

蔡志宏, 曾安, 潘丹, 等. 双模态迭代交叉注意力融合集成框架[J]. 广东工业大学学报, 2026, 43(2): 1-11. doi: 10.12052/gdutxb.250002.
Cai Zhihong, Zeng An, Pan Dan, et al. Bimodal iterative cross-attention fusion ensemble framework[J]. Journal of Guangdong University of Technology, 2026, 43(2): 1-11. doi: 10.12052/gdutxb.250002.

双模态迭代交叉注意力融合集成框架

蔡志宏¹, 曾安¹, 潘丹², 叶嘉宇¹

(1. 广东工业大学 计算机学院, 广东 广州 510006; 2. 广东技术师范大学 电子与信息学院, 广东 广州 510665)

摘要: 阿尔茨海默症(Alzheimer's Disease, AD)作为一种进行性神经退行性疾病,其早期诊断与临床干预始终面临重大挑战。在医学影像领域中,结构性磁共振成像(Structural Magnetic Resonance Imaging, sMRI)通过高分辨率解剖成像精准捕捉脑萎缩等结构性改变,而氟脱氧葡萄糖正电子发射断层扫描(Fluorodeoxyglucose Positron Emission Tomography, FDG-PET)则通过监测葡萄糖代谢水平有效反映脑功能变化,二者在AD相关脑部病理改变的检测中具有重要互补价值。然而,现有的AD多模态分类模型存在特征融合效果欠佳、模态间信息交互不充分以及特征分布不一致的问题,限制了其在AD诊断中的应用。为了解决这些问题,提出了双模态迭代交叉注意力融合集成框架(Bimodal Iterative Cross-Attention Fusion Ensemble Framework, BICAFEF)。该框架由基分类器和元分类器组成。基分类器通过残差网络(Residual Network, ResNet)模块提取sMRI和FDG-PET切块的特征,并设计基于卷积操作和自适应聚合池化操作的空间特征收缩(Spatial Feature Shrinking, SFS)模块减少模态间的冗余信息,突出关键特征。同时,构建迭代交叉注意力机制,在多轮迭代中动态捕捉和强化模态间的全局依赖关系和互补信息,解决无法充分挖掘模态间互补特性的难题,从而提升AD分类性能。最终,为了提升全脑分类精度,框架构建了元分类器对基分类器进行筛选和集成,剔除准确率低于75%的分类器,保留高性能分类器,进一步提高分类的鲁棒性和准确性。此外,通过可视化分析进一步验证了框架对关键脑区的关注,展现出该框架在sMRI和PET模态下对AD相关病变区域的有效识别能力。实验结果表明,该框架在ADvs.HC(正常组)中的五折分类准确率(Accuracy, ACC)为94.3%,敏感度(Sensitivity, SEN)为92.6%,特异度(Specificity, SPE)为96.3%,ROC曲线下面积(Area Under Curve, AUC)为97.5%,马修斯相关系数(Matthews correlation coefficient, MCC)为88.7%,优于现有先进的同类框架。

关键词: 阿尔茨海默症; 多模态融合; 迭代学习; 交叉注意力机制; 分类

中图分类号: TP391

文献标志码: A

文章编号: 1007-7162(2026)02-0001-11

Bimodal Iterative Cross-attention Fusion Ensemble Framework

Cai Zhihong¹, Zeng An¹, Pan Dan², Ye Jiayu¹

(1. School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China; 2. School of Electronics and Information Technology, Guangdong Technical Normal University, Guangzhou 510665, China)

Abstract: Alzheimer's disease (AD), as a progressive neurodegenerative disorder, presents significant challenges in early diagnosis and clinical intervention. In medical imaging, structural magnetic resonance imaging (sMRI) captures brain atrophy and structural alterations through high-resolution anatomical imaging, while fluorodeoxyglucose positron emission tomography (FDG-PET) effectively reflects functional changes by monitoring cerebral glucose metabolism. These two modalities hold complementary value in detecting AD-related pathological brain changes. However, existing multimodal AD classification models are limited by suboptimal feature fusion, insufficient inter-modal information interaction, and feature distribution discrepancies, hindering their diagnostic utility. To address these issues, a bimodal iterative cross-attention fusion ensemble framework (BICAFEF) is proposed. This framework comprises base classifiers and a meta-classifier. The base classifiers employ ResNet modules to extract features from sMRI and FDG-PET image patches. A spatial feature shrinking

收稿日期: 2025-01-05 录用日期: 2025-03-05 网络首发日期: 2025-06-17

基金项目: 国家自然科学基金资助项目(61976058); 广州市科技计划项目(202103000034, 202206010007, 202002020090); 广东省科技计划项目(2021A1515012300, 2019A050510041, 2021B0101220006)

作者简介: 蔡志宏(1999-), 男, 硕士研究生, 主要研究方向为深度学习、图像处理, E-mail: 2112205101@mail2.gdut.edu.cn

通信作者: 曾安(1978-), 女, 教授, 博士生导师, 主要研究方向为图像处理、模式识别、人工智能, E-mail: zengan@gdut.edu.cn

(SFS) module, integrating convolutional operations and adaptive aggregation pooling, is designed to reduce inter-modal redundancy and emphasize discriminative features. Additionally, an iterative cross-attention mechanism is constructed to dynamically capture and reinforce global dependencies and complementary information across modalities through multi-round iterations, thereby resolving the challenge of insufficiently exploiting inter-modal synergies and enhancing AD classification performance. To further improve whole-brain classification accuracy, the framework incorporates a meta-classifier to screen and ensemble base classifiers by discarding those with accuracy below 75%, retaining high-performance classifiers to boost robustness and precision. Visualization analyses validate the framework's focus on critical brain regions, demonstrating its capability to effectively identify AD-related pathological areas in sMRI and PET modalities. Experimental results show that the framework achieves a five-fold classification accuracy (ACC) of 94.3%, sensitivity (SEN) of 92.6%, specificity (SPE) of 96.3%, AUC of 97.5%, and Matthews correlation coefficient (MCC) of 88.7% in AD vs. healthy control (HC) classification, outperforming state-of-the-art multimodal frameworks.

