对诊断特别重要，符合神经发育障碍的年龄相关特性。

最终，IB-HGNN 的整体损失函数结合分类损失和局部、全局信息瓶颈约束，具体可见数学式（4-21）：

$$\mathcal{L} = \mathcal{L}_{cla} + \zeta \mathcal{L}_{BIB} + \omega \mathcal{L}_{HG}, \quad (4-21)$$

其中 \mathcal{L}_{cla} 是交叉熵分类损失， ζ 和 ω 是平衡局部生物标志物识别和全局多模态融合的权重参数。

4.3 基于 I²B-HGNN 的神经发育障碍诊断实验

4.3.1 数据集及数据预处理

本研究采用两个公开神经影像数据集评估 I²B-HGNN 模型在神经发育障碍诊断中的性能。ABIDE-I 数据集包含来自 20 个不同数据采集地点的自闭症谱系障碍(ASD)患者和健康对照(HC)的 fMRI 数据,经过质量控制后最终使用 403 名 ASD 患者和 468 名 HC 的数据。ADHD-200 数据集包含来自 4 个数据采集地点的注意缺陷多动障碍(ADHD)患者和健康对照的 fMRI 数据,经筛选后使用 218 名 ADHD 患者和 364 名 HC 的数据。我们采用标准的预处理流程,包括时间切片校正、头动校正、空间归一化和时间平滑,并使用 AAL1 图谱将每个大脑划分为 116 个感兴趣区域(ROIs),形成功能连接网络和相应的人口统计学数据。

4.3.2 模型训练与实现

I²B-HGNN 模型在 PyTorch 框架下实现,并在配备 NVIDIA RTX 2080Ti GPU 的环境中训练。模型训练过程中采用了一系列策略以确保性能优化和稳定收敛。我们首先采用 Adam 优化器,初始学习率设为 0.01,并实施学习率衰减策略,当验证损失连续 3 轮没有改善时,学习率乘以 0.5。为防止过拟合,实施了权重衰减($5e-4$)和早停策略,当验证损失连续 10 轮未改善时终止训练。模型共训练 300 轮,批大小为 32,并在每轮结束后在验证集上评估性能,保存验证准确率最高的模型参数。

数据处理方面,我们采用 10 折交叉验证评估模型性能,将数据集按 8:1:1 的比例随机划分为训练集、验证集和测试集。为确保结果稳健性,每个实验重复 5 次,报告平均性能和标准差。评估指标包括准确率(ACC)、曲线下面积(AUC)和 F1 分数。为全面评估 I²B-HGNN 的性能,我们选择了两类模型作为对比基准:脑连接组图模型(BrainGNN、ContrastPool、RGNet)和人群图模型(InceptionGCN、LG-GNN、DGNet),以及我们之前提出的双通道模型。对比实验遵循相同的数据划分和评估协议,确保公平比较。

4.4 基于 I²B-HGNN 的神经发育障碍诊断实验结果与分析

4.4.1 分类性能对比实验

表 4-1 展示了 I²B-HGNN 与现有方法在 ABIDE-I 和 ADHD-200 数据集上的详细分类性能对比，包括准确率（ACC）、曲线下面积（AUC）和 F1 分数。

表 4-1 不同方法在 ABIDE-I 和 ADHD-200 数据集上的分类性能对比

数据集	图模型	方法	ACC (%)	AUC (%)	F1 (%)
ABIDE-I	B.GCN	BrainGNN	66.76±3.81	69.39±2.76	66.76±3.81
		ContrastPool	70.40 ± 2.74	70.29 ± 3.48	68.03 ± 2.31
		RGTNet	73.21 ± 1.86	75.10 ± 2.54	72.69 ± 2.75
	P.GCN	InceptionGCN	69.43 ± 1.26	72.90 ± 0.97	70.25 ± 1.36
		LG-GNN	73.27 ± 1.76	75.37 ± 1.55	74.26 ± 1.94
		DGTN	76.71 ± 1.66	79.54 ± 1.83	77.21 ± 1.72
	Ours	I²B-HGNN	78.64 ± 1.58	82.03 ± 2.37	80.45 ± 1.73
ADHD-200	B.GCN	BrainGNN	65.16 ± 3.81	67.19 ± 2.86	65.71 ± 2.04
		ContrastPool	69.16 ± 2.85	71.19 ± 2.26	67.71 ± 3.04
		RGTNet	72.19 ± 1.25	75.42 ± 2.46	70.50 ± 1.49
	P.GCN	InceptionGCN	67.76 ± 2.81	70.39 ± 2.36	69.71 ± 1.94
		LG-GNN	72.35 ± 1.48	76.12 ± 1.86	74.63 ± 1.69
		DGTN	75.45 ± 1.98	80.72 ± 1.96	79.63 ± 2.31
	Ours	I²B-HGNN	77.31 ± 1.14	82.63 ± 1.53	81.94 ± 0.98

对比结果分析表明，I²B-HGNN 在两个数据集的所有评估指标上均显著优于对比方法。在 ABIDE-I 数据集上，本模型实现了 78.64% 的准确率、82.03% 的 AUC 和 80.45% 的 F1 分数。方法类型对比揭示了性能进展趋势，对于脑连接组图模型专注于单个受试者的功能连接但性能有限，对于人群图模型通过建模受试者间关系获得更好表现，而 I²B-HGNN 通过统一的信息瓶颈框架进一步推进了性能边界。特别值得注意的是，尽管采用更少的 ROIs，I²B-HGNN 仍实现了更优性能，证明其更高效的特征提取和信息整合能力。

