

基于3D多模态卷积网络与跨模态特征集成的阿尔茨海默症分类

朱厚元¹, 郑乐乐¹, 商浩¹, 臧雪峰¹, 吴少琪¹, 周广超¹, 孙建德², 乔建革¹

(1. 山东师范大学物理与电子科学学院, 济南 250307; 2. 山东师范大学信息科学与工程学院, 济南 250307)

摘要: 多模态神经影像技术为阿尔茨海默症(Alzheimer's disease, AD)的早期精准诊断提供了重要的技术支撑。然而,由于不同模态神经影像数据在成像原理和特征表达上存在固有异质性,模态间的信息融合面临挑战。针对这一问题,提出了一种基于3D ResNet架构的多模态融合网络(Multi-modal fusion network, MFN),用于AD的早期辅助诊断。该方法首先采用3D ResNet网络分别提取T1加权 and T2加权磁共振图像的特征表示,然后设计了一种创新的跨模态特征集成模块(Cross-modal feature integration module, CFIM)。相较于多模态数据直接串联,导致维度增长无法自适应调整模态权重的问题,CFIM采用分阶段融合策略,包括全局信息融合模块、局部特征学习模块和关键因素模块。最后,融合后的多模态特征通过全连接神经网络进行分类决策。相比早期拼接的固定权重叠加和后期融合的浅层聚合,该策略能更精准地筛选出疾病诊断相关的特征。通过在阿尔茨海默症神经影像倡议(ADNI)数据库上的实验结果表明,与现有方法相比,本文方法在AD分类任务中具有较高的准确率和显著优势,且消融实验进一步验证了各模块的有效性,为多模态神经影像分析提供了新的技术思路。

关键词: 阿尔茨海默症; 3D多模态融合网络; 核磁共振图像; 跨模态特征集成模块; 深度学习

中图分类号: TP391

文献标志码: A

Alzheimer's Disease Classification Based on 3D Multi-modal Convolutional Network and Cross-Modal Feature Integration

ZHU Houyuan¹, ZHENG Lele¹, SHANG Hao¹, ZANG Xuefeng¹, WU Shaoqi¹, ZHOU Guangchao¹, SUN Jiande², QIAO Jianping¹

(1. School of Physics and Electronics, Shandong Normal University, Jinan 250307, China; 2. School of Information Science and Engineering, Shandong Normal University, Jinan 250307, China)

Abstract: Multi-modal neuroimaging technology provides crucial technical support for the early and precise diagnosis of Alzheimer's disease (AD). However, due to the inherent heterogeneity in imaging principles and feature representations across different neuroimaging modalities, the fusion of inter-modal information poses significant challenges. To address this issue, this study proposes a multi-modal fusion network (MFN) based on a 3D ResNet architecture for the early auxiliary diagnosis of AD. The proposed method first employs a 3D ResNet to separately extract feature representations from T1- and T2-weighted magnetic resonance images. Subsequently, an innovative cross-modal feature integration module (CFIM)

基金项目: 济南市市校融合发展战略工程项目(JNSX2023038)。

收稿日期: 2025-05-10; 修订日期: 2025-07-07

is designed to overcome the limitations of direct concatenation. CFIM adopts a hierarchical fusion strategy, consisting of global information fusion module, local feature learning module and key factor module. Finally, the fused multimodal features are fed into a fully connected neural network for classification. Compared to early concatenation (fixed-weight fusion) and late fusion (shallow aggregation), this strategy more effectively identifies disease-relevant diagnostic features. Experiments conducted on the Alzheimer's disease neuroimaging initiative (ADNI) database demonstrate that the proposed method achieves higher accuracy and superior performance in AD classification tasks compared to existing approaches. Ablation studies further validate the effectiveness of each module, offering new technical insights for multi-modal neuroimaging analysis.

Key words: Alzheimer's disease (AD); 3D multi-modal fusion network; magnetic resonance images; cross-modal feature integration module; deep learning

引 言

阿尔茨海默症(Alzheimer's disease, AD)是一种不可逆的神经退行性疾病,发病率随着年龄的增长而增加,该疾病以记忆功能损害和认知功能障碍为主要临床表现^[1],目前全球患者人数已达5 500万,被认为是导致死亡的第七大原因^[2]。轻度认知功能障碍(Mild cognitive impairment, MCI)是AD的前驱期,其特征性表现为持续性认知功能减退^[3],尽管AD的确切发病机制尚未完全阐明,鉴于其不可治愈性的特点,早期预防和精准诊断已成为当前研究重点^[4],现在已经有很多疾病预测方法^[5-6],本文使用结构磁共振成像预测方法。