Key words: Alzheimer's disease (AD); multimodal fusion; iterative learning; cross-attention mechanism; classification

阿尔茨海默症(Alzheimer's Disease, AD)是一种由于大脑细胞代谢异常引起的不可逆神经退行性疾病^[1]。研究表明,2019年全球AD患者人数为5 500万,预计到2050年这一数字将增至1.39亿。目前,我国60岁及以上老年人中有983万AD患者,数量居全球之首^[2]。AD的发展会导致患者脑区萎缩或脑区代谢活性的退化^[3]。轻度认知障碍(Mild Cognitive Impairment, MCI)^[4]是介于健康(Healthy Control, HC)和AD之间的过渡状态。MCI患者较健康人群更有可能进展为AD,每年转变为AD的几率高达10%至15%。因此,深入研究健康人群、MCI患者和AD患者之间的差异,对于实现AD的早期诊断至关重要。

在AD诊断中,结构性磁共振成像(Structural Magnetic Resonance Imaging, sMRI)常用于非侵入性地捕捉局部脑萎缩现象,能够高分辨率地显示大脑结构细节,从而协助分析大脑的解剖学变化。该技术在AD诊断的模式识别研究中应用广泛,因为其提供的丰富解剖学信息有助于早期发现和评估疾病的进展^[5-6]。另外,氟脱氧葡萄糖正电子发射断层扫描(Fluorodeoxyglucose Positron Emission Tomography, FDG-PET)通过区域葡萄糖摄取程度表示组织代谢活动,提供了脑部代谢和活动信息。因此,PET在AD诊断中也起着重要作用,特别是在评估大脑功能变化方面^[7]。

近期研究表明,通过融合sMRI的大脑结构信息和PET的大脑功能信息能够有效提高模型在AD诊断中的性能。如Huang等^[8]提出了一种基于卷积循环神经网络的多模态数据分类框架,结合了二维卷积神经网络和循环神经网络,以学习三维MRI和三维

PET图像切分为二维切片序列之后的切片内、切片间特征,完成AD的早期诊断。Liu等^[9]提出了基于分层注意力的多任务多模态融合模型(Hierarchical Attention-based Multi-task Multi-modal Fusion Model, HAMMF),利用sMRI和PET影像数据,通过上下文分层注意力模块(Contextual Hierarchical Attention Module, CHAM)从各个维度提取细粒度的病理特征,利用Transformer捕捉多模态输入的相关特征,并通过多任务学习结合主要分类任务和次要回归任务,实现AD分类。

尽管多模态研究总体上表现优异,但基于多模态数据的AD诊断仍存在一些挑战。首先,目前许多多模态AD神经影像识别方法通过简单连接不同模态的特征来实现信息融合,然而这对于探索模态之间的互补信息并非最优的,这种方式未能充分挖掘模态之间的互补信息,忽视了不同模式特征分布不一致的问题,并可能会影响特征融合和诊断性能。此外,多模态特征存在维度较高、噪声较大等问题,如何平衡不同模态的重要性并高效实现模态间特征交互,仍是亟待解决的关键问题。

为了解决这些问题,本文提出了双模态迭代交叉注意力融合集成框架(Bimodal Iterative Cross-Attention Fusion Ensemble Framework, BICAFFEF)。该框架通过一种新型的迭代交叉注意力融合策略来弥补现有方法的不足。具体而言,迭代交叉注意力融合机制可以动态捕捉不同模态特征间的全局依赖关系,并通过迭代更新不断加强模态间的特征交互和信息补充。相比传统的简单特征拼接,该方法能够更有效地平衡和融合模态间的特征差异,减少特征冗

余和噪声,从而更高效、更准确地学习模态间的互补信息,为AD诊断提供更强大的支持。

本文的主要贡献如下:

(1) 提出BICAFEF,以双模态迭代交叉注意力融合模型(Bimodal Iterative Cross-Attention Fusion Model, BICAF)为基分类器,结合元分类器实现基于sMRI和PET的AD病情诊断。

(2) 提出迭代交叉注意力模块(Iterative Cross-Attention, ICA),以动态捕捉sMRI和PET模态间的全局依赖关系,并通过迭代优化不断加强不同模态之间的特征交互。相比传统的简单拼接方法,该方法能更好地平衡和融合不同模态之间的特征差异,从而更精确地学习模态间的互补信息。

(3) 实验数据表明,该框架在AD分类任务中优于现有的先进同类框架。可视化分析进一步验证了框架对AD关键脑区的关注,展示了其在sMRI和PET模态下对AD相关病变区域的有效识别能力,从而验证了本文框架的有效性。

本文第1节回顾了近期在AD多模态诊断方面的研究,第2节提出了一个基于一种新型的迭代交叉注意力融合策略的集成框架来进行AD诊断。第3节,使用真实数据描述了实验过程和实验结果,并且进行了消融实验探讨各模块的重要性。第4节对文章进行总结。

1 相关工作

尽管单模态生物标志物如sMRI、PET、功能性磁共振成像(Functional Magnetic Resonance Imaging, fMRI)和脑脊液(Cerebrospinal Fluid, CSF)等在AD诊断和进展评估中取得了一定成效,但由于单一模态传递的信息往往是片面的,不可能全面地模拟大脑的病理特征,所以单模态方法往往会遇到AD诊断性能的瓶颈^[10]。为弥补这些不足,许多研究已经专注于使用多模态数据进行AD分类。通过融合多模态生物标志物来实现AD诊断的研究可以分为两类,即传统机器学习方法和深度学习方法。

机器学习方法在多模态AD诊断中的研究通常包括2个部分:特征提取和分类。例如Lin等^[11]使用线性判别分析(Linear Discriminant Analysis, LDA)对包括sMRI、FDG-PET、CSF和基因的多模态数据进行评分,得到不同方式下代表AD病理进展的分数,从而实现了AD的多类诊断。Khatri等^[12]设计了一种过滤器和包器相结合的特征选择方法,通过多种生物标志物实现AD诊断,提高了AD的诊断准确率。

Walhovd等^[13]将逻辑回归分析应用于sMRI、FDG-PET和CSF数据的组合特征,以实现AD分类,同时研究哪些标志物对疾病分类最敏感。Kohannim等^[14]将机器学习分类器应用于脑成像和其他生物标志物的结合特征,以进行AD诊断。

在过去的几十年里,深度学习在计算机视觉领域取得了杰出的成就,尤其是利用卷积神经网络(Convolutional Neural Network, CNN)来学习特征。随着CNN的快速发展,各种神经网络为多模态AD的诊断提供多种思路。例如Zhang等^[15]提出了一种多模态交叉注意力AD诊断(Multi-modal Cross-attention AD Diagnosis, MCAD)框架,该框架使用sMRI、FDG-PET和CSF生物标志物的多模态数据,通过交叉模态注意力整合成像和非成像信息再进行AD分类。Golovanevsky等^[16]提出了一种多模态AD诊断框架(Multimodal Alzheimer's Disease Diagnosis framework, MADDi),使用跨模态注意力捕获模态之间的交互。根据影像、遗传和临床数据准确检测AD和MCI的存在。Sharma等^[17]提出了双分支三维卷积神经网络模型,该模型的每个分支由4个三维卷积层和实例归一化层组成,然后将2种模态的特征进行拼接,拼接后的特征输入到两个额外的卷积层进行处理,然后经过全局平均池化层操作,最终进行AD分类。

2 双模态迭代交叉注意力融合集成框架(BICAFEF)