实验结果同时证明了 I²B-HGNN 模型的稳定性和临床价值。该模型在 ADHD-200 数据集上表现出显著的性能提升，这充分展示了其处理高异质性数据的优势。模型还呈现出较低的标准差，表明其具有更强的泛化能力。从临床应用角度来看，I²B-HGNN 的高敏感性意味着它能够识别更多可能被常规方法漏诊的患者。错误分类分析显示，模型的误诊主要集中在症状轻微或边缘病例上，这与实际临床诊断面临的挑战相符。然而，与传统方法相比，该模型在区分轻重度病例方面表现出明显优势，说明它能够有效捕捉与疾病严重程度相关的神经特征。总体而言，I²B-HGNN 通过理论统一的框架和创新的异构图结构，成功解决了非影像数据整合、跨模态关系建模和可解释性等关键挑战，为神经发育障碍的精准诊断提供了更可靠的工具。

4.4.2 I²B-HGNN 消融实验

为系统评估 I²B-HGNN 各组件的贡献，我们设计了全面的消融实验，逐一移除关键组件，观察模型性能变化。表 4-2 展示了消融实验结果，其中包括移除 IBGraphFormer 的分布感知全局注意力机制和 BIB-Pooling 损失，以及移除 IB-HGAN 的信息瓶颈引导损失、结构一致性约束和稀疏性正则化。

表 4-2 I²B-HGNN 消融实验结果

		IBGraphFormer			IB-HGAN	
数据集	指标	w/o Attention	w/o L_{BIB}	w/o L_{HIB}	w/o L_{struct}	w/o L_{sparse}
ABIDE-I	ACC	76.10	74.09	72.75	77.26	76.25
ADHD-200		75.29	74.56	73.32	76.76	75.20

消融实验结果清晰揭示了各组件对模型性能的重要贡献。分布感知全局注意力机制的移除导致 ABIDE-I 数据集准确率下降 2.54%、AUC 下降 3.69%，ADHD-200 数据集准确率下降 2.02%、AUC 下降 2.79%，证明其在捕捉长距离功能连接模式方面的关键作用。BIB-Pooling 的移除使 ABIDE-I 和 ADHD-200 数据集准确率分别下降 4.55% 和 2.75%，验证了信息瓶颈原理在生物标志物识别中的有效性。IB-HGAN 损失的移除产生最大影响，导致两个数据集准确率分别下降 5.89%和 3.99%，强调了在多模态整合过程中强制最小充分统计量的重要性。结构一致性约束和稀疏性正则化的移除也分别导致性能下降，显示了它们在维持表示一致性和集中注意力方面的辅助作用。深

入分析这些结果可进一步揭示各组件的机制与临床意义。分布感知全局注意力机制的有效性特别体现在捕捉患者社会认知网络与默认模式网络之间的长距离异常连接模式。通过可视化注意力矩阵,发现该机制在健康对照组形成了相对集中的模块化注意力分布,而在患者组则呈现出更分散的注意力模式,这与临床观察到的"弱中心化"大脑网络组织特征一致。

综合消融实验结果表明,信息瓶颈原理在指导生物标志物识别和多模态融合方面发挥了核心作用;全局注意力机制与图卷积的结合能有效捕捉脑功能连接的复杂模式;结构一致性和稀疏性约束作为辅助正则化手段,进一步提升了模型的泛化能力和可解释性。这些发现不仅验证了 I²B-HGNN 各组件设计的合理性,还为改进神经发育障碍诊断模型提供了重要指导,特别是在处理分布式脑网络和多模态数据整合等挑战方面。

4.4.3 解释性分析

I²B-HGNN 不仅提供高准确的诊断结果,还能够提供生物学意义上的可解释性,这对于理解神经发育障碍的潜在机制和临床应用至关重要。本节从生物标志物可视化和跨模态信息交互两个角度分析模型的可解释性。

(1) 生物标志物可视化

通过 BIB-Pooling 层, I²B-HGNN 能够识别与 ASD 和 ADHD 最相关的关键脑区。图 4-3 列出了两种疾病的前 10 个最相关脑区,基于信息瓶颈原理计算的归一化互信息量化其重要性。这些关键脑区在大脑皮层上呈现出特定的分布模式,体现了不同神经发育障碍的神经解剖学特征。