结构磁共振成像(structural magnetic resonance imaging, sMRI)是一种流行的非侵入性神经成像技术,包含T1和T2核磁共振图像,能够有效检测阿尔茨海默症及轻度认知障碍患者的特征性脑结构改变^[7]。T1核磁共振图像具有优异的解剖结构分辨率和灰质/白质对比度,适合检测脑萎缩等结构性改变,但对早期病理变化敏感度较低;而T2加权MRI对组织含水量变化敏感,能有效识别水肿、炎症等异常信号,但在解剖对比度方面表现较差,二者在神经影像诊断中具有互补优势^[8]。有研究人员使用单模态进行AD分类,如Hassan等^[9]提出了一种轻量级堆叠卷积神经网络与通道注意力网络(Stacked CNN-based multichannel attention networks, SCCAN)相结合的MRI图像AD分类方案,通过顺序集成CNN模块构建堆叠网络,实现多层次特征提取以形成层次化特征理解;Nie等^[10]提出多平面sMRI三维引导式预训练模型,结合2D CNN和Swin Transformer的混合切片级网络来捕捉与阿尔茨海默病相关的不同脑结构萎缩之间的相互关系,然而该方法依赖于单模态的脑成像数据,限制了方法从多个结构和功能的角度同时捕获大脑特征的能力。为了解决这一问题,有研究人员通过不同的方法,结合了融合结构和功能数据的多模态神经成像技术,从多模态观点^[11-14]为早期AD诊断提供了更全面的视角,如Zhang等^[15]使用结构网络初始化图的拓扑结构,并通过结合功能信息迭代更新来融合多模态信息,属于早期多模态信息融合,但这种方法需要深度网络多次更新,找到最佳融合参数;Wang等^[16]提出利用sMRI、临床特征和遗传多态性数据准确预测MCI向AD的长期转化,在提取完成图像与临床特征后,通过拼接等方法进行特征融合,然后进行疾病预测,属于晚期多模态信息融合,但是临床特征等信息可能与sMRI图像信息不同源,较难融合。许多研究也已经成功地利用多模态脑图像在AD诊断^[17-19]方面取得了显著进展。然而现有的方法经常采用简单的融合技术,未能考虑不同融合方法对模型性能的影响。为增强多模态数据的有效融合,本文提出了一种基于3D ResNet的多模态融合网络(Multi-modal fusion network, MFN)进行脑疾病的分类,通过跨模态特征集成模块(Cross-modal feature integration

module, CFIM)的新融合机制,有效融合多模态数据特征。CFM通过全局信息融合(Global information fusion, GIF)模块、局部特征学习(Local feature learning, LFL)模块以及关键因素模块(Key factor module, KFM),从3个角度进行多模态数据的学习融合,增强了从全局、局部和潜在学习多个方面的特征表示。

本文的主要贡献如下:(1)提出一种用于AD早期诊断的三维多模态融合网络,有效地结合三维多模态成像数据建立深度学习模型,为AD诊断提供更丰富的特征;(2)跨模态特征集成通过GIF、LFL和KFM三大模块,有效捕获多模态特征之间的全局、局部和潜在关系,增强多模态特征的特征表示。

1 实验方法

本文提出基于3D ResNet和多模态融合的阿尔茨海默症早期辅助诊断方法,具体包括两个步骤:(1)多模态特征提取,利用3D ResNet18的骨干网络提取多模态特征;(2)跨模态特征集成模块,用于融合多模态特征。方法框架如图1所示。

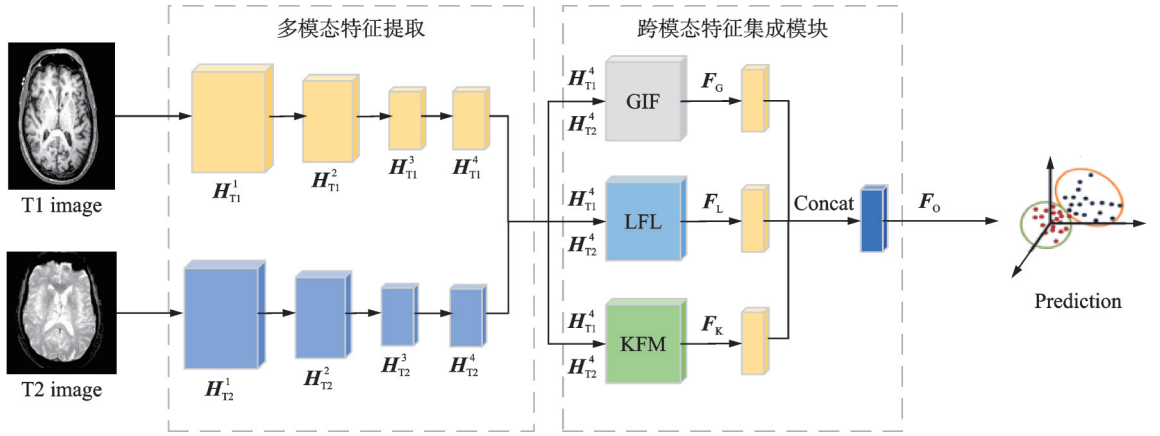


图1 本文提出的多模态融合分类算法流程图

Fig.1 Flow chart of the proposed multi-modal fusion classification algorithm

1.1 数据预处理

本文采用的数据集是阿尔茨海默氏症神经成像计划ADNI(包含ADNI1、ADNI2与ADNI3)公开数据库,多模态数据包含140例老年痴呆(AD)数据,其中104例来自ADNI2,36例来自ADNI3;136例认知正常(NC)数据,其中32例来自ADNI1,104例来自ADNI2;所有受试者均同时提供了T1和T2 MRI图像。

使用SPM12工具(<http://www.fil.ion.ucl.ac.uk/spm/software/spm12/>)对T1和T2 MRI图像进行预处理时,首先利用其6自由度(6-DoF,包含 $x/y/z$ 轴3维旋转与3维平移,无缩放操作)的刚体变换模型将每位受试者的T1与T2图像进行空间配准,以保持结构的几何一致性;接着对图像分别实施偏置校正以消除强度不均匀性,再通过线性尺度变换进行标准化;最后将图像归一化至基于MNI-152模板的蒙特利尔神经学研究所解剖自动标记(Anatomical automatic labeling, AAL)空间标准,将图像配准在统一空间坐标尺度,为后续卷积神经网络处理提供一致的空间采样框架。