BICAFEF的整体结构如图1(a)所示,由150个基分类器即BICAF和元分类器组成。BICAF(Bimodal Iterative Cross-Attention Fusion)如图1(b)所示,是一个双分支骨干网络,由单模态特征提取、双模态特征融合和全连接层三部分组成,BICAF以预处理后的sMRI和FDG-PET的切块 $\{B_M^i, B_P^i\}_{i=0, \dots, 149}$ 作为输入。首先经过残差网络(Residual Network, ResNet)模块进行特征提取得到三维特征向量 $\{F_M^i, F_P^i\}_{i=0, \dots, 149}$,接着经过空间特征收缩(Spatial Feature Shrinking, SFS)模块进行特征收缩以减少特征中的冗余信息,突出关键特征,同时减少后续的计算成本得到二维特征向量 $\{V_M^i, V_P^i\}_{i=0, \dots, 149}$,随后使用ICA(如图1(d)所示)进行特征增强,以加强sMRI和PET模态之间的特征交互,得到增强后的特征向量 $\{\hat{F}_M^i, \hat{F}_P^i\}_{i=0, \dots, 149}$ 。ICA允许框架在每一层迭代中将一个模态的特征作为查询,另一个模态的特征作为键和值,通过注意力矩阵来捕捉不同模态间的全局依赖关系。这一过程在多层次迭代中进行,以不断加强模态间和模态内

的互补信息,从而更好地学习模态间的特征交互。接着将 $\{\hat{F}_M^i, \hat{F}_P^i\}_{i=0, \dots, 149}$ 与 $\{F_M^i, F_P^i\}_{i=0, \dots, 149}$ 连接起来经过全连接层进行分类,得到在验证集上的分类结果 $\{\text{Acc}_i\}_{i=0, \dots, 149}$ 。最后为了提升全脑分类精度,框架构建

元分类器,对基分类器进行筛选和集成。元分类器由一个卷积层和一个全连接层组成,输入为准确率大于75%的基分类器的倒数第二层特征输出,在标签的指导下在训练集中训练元分类器,得到最终分类结果。

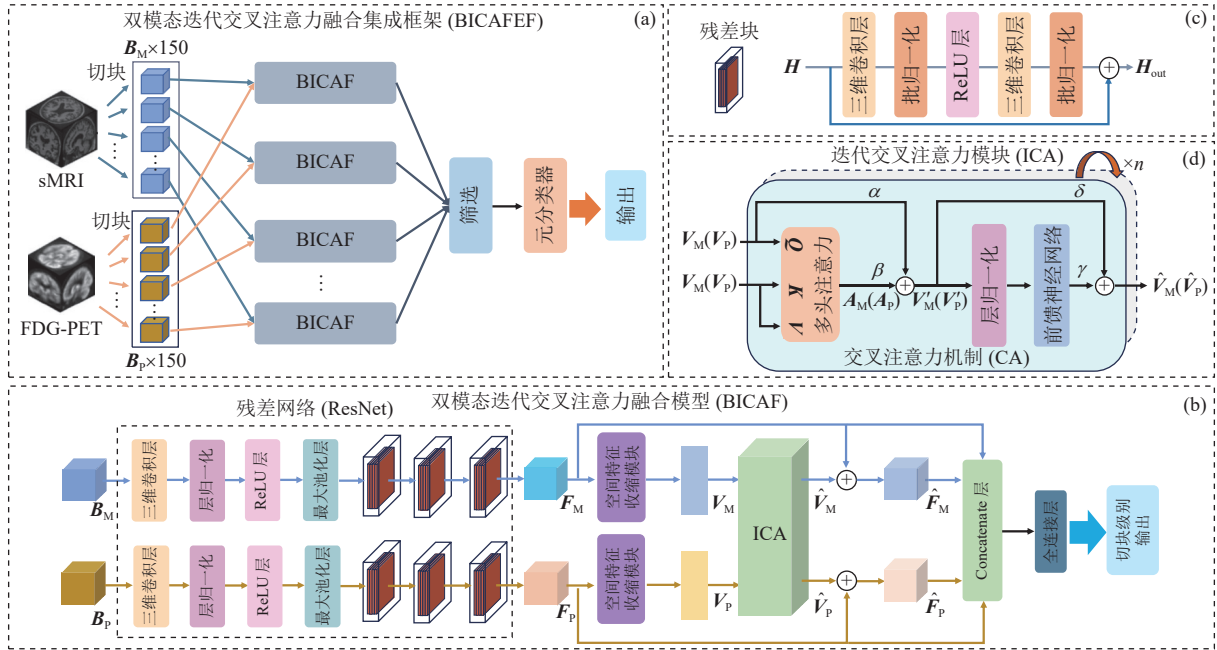


图1 整体框架

Fig.1 Overall Framework

2.1 双模态迭代交叉注意力融合模型(BICAF)

2.1.1 残差网络(ResNet) 模块

ResNet是基于ResNet-18架构的三维卷积神经网络,相比标准的3D ResNet-18^[18],本文对网络进行了裁剪和优化,使其更适用于小规模切块数据集,并在保持足够的特征能力的同时有效降低了模型的计算开销和防止过拟合。ResNet由初始卷积层和残差块组成。以sMRI分支为例,初始卷积层输入为 $B_M \in \mathbb{R}^{25 \times 25 \times 25 \times 1}$,其中 $25 \times 25 \times 25$ 为sMRI的空间维度 $H \times W \times D$,1表示通道数, B_M 首先经过一个具有64个滤波器、核大小为 $7 \times 7 \times 7$ 、步幅为2且填充为same的三维卷积操作,可表示为

$$H_0 = \text{ReLU}(\text{LN}(\text{Conv3D}(P_M, 64, 7, 2, \text{same}))) \quad (1)$$

其中, LN表示层归一化(Layer Normalization, LN)。然后进行最大池化操作,池化窗口为 $3 \times 3 \times 3$,步幅为2,可表示为

$$H_1 = \text{MaxPool3D}(H_0, 3, 2, \text{same}) \quad (2)$$

接着,经过3个残差块,可表示为 $f_{\text{res}}(H, C_{\text{out}}, s)$ 。其中, H 为输入的特征向量, C_{out} 为输出特征通道数, s 为卷积的步幅。如图1(c)所示, H 首先经过一个三维卷积

层,其输出特征通道数为 C_{out} ,核大小为 $3 \times 3 \times 3$,步幅为 s ,随后经过批归一化(Batch Normalization, BN)和ReLU激活函数,接着经过输出特征通道数为 C_{out} 、核大小为 $3 \times 3 \times 3$ 、步幅为1的三维卷积层,然后再经过一个BN得到特征向量 Y 。最后进行残差连接得到最终输出 H_{out} ,公式为

$$H_{\text{out}} = \text{ReLU}(Y + S) \quad (3)$$

$$S = \begin{cases} \text{BN}(\text{Conv3D}(H, C_{\text{out}}, 1, s), & \text{if } s \neq 1 \text{ or } C_H \neq C_{\text{out}} \\ H, & \text{otherwise} \end{cases}$$

其中, C_H 表示 H 的通道数。

为了使ResNet从特征提取的初期到高阶特征的聚合,实现逐渐增加特征表达能力的目的,3个残差块分别使用了不同的 C_{out} 和 s ,表示为

$$H_2 = f_{\text{res}}(H_1, 64, 1)$$

$$H_3 = f_{\text{res}}(H_2, 128, 2)$$

$$F_M = f_{\text{res}}(H_3, 256, 2)$$

其中, H_2 为第1个残差块的输出, H_3 为第2个残差块的输出, F_M 为第3个残差块的输出,即ResNet的最终

输出。同理,在PET分支中经过ResNet将得到PET的特征向量 F_p 。

对ResNet模块进行空间和时间复杂度分析:

(1) 空间复杂度:对于初始卷积层,其参数空间复杂度为 $O(k^3 \cdot C_{in} \cdot C_{out})$,其中 k 为固定核尺寸, C_{in} 和 C_{out} 分别为输入与输出通道数;层归一化参数随通道数线性增长,即 $O(C_{in})$ 。残差块的参数量取决于主路径的两次 $3 \times 3 \times 3$ 卷积及可能的 $1 \times 1 \times 1$ 捷径卷积,其复杂度为 $O(C_{in}^2 + C_{in} \cdot C_{out})$ 。当通道数逐层倍增时(如 $C_{out} = 2C_{in}$),整体空间复杂度简化为 $O(C_{in}^2)$ 。

(2) 时间复杂度:初始卷积层的计算量随输入空间尺寸 D_{in} 与通道数呈立方关系,即 $O(D_{in}^3 \cdot C_{in} \cdot C_{out})$ 。残差块的主路径计算量由2次 $3 \times 3 \times 3$ 卷积主导,其复杂度为 $O(D_{in}^3 \cdot C_{in}^2)$,捷径卷积的计算量则为 $O(D_{in}^3 \cdot C_{in} \cdot C_{out})$ 。当网络深度增加且通道数逐层扩展时,总时间复杂度为 $O(L \cdot D_{in}^3 \cdot C_{max}^2)$,其中 L 为残差块数量, C_{max} 为最大通道数。

2.1.2 空间特征收缩模块(SFS)

为了进一步减少特征中的冗余信息,突出关键特征,并减少后续计算成本,在ICA模块之前应用了SFS模块^[19]。SFS包含2个部分:池化操作和卷积操作。对于卷积操作,首先使用了一种基于卷积运算进行降维的方法,如式(4)所示。

$$F_{conv} = \text{conv}_{1 \times 1}(\text{Reshape}(F)) \quad (4)$$

具体来说,通过重塑特征图的维度将特征的空间信息转换为通道维度,然后用 1×1 卷积操作压缩通道维度。其中 F 表示输入特征图,conv表示通过 1×1 卷积压缩特征图。

对于池化操作,传统的平均池化和最大池化方法各有优缺点。最大池化更关注局部的显著特征(如边缘和纹理),但可能会丢失背景信息;平均池化能更平滑地保留整体信息,但容易丢细节。因此,SFS采用了一种自适应聚合平均池化和最大池化的方法,通过自适应聚合这两种方法,可以同时保留显著特征和全局背景信息,从而增强网络的特征表达能力。公式为

$$F_a = \text{AvgPooling}(F, S) \quad (5)$$

$$F_m = \text{MaxPooling}(F, S) \quad (6)$$

$$F_o = \mu \cdot F_a + (1 - \mu) \cdot F_m \quad (7)$$

其中, F 表示输入的特征图, S 表示特征图的比例因子, F_a 和 F_m 分别表示通过AvgPooling(\cdot)和

MaxPooling(\cdot)压缩的特征图, μ 为0和1之间的权重,是可学习参数。

在SFS模块中,空间复杂度和时间复杂度主要由其核心的卷积操作与自适应池化操作决定。对于卷积部分,采用 1×1 卷积对特征图的通道维度进行压缩,其时间复杂度为 $O(C_{in} \cdot C_{out} \cdot H \cdot W)$,其中 C_{in} 和 C_{out} 分别表示输入和输出的通道数, H 和 W 为特征图的空间尺寸。空间复杂度则与卷积核参数数量相关,为 $O(C_{in} \cdot C_{out})$ 。池化操作通过并行计算平均池化和最大池化,并引入可学习的权重因子进行自适应融合,其时间复杂度为 $O(H \cdot W \cdot C)$,其中 C 表示池化操作中处理的通道数。空间复杂度仅增加一个可学习的标量参数,即 $O(1)$ 。由于重塑操作仅改变张量维度而不涉及计算,其对复杂度的影响可忽略。因此,SFS模块的时间复杂度为 $O((C_{in} \cdot C_{out} + C) \cdot H \cdot W)$,空间复杂度为 $O(C_{in} \cdot C_{out} + 1)$ 。

2.1.3 迭代交叉注意力模块(ICA)

ICA(见图1(d))将一种新颖的迭代学习策略^[19]与交叉注意力(Cross-Attention, CA)模块结合起来,以加强来自模态间和模态内特征互补信息的记忆,进一步提高模型的性能。与之前捕获不同模态的局部特征的研究不同,所提出的CA使单模态能够在全局角度从辅助模态中学习更多的互补信息。CA由多头注意力模块、LN和前馈神经网络(Feedforward Neural Network, FFN)组成。给定输入特征图 F_M 和 $F_P \in \mathbb{R}^{H \times W \times D \times C}$,本文首先将每个特征图展平为一组token,并添加一个可学习的位置嵌入,这是一个维度为 $H \times W \times D \times C$ 的可训练参数,用于编码不同token之间的空间信息。随后,得到一组带有位置嵌入的sMRI和PET的token: V_M 和 V_P ,并将其作为CA的输入。由于sMRI和PET特征在空间表达上通常存在差异,本文采用双CA分别提取互补信息,以增强sMRI和PET特征。为了简洁起见,本文仅对sMRI分支的CA模块进行说明。

对于sMRI分支的CA模块,首先将sMRI的token V_M 投影到单独的矩阵 $Q_M \in \mathbb{R}^{H \times W \times D \times C}$,PET的token V_P 投影到2个单独的矩阵 $K_P, V_P \in \mathbb{R}^{H \times W \times D \times C}$ 中进行多头注意力计算,得到特征向量 A_M ,如公式(8)所示

$$A_M = \text{Concat}(\text{head}_1, \dots, \text{head}_h) W_e \quad (8)$$

其中, h 为注意力头的数量, W_e 表示用于最终输出的线性变换矩阵,第 i 个头的注意力计算公式为

$$\text{head}_i = \text{Att}(Q_M^i, K_P^i, V_P^i) = \text{softmax} \left(\frac{Q_M^i (K_P^i)^T}{\sqrt{d_k/h}} \right) V_P^i \quad (9)$$

其中, T 表示矩阵的转置, d_k 表示 K_p 的维度。

然后,将特征向量 A_M 通过非线性变换重新投影回原始空间并通过残差连接添加到输入序列中,公式为

$$V'_M = \alpha \cdot V_M + \beta \cdot A_M \quad (10)$$

随后经过LN,采用与标准Transformer^[20]中相同的两层全连接层的FFN进一步细化全局信息,以提高模型的鲁棒性和准确性,并输出增强后的特征 \hat{V}_M ,公式为

$$\hat{V}_M = \gamma \cdot \text{FFN}(\text{LN}(V'_M)) + \delta \cdot V'_M \quad (11)$$

为了适应性地从不同分支学习数据,以实现性能提升。本文在残差连接的每个分支上应用了可学习系数,其中 α 、 β 、 γ 和 δ 是可学习参数,在训练过程中初始化为1。