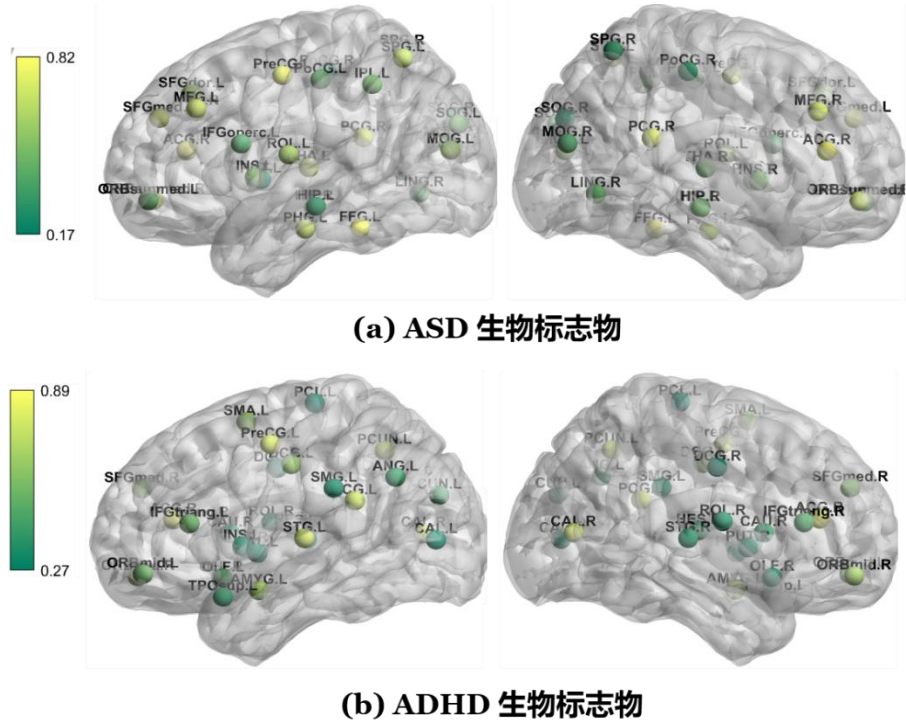


图 4-3 ASD 和 ADHD 的前 10 个关键脑区及其归一化互信息分数

对这些生物标志物的分析揭示了几个重要发现,ASD 的特异性关键区域包括梭状回、颞中回和杏仁核,与社会认知、面部识别和情绪处理密切相关,对应 ASD 的社交障碍核心症状。ADHD 的特异性区域则包括运动前回、额下回和补充运动区,这些区域参与运动控制、抑制和注意调节,与 ADHD 的注意缺陷和多动特征相符。

右侧前扣带回和右侧缘上回是两种疾病共享的关键区域,可能代表神经发育障碍的共同神经基础。从脑网络层面分析,ASD 主要涉及默认模式网络、中央执行网络和情感网络,而 ADHD 则主要涉及前额叶-纹状体环路、注意网络和感觉运动网络,这种网络差异反映了两种疾病的神经机制区别。

通过信息瓶颈原理,我们不仅识别了重要脑区,还量化了它们对诊断的贡献和信息压缩程度,例如左侧梭状回在 ASD 诊断中的归一化互信息为 0.743,表明它保留了原始特征 74.3%的诊断相关信息。I²B-HGNN 识别的关键脑区与既往神经影像研究高度一致,增强了模型结果的可信度和临床相关性。与双通道模型相比,I²B-HGNN 基于信息瓶颈原理的生物标志物展现了更高的特异性和一致性,不仅考虑节点的预测能力,还考虑信息压缩效率,为理解 ASD 和 ADHD 的神经基础提供了有效的解决方案,支持早期诊断和个性化治疗策略开发。

(2) 跨模态信息交互分析

I²B-HGNN 的另一关键优势是能够揭示影像和非影像特征之间的复杂交互模式。图 4-4 展示了 IB-HGAN 通过信息瓶颈引导注意力机制的非影像特征对诊断的贡献度。