1.2 多模态特征提取

设 $X_{T1}, X_{T2} \in \mathbb{R}^{d \times h \times w \times c}$ 分别表示预处理后的T1和T2 MRI图像,其中 d, h, w 和 c 分别表示深度、高度、宽度和通道。使用3D ResNet18为骨干网络,提取多模态特征,如图1所示。为了便于特征融合过

程的维数变换,首先使用上采样层对输入特征进行维数变换。然后设置了3D ResNet的通道数,从输入层开始依次为16、32、64、128,因此可以凭借高分辨率特征图保留脑沟、脑回等微观解剖结构细节,增大感受野,保留更多的有用信息,然后使用降维层(第0层)和4个残差层进行多模态特征提取,降维层采用1个3D卷积操作实现维度转换,卷积核大小为 $3\times3\times3$,步长为 $4\times4\times4$,通过零填充(padding=1)保持特征映射的边缘信息,同时设置批量归一化和ReLU激活函数,实现空间维度的下采样与通道维度的扩展,残差层均由2个残差块组成,每个残差块由2个3D卷积层构成,卷积层步长均为 $1\times1\times1$,且每个卷积层后均紧跟批量归一化和ReLU激活函数,残差块采用恒等映射 shortcut 连接,即输入特征直接与卷积操作输出相加,确保特征维度一致。这是通过低分辨率特征实现宏观特征捕捉与运算效率的双重优化,通过层级特征筛选机制,强化图像宏观特征的表达,同时,残差层的跳跃连接机制(Skip connection)可以解决特征分辨率降低与浅层细节保留的矛盾,确保浅层高分辨率特征与深层低分辨率特征直接融合,避免小尺寸信息的丢失。表1详细列出了每一层的输入和输出尺寸。最后,得到了多模态特征 $H_{T1}^m, H_{T2}^m \in \mathbb{R}^{dm \times hm \times wm \times cm}$ (m 表示不同的层)。

表1 3D ResNet18 参数设置
Table 1 Parameter setting of 3D ResNet18

层	输入输出尺寸($d \times h \times w \times c$)	
	输入	输出
降维层	$1 \times 128 \times 128 \times 128$	$16 \times 32 \times 32 \times 32$
残差层 1	$16 \times 32 \times 32 \times 32$	$16 \times 32 \times 32 \times 32$
残差层 2	$16 \times 32 \times 32 \times 32$	$32 \times 16 \times 16 \times 16$
残差层 3	$32 \times 16 \times 16 \times 16$	$32 \times 16 \times 16 \times 16$
残差层 4	$32 \times 16 \times 16 \times 16$	$64 \times 8 \times 8 \times 8$

1.3 跨模态特征集成模块

由于两种核磁共振模态之间关联性较弱,采用简单的融合手段往往难以有效地捕捉和利用多模态特征的互补性信息。为了解决这一限制,CFIM 模块设计用来提高多模态融合特征的可学习性,CFIM 包括3个模块:GIF、LFL和KFM。GIF利用了多头注意力机制,能够在多模态特征内实现自适应的全局学习,而早期拼接只是简单地将特征相连,形成一个新的长输入模型,这种方式并未深入挖掘图像各区域特征间的关联,只是进行了表面上的整合。GIF则不同,它能在图像特征维度,对不同区域特征赋予不同权重,对关键目标区域重点关注,实现对多模态特征内全局信息的智能筛选与有效学习。LFL利用局部感知多尺度卷积神经网络来学习不同尺度之间的特征及其相应关系。对比晚期聚合将图像各自独立经过神经网络学习,在最后的决策阶段,才将两者的输出结果进行简单合并,将图像分类概率求平均决定最终分类,而LFL却从特征学习初始阶段就发挥作用,在图像模态中,通过不同大小卷积核,对图像中从微小细节到较大结构的多尺度特征进行捕捉,并挖掘这些不同尺度特征间的联系。KFM利用多尺度池化与元素级乘法学习模态特征之间的可学习潜在信息,增强模态特征表示。然后利用GIF、LFL和KFM得到 $F_G \in \mathbb{R}^c, F_L \in \mathbb{R}^c$ 和 $F_K \in \mathbb{R}^c$,并对3个特征进行维度调整和求和,得到最终的融合特征 $F_O \in \mathbb{R}^c$ 。

(1) 全局信息融合模块(GIF):Transformer 机制中的多头注意力机制可以建立特征的长期依赖关系,这有助于捕获两个特征之间的全局关系。GIF运用多头注意力机制,在单一模态内部交互式地学习并增强不同位置的特征,从而改进和优化单独模态特征的代表能力。因此受Transformer机制的启发,GIF被设计用来学习多模态数据的全局关系如图2所示。