同理,另一个CA模块也用于增强PET分支的特征,但与sMRI分支相反, V_P 将投影到多头注意力模块中的 Q , V_M 则投影到 K , V 可更全面地捕捉模态之间的关联性,从而提高模型性能。

传统方法通常通过堆叠多个模块来提升性能,但这种增加模型深度的策略不仅会大幅增加参数量,还可能引起过拟合。为了克服传统方法的缺点,本文使用的迭代学习策略通过多次迭代并共享参数的方式来逐步加深网络深度,同时在不增加参数量的前提下,逐步优化跨模态的互补信息。公式为

$$\{\hat{V}_M^n, \hat{V}_P^n\} = f_{ICA}(\{V_M, V_P\}, n) = \underbrace{f_{CA}(\cdots f_{CA}(\{V_M, V_P\}))}_n \quad (12)$$

其中, $\{\hat{V}_M^n, \hat{V}_P^n\}$ 表示经过 n 次迭代后的输出, $\{V_M, V_P\}$ 表示ICA的输入, $f_{ICA}(\cdot)$ 表示本文提出的ICA模块, $f_{CA}(\cdot)$ 表示CA模块。在每次迭代中,CA模块通过多头注意力机制捕捉sMRI和PET模态间的互补信息。具体来说,sMRI和PET的token通过线性变换投影到不同的空间,并通过注意力机制生成新的特征表示。这些特征通过残差连接和非线性变换进行更新,并输入到下一次迭代中。在每轮迭代中,CA模块不断地优化模态间的关联性,并逐步提升模型性能。由于共享参数设计,网络深度得以增加,但参数增长得到有效控制,从而避免了过拟合。通过这种迭代策略,网络逐步增强模态间信息融合的能力,直到输出优化后的特征。公式(12)描述了该迭代学习策略的数学表达式,其中输出结果能够更好地表达跨模态信息。最后,将ICA模块的输出序列和转换为特征映射,然后通过双线性插值重新校准到特征地图的原始大小。

对ICA模块进行空间和时间复杂度分析:

(1) 空间复杂度:ICA模块的参数量主要由以下组件构成。首先,位置嵌入层为sMRI和PET模态的每个token分配可学习的空间编码。假设每个模态有 N 个token,嵌入维度为 D ,因此其参数量为 $2ND$ 。其次,双分支CA模块中的多头注意力机制通过投影矩阵(Q, K, V)生成特征映射,每个分支的投影参数量为 $3D^2$,双分支总计 $6D^2$ 。此外,FFN采用两层全连接结构,单分支参数量为 $2 \times (D \times 4D) = 8D^2$,双分支共计 $16D^2$ 。迭代学习策略通过参数共享机制实现深度扩展,迭代 n 次不增加额外参数。因此,ICA模块的总空间复杂度为 $O(ND + D^2)$,其中主导项 D^2 源于注意力投影与FFN操作。

(2) 时间复杂度:ICA的计算开销主要由多头注意力和FFN主导。对于单次迭代,多头注意力的计算复杂度为 $O(N^2D)$,其中每个头的注意力权重需计算 $N \times N$ 矩阵;FFN的时间复杂度为 $O(ND^2)$,由全连接层的矩阵乘法决定。迭代策略通过 n 次重复调用CA模块,总时间复杂度为 $O(n(N^2D + ND^2))$ 。当输入token数量 N 显著大于嵌入维度 D 时, N^2D 项成为主导,此时计算量随数据规模呈平方增长。

2.2 元分类器(Meta-Classifer)

元分类器负责对150个基分类器进行集成得到全脑级别的分类结果,由于150个脑切块中并不是所有切块都包含了病变区域,因此,并不是每个基分类器的性能都表现良好。为了避免这些表现不佳的分类器可能造成的负面影响,在验证数据集上准确率低于75%的基分类器被淘汰,只保留了准确率大于75%的基分类器用于集成。元分类器由1个一维卷积层和1个全连接层组成,输入为双模态特征融合后的特征图 F_{fuse} ,公式为

$$F_{fuse} = \text{concat}(\hat{F}_M^i, \hat{F}_P^i, F_M^i, F_P^i) \quad (13)$$

其中, i 满足第 i 个基分类器的准确率 $\text{ACC}_i \geq 75\%$, \hat{F}_M^i 和 \hat{F}_P^i 为第 i 个基分类器Concatenate操作前sMRI和PET的特征图, F_M^i 和 F_P^i 为第 i 个基分类器ResNet模块的输出特征图。 F_{fuse} 首先经过一个filters为1,卷积核大小为1的一维卷积层,接着将卷积后的输出展平成1个一维数组,最后输入到最终的全连接层得到全脑级别的分类结果。

3 实验

3.1 数据集

本文中使用的数据来自阿尔茨海默症神经影像

学倡议^[21](Alzheimer's Disease Neuroimaging Initiative, ADNI) 数据库。ADNI于2003年作为公私合作伙伴关系启动,由首席研究员Michael W.Weiner医学博士领导。

本文共从ADNI数据库中挑选了418位同时拥有sMRI和PET图像数据的受试者,受试者分为AD、HC两组(192名AD,226名HC),AD组是被临床确诊的具有AD症状的受试者,HC是没有任何认知障碍的受试者。数据集的详细信息如表1所示。具体地,表1展示了数据集中各类别的样本数、性别、年龄、简易精神状态检查(Mini Mental State Examination, MMSE)^[22]、临床痴呆评定(Clinical Dementia Rating, CDR)^[23]的情况。

表1 数据集中的受试者的详细信息
Table 1 Details of subjects included in the studied dataset

	样本数	男/女	年龄	MMSE	CDR
AD	192	118/74	74.97±8.16	22.94±2.27	0.81±0.28
HC	226	145/121	74.38±5.86	29.04±1.16	0.00±0.03

3.2 数据预处理

从ADNI数据库直接下载的sMRI和PET源图像数据是不能直接使用的,需要对它们进行预处理。首先使用mricron软件将原始图像数据格式转化为神经影像信息技术倡议(Neuroimaging Informatics Technology Initiative, NIFTI) 格式,在MATLAB中使用工具包CAT12对sMRI进行去头骨、配准到蒙特利尔神经病学研究所(Montreal Neurological Institute, MNI) 标准空间和图像平滑等一系列处理,随后对每个sMRI进行灰度归一化,使每个受试者的图像数据像素值介于0~1之间。对于PET图像的预处理,原始图像数据转化为高动态范围(High Dynamic Range, HDR)格式后,使用spm12将它们配准到PET的模板上,再把它们仿射配准到相应的sMRI,同时删除大脑外侧的体素,并进行强度归一化,然后转换为8 mm全宽半高(Full Width at Half Maximum, FWHM) 的均匀各向同性分辨率。预处理完成后的sMRI和PET图像尺寸大小均为121 mm×145 mm×121 mm,体素大小为1.5 mm×1.5 mm×1.5 mm。

为了降低神经网络模型的训练难度,减少计算资源和训练多个基分类器,本文将切块的大小设置为25 mm×25 mm×25 mm。由于121和145不是25的整数倍,因此使用边缘填充和零填充将每张大小为121×145×121的图像重新格式化为125×150×125,最终得到2个模态的150个不同的切块。