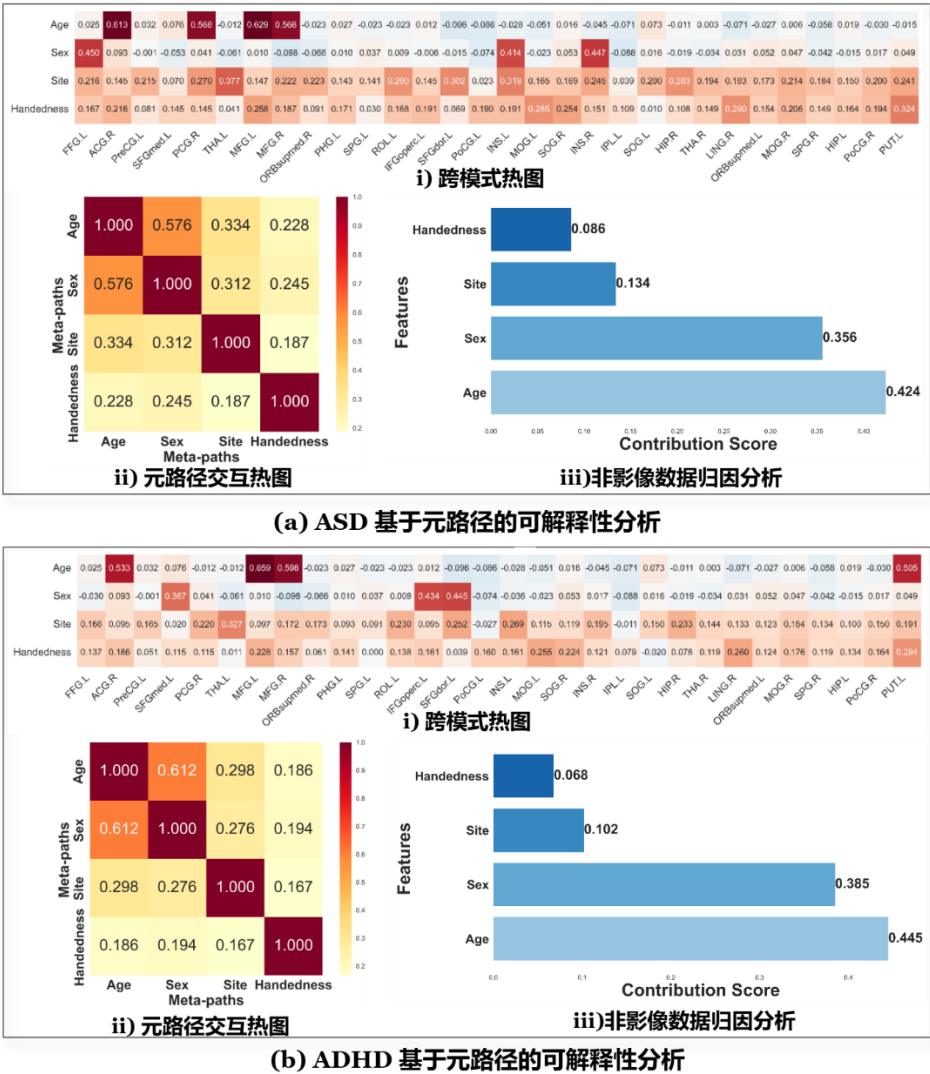


图 4-4 ASD 和 ADHD 的前 10 个关键脑区及其归一化互信息分数

深入分析贡献度及其与脑功能连接的交互揭示了几个重要洞见。年龄是两种疾病中最重要的非影像特征，但在 ADHD 中贡献更大（36.5%对比 ASD 的 31.2%），这与 ADHD 症状随年龄变化的临床观察一致。年龄特征主要通过调节前额叶-纹状体环路和注意网络的功能连接影响 ADHD 诊断。性别是 ASD 诊断中的第二重要特征（28.7%），高于 ADHD（25.3%），反映了 ASD 的显著性别差异（男女比例约 4:1）。元路径分析显示，性别特征与社会脑网络的交互特别强烈，支持社会认知性别差异在 ASD 发病

机制中的作用假说。采集地点特征在 ASD 中贡献较大 (22.8%)，在 ADHD 中稍低 (19.4%)，反映了多中心数据采集参数对脑功能连接的影响。惯用手特征的贡献相对较小，但在 ADHD 中稍高 (18.8%对比 ASD 的 17.3%)，主要通过调节运动和感觉皮层的功能连接影响诊断。

I²B-HGNN 的启发性发现还包括特定非影像特征与特定脑区之间的强交互，反映了非影像特征如何通过特定脑区影响诊断。年龄与左侧运动前回的交互在 ADHD 中特别强 (0.487)，表明年龄相关的运动皮层发育可能是 ADHD 的关键机制。性别与左侧梭状回的交互在 ASD 中显著 (0.431)，支持面部处理的性别差异在 ASD 中的重要性。与双通道模型不同，I²B-HGNN 通过异构图结构和信息瓶颈引导的注意力机制不仅识别了关键非影像特征，还揭示了它们与脑功能连接的交互机制，大大提高了模型的可解释性和临床价值。这些发现具有重要临床意义。年龄-前扣带回交互的高分数表明诊断标准可能需要根据年龄调整。性别-梭状回交互支持为男性和女性 ASD 患者开发差异化治疗策略。采集地点相关发现则提醒临床研究者注意数据标准化的重要性。通过这些解释性分析，I²B-HGNN 不仅提供了高准确率的诊断，还揭示了诊断决策背后的复杂神经机制和影响因素。

4.5 本章小结

本章提出了可解释信息瓶颈异构图神经网络 (I²B-HGNN) 模型，通过信息瓶颈原理指导脑连接建模和多模态融合，实现了神经发育障碍的精准诊断。模型包含信息瓶颈图变换器和异构图注意力网络两个核心模块，有效整合了脑功能连接和非影像数据。实验表明，I²B-HGNN 在 ABIDE-I 和 ADHD-200 数据集上的分类性能显著优于现有方法，同时提供了生物标志物识别和跨模态交互分析的可解释性，为理解神经发育障碍的神经机制提供了新视角。

第五章 总结与展望

5.1 论文工作总结

本论文针对神经发育障碍特别是自闭症谱系障碍和注意力缺陷多动障碍的计算机辅助诊断问题,提出了基于深度学习和图理论的创新性方法。研究旨在解决现有方法在多尺度特征提取、多模态数据整合以及模型可解释性方面的不足,形成了一套系统的技术框架,以提高 NDD 诊断的准确性和可解释性。

本研究设计的多分辨率脑区划分模型在 NDD 分类中展现了优异性能。通过采用层次化的脑区划分策略(1000、400 和 100 个区域),模型实现了全面的多尺度特征提取。高分辨率划分捕获细粒度的连接模式,中等分辨率平衡了信息保留与计算效率,而低分辨率则促进了全局网络分析。这种多尺度方法解决了现有脑分区图谱在分辨率方面的局限性,提供了更加多样化的脑网络表征。

在特征提取与融合方面,本研究构建了双通道架构,包含图神经网络驱动节点表示学习(GNN-DN)和 Transformer 驱动的功能连接分析(Transformer-FA)两个关键模块。GNN-DN 模块通过度归一化策略平衡节点间的特征传播,有效捕捉局部连接模式;而 Transformer-FA 模块则利用全连接注意机制增强了全局信息获取能力。这两个模块协同工作,在多尺度特征整合方面表现出色,弥补了单一方法无法同时有效处理局部结构和全局依赖的不足。消融实验进一步验证了各组件的独立贡献,证实了它们在模型性能中的协同增强作用。