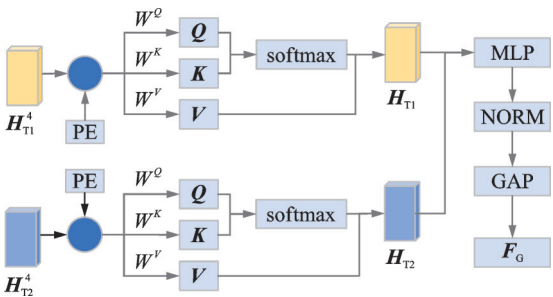


图2 全局信息融合模块

Fig.2 Global information fusion module

具体来说,改变 H_{T1} 和 H_{T2} 的尺寸,并对 H_{T1}, H_{T2} 增加位置嵌入(Positional embedding),位置嵌入使得 H_{T1}, H_{T2} 包含更丰富的空间位置信息。位置嵌入使用正弦余弦函数,为每个位置生成独特的向量表示,并将其与输入嵌入相加,使不同位置的嵌入向量在高维空间中具有可区分的相对位置关系,确保模型能捕获序列中的位置关联;对特征进行位置编码后得到 $H_{T1}, H_{T2} \in \mathbb{R}^{c \times n}$,其中 $n=dhw$ 。然后,对 H_{T1}, H_{T2} 分别计算查询(Query)、键(Key)和值(Value)矩阵,即

$$Q = H \times W^Q \quad (1)$$

$$K = H \times W^K \quad (2)$$

$$V = H \times W^V \quad (3)$$

式中: $W^Q, W^K, W^V \in \mathbb{R}^{c \times c'}$ 分别为查询、键和值权重矩阵,其中 c' 为映射后的特征维度,通常满足 $c' = c/h$, h 为多头注意力的头数,为平衡计算效率与特征表达能力, h 取8。因为较多的头数可捕捉更丰富的特征交互模式,但会增加计算量;较少的头数则可能限制信息提取的多样性。 Q, K, V 矩阵将作为后续注意力机制或自注意力机制中的关键组成部分。通过这些矩阵,模型能够捕捉到不同模态数据之间的关联性,并基于位置嵌入所引入的空间位置信息,实现更为精准和丰富的特征表示。然后,使用多头注意力机制计算不同模态间的全局关系,并获得不同模态的增强特征 $H_{T1}, H_{T2} \in \mathbb{R}^{c \times n}$,即

$$H_{T1} = \text{softmax}(Q_{T1} K_{T1} / \sqrt{d_k}) V_{T1} \quad (4)$$

$$H_{T2} = \text{softmax}(Q_{T2} K_{T2} / \sqrt{d_k}) V_{T2} \quad (5)$$

式中: H_{T1} 和 H_{T2} 为计算后的模态特征, d_k 是一个超参数,由输入特征维度 dim 与多头注意力的头数 h 计算得到,输入特征维度 $\text{dim}=512$,则 $d_k = \text{dim}/h=64$,因此 d_k 为64,此维度确保每个头在高维空间中独立学习特征子空间,同时避免维度过高导致的过拟合问题。 softmax 用来归一化维度,对于缩放点积注意力公式(4,5)的输出矩阵, softmax 将对每个查询向量(Q)对应的键向量(K)维度进行概率分布转换,确保同一查询位置下不同键位置的注意力权重之和为1。然后,对模态增强特征 H_{T1} 和 H_{T2} 进行融合,得到融合特征 F_G ,即

$$F_G = \text{MLP}(\text{softmax}(H_{T1} (H_{T2})^T)) \quad (6)$$

式中:MLP为多层感知器,由两个全连接层构成,这两层之间嵌入了一个高斯误差线性单位(Gaussian error linear unit, GELU)作为激活函数。GELU激活函数的作用在于减少内部协变量的偏移现象,并增强模型的非线性表达能力。之后,得到了自适应融合特征 $F_G \in \mathbb{R}^{c \times n}$ 。改变 F_G 维度并进行全局平均池化(Global average pooling, GAP)操作,得到最终的融合特征 $F_G \in \mathbb{R}^c$ 。

(2)局部特征学习模块(LFL):LFL旨在通过编码不同尺度下的聚合特征来增强模态表示,如图3所示。在对多模态的MRI图像提取特征之后,为了捕捉不同尺度的信息,使用不同尺寸的卷积核对这些特征进行混合卷积操作,并引入残差连接,解决深层网络梯度消失问题,增强数据表示。这一步骤实现了通道间的信息转换,还通过卷积运算学习了特征的权重。随后,将卷积后的低尺度特征与高尺度特征进行融合。融合后的特征再次经过不同尺寸卷积核的卷积操作,以进一步提取和融合特征信息。使用LFL旨在高效整合与利用多尺度特征资源,从而提升模型的代表与学习能力。上述操作可定义为

$$S_1(s_1, s_2) = \text{Concat}(s_2, \sigma(\text{Conv}(s_1)) \times s_1) \quad (7)$$