3.3 实验设置

本文使用5折分层交叉验证策略来训练和测试所提出的集成模型,以此减少随机因素的影响。将数据集随机分为6个子集,随后取出一个子集固定地作为测试集,接着从剩下的5个子集中轮流选4个子集作为训练集,剩下的1个子集作为验证集。训练集、测试集和验证集的比例为4:1:1。另外,为了增加基础分类器训练期间的数据多样性,本文对输入的数据进行了数据增广,数据增广方法为随机取50%的切块沿X轴、Y轴或Z轴随机翻转。

实验所使用的是 NVIDIA GeForce RTX 4090 Ti 显卡,深度学习框架为TensorFlow2.5。训练分为基分类器的训练和元分类器的训练两个阶段。基分类器训练批次大小设置为64,学习率设置为 5×10^{-6} ,使用Adam优化器进行参数更新,训练轮数为1 000。为了防止过拟合,使用了Early Stopping技术来监控模型在验证集上的性能,当模型在验证集上的性能连续100轮内没有提升,则停止模型的训练。元分类器训练批次大小设置为16,学习率设置为0.000 5,使用了Adam优化器,训练轮数为50。实验采用分类准确率(Accuracy, ACC)、ROC曲线下面积(Area Under Curve, AUC)、敏感度(Sensitivity, SEN)、特异度(Specificity, SPE)、马修斯相关系数(Matthews correlation coefficient, MCC) 五种分类指标衡量模型性能。公式为

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (14)$$

$$AUC = P\left(\frac{TP}{TP + FN} > \frac{FP}{FP + TN}\right) \quad (15)$$

$$SEN = \frac{TP}{TP + FN} \quad (16)$$

$$SPE = \frac{TN}{FP + TN} \quad (17)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (18)$$

其中TP、TN、FP和FN分别代表真阳性数量、真阴性数量、假阳性数量和假阴性数量。真阳性(true positive, TP) 表示正类别中正确分类样本的数量,真阴性(true negative, TN) 表示负类别中正确分类样本的数量,假阳性(false positive, FP) 表示正类别中错误分类样本的数量,假阴性(false negative, FN) 是负类别中错误分类样本的数量, $P(\cdot)$ 表示事件发生的概率。

3.4 实验结果与分析

为了验证BICAFEF的AD诊断性能,本文首先将单模态与双模态融合进行实验对比,结果如表2所示。

在ADvs.HC的实验中,PET模态的分类效果整体优于sMRI模态,尽管双模态融合在SEN指标上低于单模态模型,但在ACC、SPE、AUC和MCC评估指标上表现优于sMRI模态和PET模态。ACC相较于

sMRI和PET分别有6.6个百分点和1个百分点的提升,AUC分别有2.9个百分点和0.6个百分点的提升,MCC分别提升13个百分点和2个百分点,并且双模态融合具有更低的标准差。表明本文提出的双模态融合方法能够有效结合sMRI提供的结构信息和PET提供的功能及代谢信息,从而实现信息互补,提升AD诊断的准确性。

表2 单模态与双模态融合的分类结果
Table 2 Classification results of unimodal and bimodal fusion

模态	ACC	SEN	SPE	AUC	MCC
sMRI	0.877±0.039	0.886±0.025	0.866±0.057	0.946±0.030	0.751±0.080
PET	0.933±0.029	0.963±0.023	0.897±0.057	0.969±0.009	0.867±0.059
sMRI+PET	0.943±0.015	0.926±0.015	0.963±0.026	0.975±0.004	0.887±0.031

为了更好地说明不同模态如何影响网络性能并使模型更具可解释性,本文利用Grad-CAM^[24]显示单模态和双模态融合模型关注的区域。

可视化结果如图2所示,结果表明,单模态模型对sMRI和PET关注的区域区别较大,这表明模型采用不同的模态提取到了不同的脑区信息。双模态融合模型关注的区域与单模态关注的区域有一定的重叠,同时也呈现出更多的综合信息。为了进一步显示单模态模型与双模态融合模型关注的脑区,本文使用Brainnetome Atlas^[25]对第106个切块(Patch106)进行脑区标注,并对比了不同模态下模型关注的具体脑区,如表3所示。

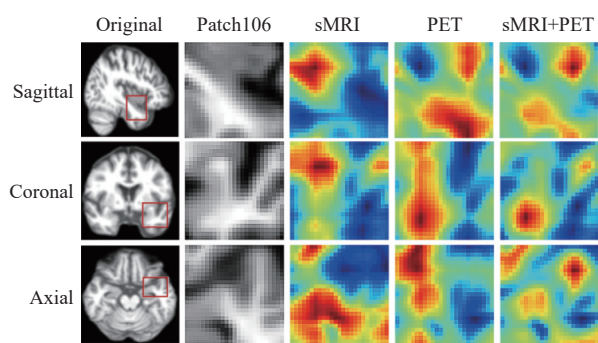


图2 单模态与双模态融合模型可视化结果

Fig.2 Visualization results of single-modal and bimodal fusion models

结果表明,sMRI关注的主要脑区包括与脑区形态、结构变化相关的脑区,如颞上回(73)、前额叶皮层(77、79)以及与记忆和情绪处理相关的区域如海马(215)。而PET则偏向于那些与代谢活动、血流和神经递质分布相关的脑区,例如腹侧岛皮层(165)和背侧岛皮层(167),这些区域与大脑的认知和情感调节

有关。双模态融合模型关注的脑区不仅包含了sMRI和PET各自的显著区域,还涵盖了多个脑区的交集和更广泛的脑功能联系。表明通过融合sMRI和PET数据,双模态融合模型能够捕捉到更加全面的脑区特征,提升了对复杂脑部病变的识别能力。与单一模态相比,双模态融合模型能够更好地反映出不同脑区之间的相互作用,并可能揭示出单模态无法识别的细微差别。

表3 不同模态下模型关注的脑区

Table 3 Brain regions focused by the model under different modalities

模态	关注的脑区
sMRI	73, 77, 79, 87, 169, 211, 213, 215, 221, 223, 225, 229
PET	109, 165, 213, 215, 217, 219, 223, 225, 229
sMRI + PET	73, 77, 79, 109, 165, 167, 169, 173, 213, 215, 219, 223

接着,将本文方法与当前主流的多模态融合方法进行比较,得到对比实验结果,如表4所示。对比方法包括Huang等^[8]提出的2DCNN+BGRU方法,Sharma等^[17]提出的双分支三维卷积神经网络模型,Zhang等^[15]提出的MCAD框架、Golovanevsky等^[16]提出的MADDi方法、Qiu等^[26]提出的一种基于多融合和疾病诱导学习的多模态诊断网络(Multimodal Diagnosis Network Based on Multi-fusion and Disease-induced Learning, MDL-Net)、Tang等^[27]提出的交叉增强融合模块(Cross-Enhanced Fusion Module, CEFM)和Shen等^[19]提出的双模态特征融合(Dual-modal Feature Fusion, DMFF)方法。由表4可知,在使用相同数据集的情况下,本文方法性能均优于其他方法。说明本方法能够更好地捕捉数据中的重要特征,从而提高分类的精确性。此外,AUC的提升进一步反映了