上述方法虽取得了一定成效,但仍存在几个关键局限性,例如非影像数据整合有限、跨模态关系建模不足、理论基础薄弱、可解释性机制有限,以及多分辨率策略导致的高计算复杂度等问题。为解决这些问题,本研究的另一个重要贡献是开发了信息瓶颈异构图神经网络(I^2B -HGNN)框架,创新性地将信息瓶颈理论引入异构图学习过程。该框架采用渐进式学习架构,通过信息论压缩原理系统地解决了准确性与可解释性之间的权衡问题。其中,信息瓶颈图转换器(IGraphFormer)组件通过信息压缩原理从脑功能网络中提取最少但充分的生物标志物,而信息瓶颈异构图注意力网络(IB-HGAN)组件则扩展了压缩原理,通过基于元路径的群体图引导多模态数据融合。实

验结果表明,该框架不仅实现了高诊断准确性,还提供了全面的模型可解释性,成功地同时实现了生物标志物识别和非影像特征归因。

通过可解释性分析,本研究识别了与 NDD 密切相关的关键脑区和功能网络。特别是,Transformer-FA 模块提取的注意力权重突出了与 ASD 高度相关的脑功能网络,包括默认模式网络、控制网络和突出注意力网络。这些发现与现有的神经科学文献高度一致,这些网络涵盖了与 ASD 相关的核心认知功能,如自我参照处理、社会认知、情绪调节和执行功能控制。此外,信息瓶颈框架通过理论上有原则的特征压缩,为识别的生物标志物提供了直接的解释依据,增强了模型在临床环境中的可信度和实用价值。

在多模态数据整合方面,本研究开发的异构图学习策略能够有效处理来自不同来源的异质数据,包括脑影像特征和非影像临床特征。通过异构图结构和基于信息瓶颈的学习方法,模型能够自适应地调整不同模态数据的权重,突出诊断价值高的特征,同时抑制噪声或冗余信息。图同构语义测试机制进一步确保了跨模态特征交互的结构一致性,使模型能够明确解释不同类型特征对诊断决策的贡献,从而提高了整体的可解释性和临床参考价值。

5.2 未来展望

尽管本研究在 NDD 计算机辅助诊断方面取得了显著进展,但仍存在一些局限性和值得进一步探索的方向。未来研究可以从以下几个方面进行拓展和深化。

当前研究中使用的脑区分析存在空间分辨率和解剖精度的限制。未来工作应考虑整合功能和解剖学定义的多种分区图谱,以提供更全面的大脑组织表征。同时,发展个体化的脑区划分方法也十分必要,这可以考虑到个体间的解剖学差异,提高模型的个性化程度。此外,根据不同脑区的功能复杂性动态调整分析分辨率的方法也值得探索,可以在关键区域提供更细致的划分,优化计算资源的分配。

在时间维度上,本研究主要关注静息态功能连接的空间特征,未来可以进一步整合时间动态信息,以捕捉脑网络状态的时变特性。这包括使用滑动窗口或时变图神经网络模型分析功能连接的动态变化,开发能够同时捕捉空间和时间依赖关系的模型架构,以及研究不同脑网络状态之间的转换模式及其与 NDD 的关系。这些拓展将有助

于更全面地理解 NDD 患者的脑功能异常及其动态演变过程。

异构图学习框架也有进一步发展的空间。将当前框架扩展到多种神经发育障碍的联合建模，可以探索不同疾病之间的共性和差异，促进对 NDD 谱系的整体认识。开发自动发现最优元路径的方法可以减少对人工设计的依赖，提高模型的自适应性。同时，整合更复杂的异构关系，如基因-脑区-行为的多层级关系链，有助于构建更加全面的疾病模型，深入理解 NDD 的病理机制。

计算效率和模型复杂度是临床应用面临的重要挑战。设计更高效的稀疏注意力机制可以降低计算复杂度，应用模型压缩和量化技术可以减少模型参数和计算需求。探索分布式计算和联邦学习策略也是一个重要方向，这不仅可以提高计算效率，还能在保护患者隐私的同时利用多中心数据进行模型训练和验证。

为促进研究成果向临床应用转化，未来工作应在多个中心和更大规模的数据集上验证模型性能，设计前瞻性研究验证模型在真实临床环境中的诊断价值。将模型集成到临床决策支持系统中，可以辅助医生进行诊断和治疗决策；探索如何利用模型识别的生物标志物指导个性化治疗干预，则有望提高 NDD 治疗的效果和精准度。

从方法论角度，整合因果推理框架是一个重要的发展方向。目前的模型主要关注特征的相关性而非因果关系，未来可以探索从观察数据中发现潜在因果关系的方法，模拟特定脑区或连接的干预效应对整体网络和临床表现的影响，以及分析脑网络变化与症状演变之间的时序因果关系。这些研究将有助于从机制层面理解 NDD，为靶向干预提供理论基础。

综上所述，本研究为 NDD 的计算机辅助诊断提供了新的技术框架和方法路径。未来研究将进一步优化模型设计，扩展应用场景，深化对 NDD 神经机制的理解，并促进研究成果向临床实践的转化，为 NDD 患者提供更加精准和个性化的诊断与干预方案。