$$S_2(s) = \text{Conv}(\sigma(\text{Conv}(s)) \times s) \quad (8)$$

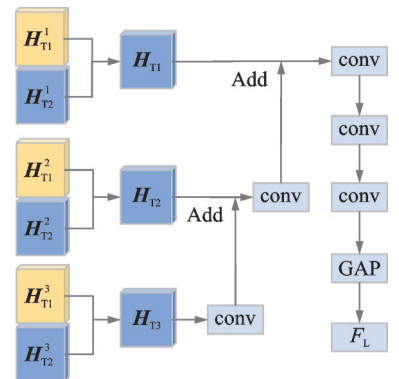


图3 局部特征学习模块

Fig.3 Local feature learning module

$$F_1 = S_{\text{GAP}}(S_2(S_1(S_1(H_{T1}, H_{T2})))) \quad (9)$$

式中: $S_1(s_1, s_2)$ 和 $S_2(s)$ 分别表示不同卷积核大小的LFL块;“ $\times$ ”表示跳跃连接; s_1, s_2 和 s 表示不同的输入特征; S_{GAP} 表示由GAP和特征展平组成的操作。Conv卷积的平均核大小为 $3 \times 3 \times 3, 1 \times 1 \times 1$ 和 $2 \times 2 \times 2$,其中 $3 \times 3 \times 3$ 卷积核聚焦目标区域的局部纹理、边缘及微小结构(如脑部MRI中海马体的局部形态变化等),实现对局部空间模式的精细刻画; $1 \times 1 \times 1$ 卷积核不改变空间维度,通过对输入特征的所有通道进行线性组合,实现跨通道信息的交互,融合不同模态下的共性病理特征等关联性信息,同时降低计算复杂度; $2 \times 2 \times 2$ 卷积核则侧重捕捉中等范围的局部结构特征,相较于 $3 \times 3 \times 3$ 卷积核感受野更小,能更精准地定位较小尺寸的目标结构,在保留局部特征连续性的基础上平衡特征分辨率与计算效率。然后,可以得到LFL模块得到的局部特性 $F_1 \in \mathbb{R}^c$ 。

(3)关键因素模块(KFM):关键因素模块是利用多尺度池化与元素级乘法对特征进行学习的模块,如图4所示。具体来说,首先获得不同模态的显著特征 H_{T1}, H_{T2} ,同时使用最高池化(Max pooling)与平均池化(Average pooling)降维单元分别对特征进行降维,以获得降维后的尺度特征 $H_{T1M}, H_{T1A}, H_{T2M}, H_{T2A} \in \mathbb{R}^{c \times d}$ 。最高池化(Max pooling)提取局部极值特征,对细节敏感;平均池化提取全局平均特征,对整体结构鲁棒。两者结合,通过元素级乘法强化特征间的关联性,使用乘法融合的方法将 H_{T1M} 与 H_{T1A} 及 H_{T2M} 与 H_{T2A} 对应位置的元素相乘,从而得到新的融合特征向量 H_{T1Z} 和 H_{T2Z} ,即

$$H_{T1Z} = H_{T1M} \times H_{T1A} \quad (10)$$

$$H_{T2Z} = H_{T2M} \times H_{T2A} \quad (11)$$

式中“ $\times$ ”表示元素级的乘法运算。这样,就得到了两个新的融合特征向量 H_{T1Z} 和 H_{T2Z} ;乘法融合可以突显那些在所有模态或特征上都较为重要的信息,保留特征中的细节内容。然后对新的融合特征向量 H_{T1Z} 和 H_{T2Z} 进行Flatten特征展平操作,对融合特征进行求和,利用池化和 L_2 归一化得到最终的外部特征 $F_K = \|\text{sumpooling}(H_{T1} + H_{T2})\|_2$ 。最后,将 F_s, F_L 和 F_K 拼接,得到融合特征 F_c ,进行AD脑疾病分类。

2 实验结果及分析

2.1 实验设置

在MFN模型训练中,初始学习率设置为0.001,并通过Lion算法进行对模型参数动态调整。训练时采用固定的批次大小(Batch size=16),总训练轮次(epochs)设置为100。特别地,在前10个训练轮次中,运用Lion算法动态调整学习率,以促进模型的初始收敛;后续训练阶段中,继续沿用该优化策略以进一步提升模型性能。此外,模型采用交叉熵损失函数进行优化,以增强泛化能力并减少过拟合风险。

为了确保实验结果的可靠性,采用五折交叉验证方法进行模型训练与评估,所有对比实验均在相同的数据集上完成。使用3个关键指标评估模型性能:准确率(ACC)反映整体分类正确率,计算公式为 $(TP+TN)/(TP+TN+FP+FN)$;敏感度(SEN)衡量模型识别正类样本的能力,计算公式为 $TP/(TP+FN)$;特异性(SPE)评估模型识别负类样本的能力,计算公式为 $TN/(TN+FP)$,其中TP、TN、FP和FN分别表示真正例、真负例、假正例和假负例。