模型在不同阈值下的分类能力,证明了其在实际应用中的可靠性。MCC的提升表明本文提出的方法在处理类别不平衡问题时具有更强的鲁棒性。同时也

表明了该方法在处理复杂数据集时的有效性,进一步增强了其在实际应用场景中的价值。本文方法性能均优于其他方法。

表4 对比实验分类结果
Table 4 Comparison of experimental classification results

方法	ACC	SEN	SPE	AUC	MCC
2DCNN+BGRU ^[8]	0.894±0.083	0.885±0.056	0.900±0.073	0.953±0.038	0.788±0.184
Sharma ^[17]	0.894±0.048	0.926±0.035	0.857±0.076	0.965±0.040	0.792±0.195
MCAD ^[15]	0.919±0.023	0.928±0.056	0.909±0.031	0.918±0.022	0.840±0.047
MADDi ^[16]	0.935±0.024	0.925±0.043	0.937±0.022	0.968±0.026	0.879±0.058
MDL-Net ^[26]	0.934±0.031	0.926±0.039	0.943±0.044	0.966±0.028	0.869±0.070
CEFM ^[27]	0.941±0.020	0.929±0.040	0.961±0.036	0.971±0.014	0.880±0.038
DMFF ^[19]	0.928±0.034	0.921±0.027	0.938±0.054	0.968±0.014	0.857±0.054
本文模型	0.943±0.015	0.926±0.015	0.963±0.026	0.975±0.004	0.887±0.031

实验结果表明,本文框架使用的结合迭代学习策略和CA的方法可以有效加强sMRI的结构信息与PET的代谢特征之间的交互融合。具体来说,共享参数的迭代学习策略能够逐步优化两个模态的特征表示,使得网络在融合过程中对重要特征的捕捉更加精准,减少了冗余信息的干扰。此外,CA通过动态权重分配,使得框架能够在sMRI和PET间进行信息筛选和特征增强,有效提高了框架对多模态数据的理解能力。两者结合不仅加强了来自模态间和模态内特征的互补信息的记忆,还增强了模态间的协同作用,从而提高了阿尔茨海默症的多模态诊断性能,获得更优的分类效果。

3.5 消融实验

为了验证SFS、ICA的有效性,设计如下消融实验:(1)屏蔽SFS模块,sMRI和PET的特征向量 F_M 和 F_P 直接输入到ICA模块,最终获得AD分类结果。(2)使用SFS和ICA,设置不同的迭代次数(n),进行AD分类实验。(3)保留SFS模块,屏蔽ICA,进行AD分类实验。(4)屏蔽SFS和ICA,特征向量 F_M 和 F_P 直接输入到全连接层进行分类,作为基线实验。消融实验结果如表5所示。

在屏蔽不同模块的消融实验中,相对于基线实验,单独使用ICA模块后框架准确率(ACC)提升1.7个百分点,单独使用SFS模块后框架准确率(ACC)提升1.6个百分点,没有屏蔽模块的情况下框架准确率提升2.9个百分点。消融实验结果表明,本文框架的准确率和泛化能力的提升,得益于SFS和ICA模块的保留以及这两个模块在不同功能上的互补作用。ICA模块通过交叉注意力机制,更好地融合了多模态信息,从而增强了框架对复杂特征的捕捉能力。而

SFS模块则通过特征收缩策略,有效降低了冗余特征的干扰,提高了特征的区分度。此外,在没有屏蔽模块的情况下,框架的表现进一步提升,表明ICA和SFS模块的结合在实现sMRI和PET信息的深度融合方面具有重要作用。总体而言,该消融实验验证了SFS和ICA模块的有效性及其对框架性能的贡献。

表5 消融实验结果
Table 5 The result of ablation experimen

SFS	ICA	ACC	SEN	SPE	AUC	MCC	时间/s
		0.914	0.895	0.938	0.968	0.830	59.0
	√	0.931	0.926	0.936	0.970	0.862	83.1
√		0.930	0.913	0.951	0.970	0.862	72.6
√	√	0.943	0.926	0.963	0.975	0.887	95.8

为了进一步验证SFS模块的优越性,本文对SFS的池化操作部分进行了消融实验并与平均池化-最大池化组合和混合池化两种不同的聚合方法进行对比。其中平均池化-最大池化组合需要先进行平均池化后进行最大池化,混合池化则将平均池化和最大池化的结果进行拼接。实验结果如表6所示。

表6 不同池化操作的实验结果
Table 6 Experimental results of different pooling operations.

方法	ACC	SEN	SPE	AUC	MCC	时间/s
平均池化	0.937	0.929	0.947	0.973	0.874	93.6
最大池化	0.932	0.924	0.944	0.970	0.866	95.0
平均池化-最大池化组合	0.934	0.916	0.956	0.972	0.870	95.2
混合池化	0.942	0.941	0.943	0.974	0.883	114.8
自适应聚合	0.943	0.926	0.963	0.975	0.887	95.8

实验结果表明,自适应聚合方法在大部分指标上都表现优于其他池化方式,尤其是在ACC、SPE和

MCC上有着更好的性能。相比之下,混合池化方法也取得了较好的表现,但仍稍逊于自适应聚合。其他池化方法(如平均池化和最大池化)相较于自适应聚合和混合池化在性能上有所落后。这些结果验证了自适应聚合方法在SFS模块中的优势。

设置不同迭代次数的实验结果如表7所示,当迭代次数为2时框架的性能最好,比只迭代1次的准确率(ACC)高1.6个百分点。值得注意的是,迭代1次的准确率反而比迭代0次(即不使用ICA)的低,而迭代3次的准确率低于迭代2次。这表明迭代过少和过多的次数都不会提升性能,只有适当的迭代次数才能在框架中充分融合多模态信息并提取有用特征,而不过多地引入噪声或冗余信息。当迭代次数过少时,框架未能充分利用交叉注意力机制的优势,导致特征融合不足。而当迭代次数过多时,框架可能会过拟合到训练数据,或者在交叉注意力模块中累积更多的噪声,从而降低泛化能力。实验结果表明,本文方法在迭代2次时能够在信息融合和噪声控制之间取得良好平衡,并提升框架的预测精度。此外,由于迭代学习技术使用共享参数,更多的迭代不会产生额外的参数或内存成本。

表7 不同迭代次数的实验结果

Table 7 Experimental results with different numbers of iterations

n	Params	ACC	SEN	SPE	AUC	MCC	时间/s
0	8.8×10^6	0.930	0.913	0.951	0.970	0.862	72.6
1	8.8×10^6	0.927	0.909	0.949	0.974	0.857	82.5
2	8.8×10^6	0.943	0.926	0.963	0.975	0.887	95.8
3	8.8×10^6	0.928	0.931	0.925	0.975	0.856	109.2