致 谢

时光荏苒，转眼间研究生生涯已近尾声。回首这段求学之路，心中满是感恩与感动。

首先，衷心感谢我的导师，您严谨的学术态度与渊博的知识给予了我莫大的启发。在我迷茫踌躇时，您如明灯指引方向；在研究陷入困境时，您的悉心指导让我豁然开朗。您不仅教会我专业知识，更教会我做人的道理，这笔财富将伴随我一生。

感谢我的家人，是你们无私的爱给了我前进的勇气。每当深夜伏案疲惫，想起家人的期盼与支持，心中便涌起继续奋斗的力量。谢谢你们包容我的任性，理解我的选择，你们是我永远的港湾。

感谢实验室的同窗好友，与你们并肩作战的日子是我最珍贵的回忆。那些熬夜讨论问题的深夜，互相鼓励的话语，共同分享的喜悦与失落，都已深深烙印在我心中。

也要感谢自己，感谢这个不曾轻言放弃的自己。在挫折面前坚持，在成功时保持谦虚，一路跌跌撞撞却始终向前。

点滴之恩，难以言表；长风破浪会有时，唯愿以筑梦之帆，横越学海，以今日所学，报昨日之恩，如春蚕吐丝，蜡炬成灰，将所得馈赠桑梓，以求无愧于心。

参考文献

- [1] Jiang R, Woo C W, Qi S, et al. Interpreting brain biomarkers: Challenges and solutions in interpreting machine learning-based predictive neuroimaging[J]. IEEE signal processing magazine, 2022, 39(4): 107-118.
- [2] Nisar S, Haris M. Neuroimaging genetics approaches to identify new biomarkers for the early diagnosis of autism spectrum disorder[J]. Molecular psychiatry, 2023, 28(12): 4995-5008.
- [3] 韩慧玲, 张杉杉, 陈芙蓉, 等. 注意缺陷多动障碍发病机制及治疗药物研究进展[J]. 中国药业, 2023, 32(21): 149-155.
- [4] Sharma A, Kala S, Kumar A, et al. Deep Learning in Genomics, Personalized Medicine, and Neurodevelopmental Disorders[J]. Intelligent Data Analytics for Bioinformatics and Biomedical Systems, 2024: 235-264.
- [5] Song D Y, Topriceanu C C, Ilie-Ablachim D C, et al. Machine learning with neuroimaging data to identify autism spectrum disorder: a systematic review and meta-analysis[J]. Neuroradiology, 2021, 63: 2057-2072.
- [6] Yin W, Mostafa S, Wu F X. Diagnosis of autism spectrum disorder based on functional brain networks with deep learning[J]. Journal of Computational Biology, 2021, 28(2): 146-165.
- [7] Han J, Jiang G, Ouyang G, et al. A multimodal approach for identifying autism spectrum disorders in children[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2022, 30: 2003-2011.
- [8] Morra J H, Tu Z, Apostolova L G, et al. Comparison of AdaBoost and support vector machines for detecting Alzheimer's disease through automated hippocampal segmentation[J]. IEEE transactions on medical imaging, 2009, 29(1): 30-43.
- [9] Abraham A, Milham M P, Di Martino A, et al. Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example[J]. NeuroImage, 2017, 147: 736-745.

- [10] Lee S H, Yu D, Bachman A H, et al. Application of fused lasso logistic regression to the study of corpus callosum thickness in early Alzheimer's disease[J]. *Journal of neuroscience methods*, 2014, 221: 78-84.
- [11] Sapey-Triomphe L A, Temmerman J, Puts N A J, et al. Prediction learning in adults with autism and its molecular correlates[J]. *Molecular autism*, 2021, 12: 1-17.
- [12] Larson C, Thomas H R, Crutcher J, et al. Language Networks in Autism Spectrum Disorder: A systematic review of connectivity-based fMRI studies[J]. *Review Journal of Autism and Developmental Disorders*, 2023: 1-28.
- [13] Bayram M A, Özer İ, Temurtaş F. Deep learning methods for autism spectrum disorder diagnosis based on fMRI images[J]. *Sakarya University Journal of Computer and Information Sciences*, 2021, 4(1): 142-155.
- [14] Khosla M, Jamison K, Kuceyeski A, et al. Ensemble learning with 3D convolutional neural networks for functional connectome-based prediction[J]. *NeuroImage*, 2019, 199: 651-662.
- [15] Tang Y, Chen D, Wu J, et al. Functional connectivity learning via Siamese-based SPD matrix representation of brain imaging data[J]. *Neural Networks*, 2023, 163: 272-285.
- [16] Salim I, Hamza A B. Classification of developmental and brain disorders via graph convolutional aggregation[J]. *Cognitive Computation*, 2024, 16(2): 701-716.
- [17] Zhu Q, Yang Q, Wang M, et al. Multi-discriminator active adversarial network for multi-center brain disease diagnosis[J]. *IEEE Transactions on Big Data*, 2023, 9(6): 1575-1585.
- [18] Wang Z, Xu Y, Peng D, et al. Brain functional activity-based classification of autism spectrum disorder using an attention-based graph neural network combined with gene expression[J]. *Cerebral Cortex*, 2023, 33(10): 6407-6419.
- [19] Ajith M, Aycok D M, Tone E B, et al. A deep learning approach for mental health quality prediction using functional network connectivity and assessment data[J]. *Brain Imaging and Behavior*, 2024, 18(3): 630-645.
- [20] Sun Y, Jin W, Si X, et al. Continuous seizure detection based on transformer and long-term iEEG[J]. *IEEE Journal of Biomedical and Health Informatics*, 2022, 26(11): 5418-

5427.