2.2 对比实验

为了验证本文方法的有效性,将提出的网络模型与下列方法进行对比。机器学习对比方法使用支

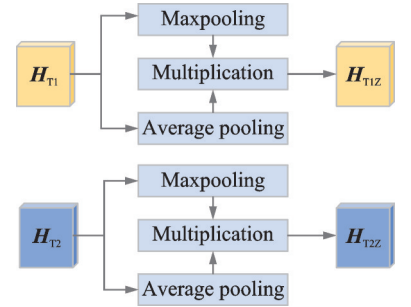


图4 关键因素模块

Fig.4 Key factor module

持向量机(Support vector machine, SVM)和随机森林(Random forest, RF)。SVM是通过核函数将数据映射到高维空间,寻找最优超平面实现分类;RF通过构建多棵决策树并投票输出结果,具有抗过拟合和高鲁棒性的特点,适用于高维数据和复杂分类任务。深度学习方法使用6种比较方法,具体包括:(1)CNN-EL^[20]通过集成多平面(矢状面、冠状面、横断面)MRI训练的CNN模型,并利用交点最多的脑区定位AD早期诊断的关键区域,从而实现AD疾病分类;(2)MADNet^[11]使用多模态数据,输入至ResNet网络,并采用双分支结构并行提取图像特征,最后实现AD分类;(3)MUSAN^[21]以全脑sMRI为输入,实现脑区特征的自动提取和不同尺度脑区之间相关性的建模,实现了对脑区的个性化疾病解读;(4)Ravi等^[22]选择对sMRI中的2D图像进行分析,利用ResNet-50v2模型中的残差块来进行疾病分类;(5)Zhang等^[23]设计了一个基于ResNet的双网络模型,分别提取多模态图像特征,通过注意力机制融合多层特征以捕捉细微变化,并采用空间金字塔池化降维,最终结合多模态特征实现AD诊断;(6)Tang等^[12]使用3DCNN来提取多模态数据的特征,然后使用Transformer的编码器块来提取各种特征之间的交互信息,最终将特征用于疾病分类。最终分类任务对比结果如表2所示。由表2可以看出,相比于其他AD疾病分类方法,本文提出的方法准确率最高,达到93.8%,训练集正确率及损失曲线如图5所示。

2.3 消融实验

为评估不同模块对分类性能的影响,可以通过消融实验验证单个模块及其组合对提升模型预测的能力。本文在ADNI数据集上进行了消融实验,结果如表3所示,其中‘√’表示使用此组件。分类结果显示,仅使用单一模态数据,神经网络的分类指标均出现不同程度的下降,仅使用T2模态相较于仅使用T1模块效果好,这是因为T1图像相比于T2图像具有较大尺寸,所以在输入卷积网络时,模型难以捕获全脑范围内的细微特征变化,而T2模态的AD患者的脑室扩张、白质高信号等病变在图像中呈现显著的高对比度变化,这种局部显著特征更易被卷积网络的卷积核捕捉。从临床角度看,AD患者的核心病理改变不仅包括海马体萎缩等结构性病变,还伴随脑室扩张、白质高信号等典型影像学特征,这些病变在T2模态中呈现出更显著的对比度。相比之下,T1模态虽能清晰显示脑解剖结构,但AD早期的细微结构改变(如轻度海马体萎缩)在T1图像中对比较低,且全脑范围内的结构性特征分布较广,模型需处理更大尺寸的图像信息,难以高效聚焦于早期病变的细微变化。这种临床病理特征与模态成像特性的差异,导致仅使用T2模态时,模型能更精准地捕捉AD患者的典型病变信号,从而获得更优的分类效果。使用多模态数据相比于使用单模态数据更有助于提升网络的疾病识别率。本文同时对不同

表2 本文方法与已有方法的比较结果

Table 2 Comparison results of the proposed method and existing methods				%
方法	方法分类	ACC	SEN	SPE
SVM	Machine learning	62.2	62.9	61.8
RF	Machine learning	64.7	65.0	64.0
CNN-EL ^[20]	Deep learning	82.4	82.1	81.6
MADNet ^[11]	Deep learning	91.3	89.2	92.7
MUSAN ^[21]	Deep learning	92.0	90.7	93.3
ResNet-50v2 ^[22]	Deep learning	89.9	90.1	89.7
Zhang等 ^[23]	Deep learning	90.9	92.8	88.9
Tang等 ^[12]	Deep learning	92.7	92.8	92.6
Ours	Deep learning	93.8	93.5	94.1

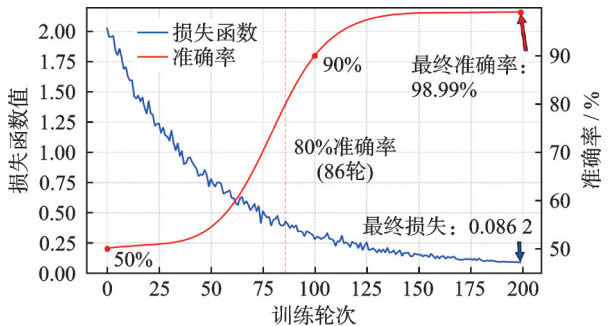


图5 训练集正确率及损失曲线

Fig.5 Training set accuracy and loss curves

的融合方法进行了消融实验,使用两种融合方法相比于同时使用全部的融合方法,神经网络的分类指标均出现不同程度的下降。实验结果表明,同时使用多种融合方法可获得最佳的性能。此外,使用 GIF 和 LFL 的性能仅仅次于同时使用这 3 种方法,这可能是由于全局和局部的互补优势,可以有效地提高网络分类性能。对于 LFL 模块卷积核大小的选择,设计了组合 A: $[3\times3\times3,1\times1\times1,2\times2\times2]$ 、组合 B: $[3\times3\times3,5\times5\times5,7\times7\times7]$ 、组合 C: $[1\times1\times1,3\times3\times3,4\times4\times4]$ 和组合 D: $[2\times2\times2,4\times4\times4,6\times6\times6]$ 这 4 种方案,在 ADNI 数据集上进行阿尔茨海默病分类任务实验。实验结果显示,组合 A 在 ACC、SPE 和 SEN 等多项指标上均优于其他组合,准确率达到 93.8%,而组合 B、C、D 的最高准确率分别为 84.5%、85.8%、83.1%。