4 结论

为了加强来自模态间和模态内特征的互补信息的记忆,从而更高效、更准确地学习不同模态的互补信息进行AD诊断,本文提出了BICAFEF。该框架采用一种新型的迭代交叉注意力融合策略,通过多次迭代的交叉注意力机制在不同模态之间进行信息交互与融合。具体来说,BICAFEF利用ICA模块,使得来自不同模态的特征可以相互参考和强化,从而更好地捕捉模态间的互补信息。每次迭代交叉注意力机制会动态调整特征的权重,逐步优化多模态特征的融合效果。此外,框架还引入了SFS模块,有效减少冗余信息,突出关键特征,进一步提高信息融合的精度和效率。通过更好地融合来自不同成像模态的信

息,BICAFEF有效提高了对疾病相关特征的识别能力,从而有助于实现更早期和更可靠的阿尔茨海默症诊断。这些改进不仅有助于提高临床诊断的准确性,还可能为个性化治疗方案的制定提供更有价值的参考。

参考文献:

- [1] TREJO-LOPEZ J A, YACHNIS A T, PROKOP S. Neuro-pathology of Alzheimer's disease[J]. *Neurotherapeutics*, 2022, 19(1): 173-185.
- [2] REN R J, QI J L, LIN S H, *et al.* The China Alzheimer report 2022[J]. *General Psychiatry*, 2022, 35(1): e100751.
- [3] BUCKNER R L. Memory and executive function in aging and AD: multiple factors that cause decline and reserve factors that compensate[J]. *Neuron*, 2004, 44(1): 195-208.
- [4] DUNNE R A, AARSLAND D, O'BRIEN J T, *et al.* Mild cognitive impairment: the Manchester consensus[J]. *Age and Ageing*, 2021, 50(1): 72-80.
- [5] CHEN Q H, FU Q, BAI H, *et al.* Longformer: longitudinal transformer for Alzheimer's disease classification with structural MRIs[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE/CVF, 2024: 3575-3584.
- [6] 徐萍萍, 黄国恒, 赵钦, 等. AFEM-Transformer: 基于自适应特征提取的Transformer阿尔兹海默症早期诊断研究[J]. *广东工业大学学报*, 2025, 42(1): 51-59.
- [7] XU P P, HUANG G H, ZHAO Q, *et al.* AFEM-Transformer: early diagnosis of Alzheimer's disease based on adaptive feature extraction with transformer[J]. *Journal of Guangdong University of Technology*, 2025, 42(1): 51-59.
- [8] DUAN J W, LIU Y, WU H H, *et al.* Broad learning for early diagnosis of Alzheimer's disease using FDG-PET of the brain[J]. *Frontiers in Neuroscience*, 2023, 17: 1137567.
- [9] 黄殷. 基于卷积循环神经网络的AD多模态生物标记物发现[D]. 广州: 广东工业大学, 2021.
- [10] LIU X, LI W M, MIAO S, *et al.* HAMMF: hierarchical attention-based multi-task and multi-modal fusion model for computer-aided diagnosis of Alzheimer's disease[J]. *Computers in Biology and Medicine*, 2024, 176: 108564.
- [11] LIU F Y, YUAN S Z, LI W M, *et al.* Patch-based deep multi-modal learning framework for Alzheimer's disease diagnosis using multi-view neuroimaging[J]. *Biomedical Signal Processing and Control*, 2023, 80: 104400.
- [12] LI W M, GAO Q Q, DU M, *et al.* Multiclass diagnosis of stages of Alzheimer's disease using linear discriminant analysis scoring for multimodal data[J]. *Computers in Biology and Medicine*, 2021, 134: 104478.
- [13] KHATRI U, KWON G R. An efficient combination among sMRI, CSF, cognitive Score, and APOE ϵ 4 biomarkers for classification of AD and MCI using extreme learning ma-

- chine[J]. *Computational Intelligence and Neuroscience*, 2020, 2020(1): 8015156.
- [13] WALHOVD K B, FJELL A M, BREWER J, *et al.* Combining MR imaging, positron-emission tomography, and CSF biomarkers in the diagnosis and prognosis of Alzheimer disease[J]. *American Journal of Neuroradiology*, 2010, 31(2): 347-354.
- [14] KOHANNIM O, HUA X, HIBAR D P, *et al.* Boosting power for clinical trials using classifiers based on multiple biomarkers[J]. *Neurobiology of Aging*, 2010, 31(8): 1429-1442.
- [15] ZHANG J, HE X H, LIU Y, *et al.* Multi-modal cross-attention network for Alzheimer's disease diagnosis with multimodality data[J]. *Computers in Biology and Medicine*, 2023, 162: 107050.
- [16] GOLOVANEVSKY M, EICKHOFF C, SINGH R. Multimodal attention-based deep learning for Alzheimer's disease diagnosis[J]. *Journal of the American Medical Informatics Association*, 2022, 29(12): 2014-2022.
- [17] SHARMA R, SIBILLE L, FAHMI R. Multi - branch convolutional neural network for Alzheimer's disease versus normal control classification using PET images[J]. *Alzheimer's & Dementia*, 2023, 19: e061092.
- [18] EBRAHIMI A, LUO S H, CHIONG R. Introducing transfer learning to 3D ResNet-18 for Alzheimer's disease detection on MRI images[C]//2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ) . Wellington: IEEE, 2020: 1-6.
- [19] SHEN J F, CHEN Y F, LIU Y, *et al.* ICAFusion: iterative cross-attention guided feature fusion for multispectral object detection[J]. *Pattern Recognition*, 2024, 145: 109913.
- [20] HAN K, XIAO A, WU E H, *et al.* Transformer in transformer[J]. *Advances in Neural Information Processing Systems*, 2021, 34: 15908-15919.
- [21] MUELLER S G, WEINER M W, THAL L J, *et al.* The Alzheimer's disease neuroimaging initiative[J]. *Neuroimaging Clinics*, 2005, 15(4): 869-877.
- [22] FOLSTEIN M F, ROBINS L N, HELZER J E. The minimal state examination[J]. *Archives of General Psychiatry*, 1983, 40(7): 812-812.
- [23] BERG L. Clinical dementia rating[J]. *The British Journal of Psychiatry*, 1984, 145(3): 339-339.
- [24] SELVARAJU R R, COGSWELL M, DAS A, *et al.* Grad-cam: visual explanations from deep networks via gradient-based localization[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 618-626.
- [25] FAN L Z, LI H, ZHUO J J, *et al.* The human brainnetome atlas: a new brain atlas based on connectonal architecture[J]. *Cerebral Cortex*, 2016, 26(8): 3508-3526.
- [26] QIU Z F, YANG P, XIAO C L, *et al.* 3D multimodal fusion network with disease-induced joint learning for early Alzheimer's disease diagnosis[J]. *IEEE Transactions on Medical Imaging*, 2024, 43(9): 3161-3175.
- [27] TANG C S, XI M B, SUN J D, *et al.* MACFNet: detection of Alzheimer's disease via multiscale attention and cross-enhancement fusion network[J]. *Computer Methods and Programs in Biomedicine*, 2024, 254: 108259.

(责任编辑: 张玮欣 英文审核: 熊荣斌)