- [21] Zhu H, Wang J, Zhao Y P, et al. Contrastive multi-view composite graph convolutional networks based on contribution learning for autism spectrum disorder classification[J]. IEEE Transactions on Biomedical Engineering, 2022, 70(6): 1943-1954.
- [22] Wang Z, Xu Y, Peng D, et al. Brain functional activity-based classification of autism spectrum disorder using an attention-based graph neural network combined with gene expression[J]. Cerebral Cortex, 2023, 33(10): 6407-6419.
- [23] Liang Y, Liu B, Zhang H. A convolutional neural network combined with prototype learning framework for brain functional network classification of autism spectrum disorder[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2021, 29: 2193-2202.
- [24] Pramono R R A, Chen Y T, Fang W H. Spatial-temporal action localization with hierarchical self-attention[J]. IEEE Transactions on Multimedia, 2021, 24: 625-639.
- [25] Xu L, Chen C L P, Han R. Graph-based sparse bayesian broad learning system for semi-supervised learning[J]. Information Sciences, 2022, 597: 193-210.
- [26] Cole M W, Ito T, Cocuzza C, et al. The functional relevance of task-state functional connectivity[J]. Journal of Neuroscience, 2021, 41(12): 2684-2702.
- [27] 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.
- [28] Hu S, Lou Z, Yan X, et al. A survey on information bottleneck[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(8): 5325-5344.
- [29] Tang P, Peng K, Dong J. Nonlinear quality-related fault detection using combined deep variational information bottleneck and variational autoencoder[J]. ISA transactions, 2021, 114: 444-454.
- [30] Chen L, Li Y, Lei Y, et al. Metarelation2vec: A Metapath-Free Scalable Representation Learning Model for Heterogeneous Networks[J]. Tsinghua Science and Technology, 2023, 29(2): 553-575.
- [31] Bing R, Yuan G, Zhu M, et al. Heterogeneous graph neural networks analysis: a survey of techniques, evaluations and applications[J]. Artificial Intelligence Review, 2023,

- 56(8): 8003-8042.
- [32] Zhao L, Dong Q, Luo C, et al. DeepOmix: a scalable and interpretable multi-omics deep learning framework and application in cancer survival analysis[J]. Computational and structural biotechnology journal, 2021, 19: 2719-2725.
- [33] Alsubaie M G, Luo S, Shaukat K. Alzheimer's disease detection using deep learning on neuroimaging: a systematic review[J]. Machine Learning and Knowledge Extraction, 2024, 6(1): 464-505.
- [34] Toga A W, Thompson P M, Mori S, et al. Towards multimodal atlases of the human brain[J]. Nature Reviews Neuroscience, 2006, 7(12): 952-966.
- [35] Logothetis N K. What we can do and what we cannot do with fMRI[J]. Nature, 2008, 453(7197): 869-878.
- [36] Chen J J, Uthayakumar B, Hyder F. Mapping oxidative metabolism in the human brain with calibrated fMRI in health and disease[J]. Journal of Cerebral Blood Flow & Metabolism, 2022, 42(7): 1139-1162.
- [37] Drew P J. Neurovascular coupling: motive unknown[J]. Trends in neurosciences, 2022, 45(11): 809-819.
- [38] Lu Y, Cui Y, Cao L, et al. Macaque Brainnetome Atlas: A multifaceted brain map with parcellation, connection, and histology[J]. Science Bulletin, 2024, 69(14): 2241-2259.
- [39] Li Y, Liu A, Li L, et al. Connectivity-based brain parcellation for parkinson's disease[J]. IEEE Transactions on Biomedical Engineering, 2022, 70(5): 1539-1552.
- [40] Evans A C, Janke A L, Collins D L, et al. Brain templates and atlases[J]. Neuroimage, 2012, 62(2): 911-922.
- [41] Sorrentino P, Petkoski S, Sparaco M, et al. Whole-brain propagation delays in multiple sclerosis, a combined tractography-magnetoencephalography study[J]. Journal of Neuroscience, 2022, 42(47): 8807-8816.
- [42] Rolls E T, Joliot M, Tzourio-Mazoyer N. Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas[J]. Neuroimage, 2015, 122: 1-5.
- [43] Schaefer A, Kong R, Gordon E M, et al. Local-global parcellation of the human cerebral