表 3 不同模块对网络性能的影响
Table 3 Influence of different modulus on network performance

分类任务	模态		方法			ACC	SPE	SEN
	T1	T2	GIF	LFL	KFM			
AD/CN	✓		✓	✓	✓	74.2	73.5	74.4
		✓	✓	✓	✓	76.7	77.1	76.4
	✓	✓		✓	✓	84.6	85.0	84.5
	✓	✓	✓		✓	81.5	81.4	80.8
	✓	✓	✓	✓		86.7	86.4	86.0

3 结束语

本文提出了一种新的多模态融合分类算法用于 AD 的早期诊断,利用基于 3D ResNet18 的骨干网络提取多模态特征,通过跨模态特征集成模块中的全局信息融合模块(GIF)、局部特征学习模块(LFL)以及关键因素模块(KFM)三个子模块,有效融合多模态数据特征,增强了从全局、局部和潜在学习多个方面的特征表示。ADNI 数据的实验表明,本文方法在 AD 诊断相关的分类任务中优于其他方法。

在消融实验部分,发现多模态图像的预测准确率要高于单模态图像,这是因为 T1 加权 MRI 侧重呈现脑解剖结构细节,能清晰显示 AD 患者典型的脑萎缩病理改变,如海马体、内嗅皮层等关键脑区的体积缩小、灰质密度降低,但 T1 模态对组织微观病理的敏感性较低,难以早期识别尚未出现明显结构萎缩的病变。T2 加权 MRI 则对组织含水量变化高度敏感,可有效捕捉 AD 患者脑内白质高信号、小血管病变及局部炎症反应等病理特征,这些特征与 AD 的神经退行性变、血管因素协同作用密切相关,且常早于明显的结构萎缩出现。然而,T2 模态的空间分辨率较低,对解剖结构的精细定位能力弱于 T1 模态。现有单一模态分析或简单融合方法(如早期拼接、晚期聚合)无法充分发挥两者的互补性:仅用 T1 可能遗漏早期白质病变,仅用 T2 难以精确定位萎缩脑区;而 CFIM 通过 3 个模块的协同设计,实现了对两种模态特征的深度整合——GIF 的全局注意力机制可聚焦海马体萎缩(T1)与白质高信号(T2)的空间关联,LFL 的多尺度卷积能同时捕捉 T1 的结构细节和 T2 的局部信号异常,KFM 则挖掘出大脑结构萎缩区域的白质信号强度变化。这种融合不仅整合了两种模态的显性特征,更通过模态间潜在关系的建模提升了 AD 疾病预测准确性。

在特征融合上,MUSAN^[21]使用单模态方法,单模态数据仅依赖单一来源(仅使用 T1 MRI 图像或 T2 MRI 图像),相比多模态数据整合 T1、T2 及正电子发射断层扫描(Positron emission computed tomography, PET)等多源信息,其疾病特征维度有限,分类效果也不如多模态数据;在多模态数据方面,

相比于本文提出的跨模态特征集成模块, Wang等^[16]使用利用sMRI、临床特征和遗传多态性数据准确预测MCI向AD的长期转化,但图像和临床信息属于不同类型的数据,其内在的特征关联较为复杂,通过拼接等方法进行特征融合,后进行疾病预测,可能对多模态的信息融合不完全,难以捕获不同信息; Tang等^[12]等使用3DCNN来提取多模态数据的特征,然后使用Transformer的编码器块来提取各种特征之间的交互信息,用于疾病分类,相比于本文方法,使用了QKV机制,未使用LFL和KFM模块,对特征的局部信息与关键信息可能较难提取,从而降低了对AD疾病的分辨率。

然而,本研究仍存在局限性,未来需进一步探索改进方向。实际临床场景中,多模态磁共振成像(MRI)数据采集成本高、流程复杂,限制了算法的普及性,模型推理速度与实时性需求存在差距,本研究仅融合T1与T2模态数据,未来可纳入PET数据,进一步提升AD病理标志物的捕捉能力,增强模型对早期微小病变的敏感性。

参考文献:

- [1] JACK C R, JR, HOLTZMAN D M. Biomarker modeling of Alzheimer's disease [J]. *Neuron*, 2013, 80(6): 1347-1358.
- [2] 戴建英, 王辉, 刘敏, 等. 阿尔茨海默病的病理模型及代谢组学研究进展[J]. *药学服务与研究*, 2021, 21(4): 241-247, 320.
DAI Jianying, WANG Hui, LIU Min, et al. Research progress in pathological models and metabolomics of Alzheimer's disease[J]. *Pharmaceutical Care and Research*, 2021, 21(4): 241-247, 320.
- [3] PORTET F, OUSSET P J, VISSER P J, et al. Mild cognitive impairment (MCI) in medical practice: A critical review of the concept and new diagnostic procedure. Report of the MCI Working Group of the European Consortium on Alzheimer's Disease [J]. *Journal of Neurol Neurosurg Psychiatry*, 2006, 77(6): 714-718.
- [4] MCDADE E, BATEMAN R J. Stop Alzheimer's before it starts[J]. *Nature*, 2017, 547(7662): 153-155.
- [5] 郝小可, 谭麒豪, 李家旺, 等. 基于噪声混淆增强特征鲁棒性的脑疾病预测[J]. *数据采集与处理*, 2023, 37(4): 776-786.
HAO Xiaoke, TAN Qihao, LI Jiawang, et al. Brain disease prediction based on noise confusion to enhance robustness of features[J]. *Journal of Data Acquisition and Processing*, 2023, 37(4): 776-786.
- [6] 徐鹤, 郑群力, 谢作玲, 等. 基于知识表示向量的可解释深度学习模型及其疾病预测应用[J]. *数据采集与处理*, 2024, 38(4): 777-792.
XU He, ZHENG Qunli, XIE Zuoling, et al. Interpretable deep learning model based on knowledge representation vectors and its application in disease prediction[J]. *Journal of Data Acquisition and Processing*, 2024, 38(4): 777-792.
- [7] DAMULINA A, PIRPAMER L, SOELLRADL M, et al. Cross-sectional and longitudinal assessment of brain iron level in Alzheimer disease using 3-T MRI[J]. *Radiology*, 2020, 296(3): 619-626.
- [8] JIANG Y, LI W, QIN Y, et al. In vivo characterization of magnetic resonance imaging-based T1w/T2w ratios reveals myelin-related changes in temporal lobe epilepsy[J]. *Human Brain Mapping*, 2023, 44(6): 2323-2335.
- [9] HASSAN N, MIAH A S M, SUZUKI K, et al. Stacked CNN-based multichannel attention networks for Alzheimer disease detection[J]. *Scientific Reports*, 2025, 15(1): 5815.
- [10] NIE Y, CUI Q, LI W, et al. MHAGuideNet: A 3D pre-trained guidance model for Alzheimer's disease diagnosis using 2D multi-planar sMRI images[J]. *BMC Medical Imaging*, 2024, 24(1): 338.
- [11] LI Y, CHEN G, WANG G, et al. Dominating Alzheimer's disease diagnosis with deep learning on sMRI and DTI-MD[J]. *Frontiers in Neurology*, 2024, 15: 1444795.
- [12] TANG Y, XIONG X, TONG G, et al. Multimodal diagnosis model of Alzheimer's disease based on improved Transformer [J]. *Biomedical Engineering Online*, 2024, 23(1): 8.
- [13] ODUSAMI M, DAMAŠEVIČIUS R, MILIEŠKAITĖ-BELOUSOVIEŅĒ E, et al. Alzheimer's disease stage recognition from MRI and PET imaging data using Pareto-optimal quantum dynamic optimization[J]. *Heliyon*, 2024, 10(15): e34402.
- [14] 樊连玺, 刘彦北, 王雯, 等. 基于多模态表示学习的阿尔兹海默症诊断算法[J]. *计算机科学*, 2021, 48(10): 107-113.
FAN Lianxi, LIU Yanbei, WANG Wen, et al. Multimodal representation learning for Alzheimer's disease diagnosis[J]. *Computer Science*, 2021, 48(10): 107-113.

[15] ZHANG L, WANG L, GAO J, et al. Deep fusion of brain structure-function in mild cognitive impairment[J]. Medical Image Analysis, 2021, 72: 102082.

[16] WANG Y, GAO R, WEI T, et al. Predicting long-term progression of Alzheimer's disease using a multimodal deep learning model incorporating interaction effects[J]. Journal of Translational Medicine, 2024, 22(1): 265.

[17] LIU M, CHENG D, WANG K, et al. Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis[J]. Neuroinformatics, 2018, 16(3): 295-308.

[18] QIU S, MILLER M I, JOSHI P S, et al. Multimodal deep learning for Alzheimer's disease dementia assessment[J]. Nature Communications, 2022, 13(1): 3404.

[19] TONG T, GRAY K, GAO Q, et al. Multi-modal classification of Alzheimer's disease using nonlinear graph fusion[J]. Pattern Recognition, 2017, 63: 171-181.

[20] PAN D, ZENG A, JIA L, et al. Early detection of Alzheimer's disease using magnetic resonance imaging: A novel approach combining convolutional neural networks and ensemble learning[J]. Frontiers in Neuroscience, 2020, 14: 259.

[21] FAN X, LI H, LIU L, et al. Early diagnosing and transformation prediction of Alzheimer's disease using multi-scaled self-attention network on structural MRI Images with occlusion sensitivity analysis[J]. Journal of Alzheimers Disease, 2024, 97(2): 909-926.

[22] L S, V S, RAVI V, et al. Deep learning-based approach for multi-stage diagnosis of Alzheimer's disease[J]. Multimedia Tools and Applications, 2024, 83(6): 16799-16822.

[23] ZHANG Y, HE X, LIU Y, et al. An end-to-end multimodal 3D CNN framework with multi-level features for the prediction of mild cognitive impairment[J]. Knowledge-Based Systems, 2023, 281: 111064.

作者简介:



朱厚元(2001-),男,硕士研究生,研究方向:医学人工智能,E-mail:1017670799@qq.com。



郑乐乐(2000-),男,硕士研究生,研究方向:医学人工智能。



商浩(2000-),男,硕士研究生,研究方向:医学人工智能。



臧雪峰(2000-),男,硕士研究生,研究方向:医学人工智能。



吴少琪(2000-),女,硕士研究生,研究方向:医学人工智能。



周广超(2002-),男,硕士研究生,研究方向:医学人工智能。



孙建德(1978-),通信作者,男,教授,研究方向:多模态数据处理、分析理论及应用研究,E-mail:jiandesun@sdnu.edu.cn。



乔建苹(1981-),通信作者,女,副教授,研究方向:多模态脑影像处理与分析,E-mail: qiaojianping@sdnu.edu.cn。

(编辑:王静)