- cortex from intrinsic functional connectivity MRI[J]. *Cerebral cortex*, 2018, 28(9): 3095-3114.
- [44] Yeo B T T, Krienen F M, Sepulcre J, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity[J]. *Journal of neurophysiology*, 2011, 106: 1125-1165.
- [45] Fotiadis P, Parkes L, Davis K A, et al. Structure–function coupling in macroscale human brain networks[J]. *Nature Reviews Neuroscience*, 2024, 25(10): 688-704.
- [46] Huang C C, Rolls E T, Feng J, et al. An extended Human Connectome Project multimodal parcellation atlas of the human cortex and subcortical areas[J]. *Brain Structure and Function*, 2022, 227(3): 763-778.
- [47] Demeter D V, Greene D J. The promise of precision functional mapping for neuroimaging in psychiatry[J]. *Neuropsychopharmacology*, 2024: 1-13.
- [48] Hempel M, Barnhofer T, Domke A K, et al. Aberrant associations between neuronal resting-state fluctuations and working memory-induced activity in major depressive disorder[J]. *Molecular Psychiatry*, 2025, 30(1): 4-12.
- [49] Ye Z, Kumar Y J, Sing G O, et al. A comprehensive survey of graph neural networks for knowledge graphs[J]. *IEEE Access*, 2022, 10: 75729-75741.
- [50] Zhou J, Cui G, Hu S, et al. Graph neural networks: A review of methods and applications[J]. *AI open*, 2020, 1: 57-81.
- [51] Khoshraftar S, An A. A survey on graph representation learning methods[J]. *ACM Transactions on Intelligent Systems and Technology*, 2024, 15(1): 1-55.
- [52] Kratsios A, Papon L. Universal approximation theorems for differentiable geometric deep learning[J]. *Journal of Machine Learning Research*, 2022, 23(196): 1-73.
- [53] Liu H, Yang B, Li D. Graph collaborative filtering based on dual-message propagation mechanism[J]. *IEEE Transactions on Cybernetics*, 2021, 53(1): 352-364.
- [54] Qureshi S. Limits of depth: Over-smoothing and over-squashing in GNNs[J]. *Big Data Mining and Analytics*, 2023, 7(1): 205-216.
- [55] Wang H, Wang J, Wang J, et al. Graphgan: Graph representation learning with generative adversarial nets[C]//*Proceedings of the AAAI conference on artificial*

- intelligence. 2018, 32(1): 2508-2515.
- [56] Yang C, Xiao Y, Zhang Y, et al. Heterogeneous network representation learning: A unified framework with survey and benchmark[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(10): 4854-4873.
- [57] Wang X, Ji H, Shi C, et al. Heterogeneous graph attention network[C]//The world wide web conference. 2019: 2022-2032.
- [58] Zhu S, Wang X, Lai S, et al. Multi-graph aggregated graph neural network for heterogeneous graph representation learning[J]. International Journal of Machine Learning and Cybernetics, 2025, 16(2): 803-818.
- [59] Hu Z, Dong Y, Wang K, et al. Heterogeneous graph transformer[C]//Proceedings of the web conference 2020. 2020: 2704-2710.
- [60] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Advances in Neural Information Processing Systems. 2017;30:5998-6008.
- [61] Yun S, Jeong M, Yoo S, et al. Graph Transformer Networks: Learning meta-path graphs to improve GNNs[J]. Neural networks, 2022, 153: 104-119.
- [62] Sun C, Li C, Lin X, et al. Attention-based graph neural networks: a survey[J]. Artificial intelligence review, 2023, 56(Suppl 2): 2263-2310.
- [63] Chen C, Wu Y, Dai Q, et al. A survey on graph neural networks and graph transformers in computer vision: A task-oriented perspective[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 10297-10318.
- [64] Wang Z, Wang Y, Yuan C, et al. Empirical analysis of performance bottlenecks in graph neural network training and inference with GPUs[J]. Neurocomputing, 2021, 446: 165-191.
- [65] Yuan Q, Chen J, Zhao H, et al. Structure-aware protein-protein interaction site prediction using deep graph convolutional network[J]. Bioinformatics, 2022, 38(1): 125-132.
- [66] Li J, Sun Q, Zhang F, et al. Meta-structure-based graph attention networks[J]. Neural Networks, 2024, 171: 362-373.
- [67] Farrant K, Uddin L Q. Atypical developmental of dorsal and ventral attention networks

- in autism[J]. *Developmental science*, 2016, 19(4): 550-563.
- [68] Attanasio M, Mazza M, Le Donne I, et al. Salience Network in Autism: preliminary results on functional connectivity analysis in resting state[J]. *European Archives of Psychiatry and Clinical Neuroscience*, 2024: 1-14.
- [69] Sato M, Nakai N, Fujima S, et al. Social circuits and their dysfunction in autism spectrum disorder[J]. *Molecular Psychiatry*, 2023, 28(8): 3194-3206.
- [70] Uddin L Q. Cognitive and behavioural flexibility: neural mechanisms and clinical considerations[J]. *Nature Reviews Neuroscience*, 2021, 22(3): 167-179.
- [71] Harikumar A, Evans D W, Dougherty C C, et al. A review of the default mode network in autism spectrum disorders and attention deficit hyperactivity disorder[J]. *Brain connectivity*, 2021, 11(4): 253-263.

攻读硕士期间发表的论文及获奖情况

硕士期间发表的学术论文

- [1] **Chen Lei**, Dong Wenhao, Zeng Weiming, et al. Recognition of Autism Spectrum Disorder via Dual Channel Feature Fusion and Multi-Resolution Brain Region Segmentation [C]//2024 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML). IEEE, 2024. (EI 会议, 已录用)
- [2] Li Y, Zeng W, Dong W, D Han, **L Chen**, et al. A tale of single-channel electroencephalogram: Devices, datasets, signal processing, applications, and future directions[J]. IEEE Transactions on Instrumentation and Measurement, 2025.

获奖情况

- [1] “华为杯”第二十一届中国研究生数学建模竞赛 全国二等奖
- [2] 2023 年研究生新生入学奖学金二等奖
- [3] 2024 年研究生学业奖学金二等奖