

融合多模态信息与位置编码的阿尔茨海默病诊断

刘蓉, 刘汝璇, 李广昶, 柴新宇, 谭桂梅, 唐奇伶*

(中南民族大学 a. 生物医学工程学院; b. 认知科学国家民委重点实验室, 武汉 430074)

摘要 阿尔茨海默病(Alzheimer's Disease, AD)作为一种致命的神经退行性疾病,其早期诊断与病理区域的精确预测对于延缓病情进展和改善患者预后具有极其重要的意义,尽管过去的研究已经在自动化诊断技术上取得了进展,现有方法在疾病的诊断准确率已经有着不错的水准,但其模型的可解释性仍是困扰临床研究的最大问题. 针对这一背景,提出了一种结合三维位置编码与多模态的阿尔茨海默病诊断模型,该模型将三维位置编码、Transformer自注意力机制和全卷积网络(FCN)有机结合,能够从三维医学影像数据中自动提取有效特征,生成代表整个大脑的高分辨率疾病概率图,并通过多模态注意力机制将此概率图与客观临床信息有机融合,实现对AD的精准预测诊断的同时,为模型决策过程提供更多的可解释层面.

关键词 阿尔茨海默病; 磁共振影像; 全卷积网络; 三维位置编码; 多模态注意力

中图分类号 TP391 文献标志码 A 文章编号 1672-4321(2026)02-0212-09

doi: 10.20056/j.cnki.ZNMDZK.20250843

Diagnosis of Alzheimer's disease via fusion of multimodal information and position encoding

LIU Rong, LIU Ruxuan, LI Guangchang, CHAI Xinyu, TAN Guimei, TANG Qiling*

(South-Central Minzu University, a. School of Biomedical Engineering; b. Key Laboratory of Cognitive Science of State Ethnic Affairs Commission, Wuhan 430074, China)

Abstract Alzheimer's Disease (AD), as a fatal neurodegenerative disease, holds immense significance for early diagnosis and precise prediction of pathological regions in delaying disease progression and improving patient prognosis. Although past research has made progress in automated diagnostic technologies, the interpretability of existing methods remains the most significant issue troubling clinical studies, despite their commendable diagnostic accuracy. Against this backdrop, a diagnostic model for Alzheimer's Disease is proposed that integrates three-dimensional position encoding with multimodal data. The model combines three-dimensional position encoding, Transformer self-attention mechanisms and Fully Convolutional Networks (FCN) to automatically extract effective features from three-dimensional medical imaging data, generating high-resolution disease probability maps representing the entire brain. Through a multimodal attention mechanism, the probability map is organically integrated with objective clinical information, achieving precise predictive diagnosis of AD while providing more interpretable aspects for the model's decision-making process.

Keywords Alzheimer's disease; magnetic resonance imaging; Fully Convolutional Network; three-dimensional position encoding; multimodal attention

阿尔茨海默病(Alzheimer's Disease, AD)是一种慢性神经退行性疾病,随着全球人口老龄化进程的不断加速,患者数量持续增加. 根据世界卫生组织的数据,目前全球约有5000万人饱受阿尔茨海默

病之苦,预计到2050年,这一数字将近乎翻倍. 阿尔茨海默病给患者带来认知、记忆和行为能力的重大障碍^[1],推动阿尔茨海默病的早期诊断和科学治疗策略的研发显得尤为紧迫^[2].

收稿日期 2025-03-11

* 通信作者 唐奇伶(1973-),男,副教授,博士,研究方向:医学图像分析, E-mail: qltang@mail.scuec.edu.cn

基金项目 湖北省重点研发计划资助项目(2022BAA037);中央高校基本科研业务费专项资金资助项目(CZQ24015)

在阿尔茨海默病患者的大脑中,异常蛋白质以淀粉样斑块和神经纤维缠结的形态积聚,引发神经元的丧失.这些脑部变化通常在记忆问题显现之前就已出现,主要影响到颞叶和海马体等区域.随着医学影像技术的快速发展,磁共振成像(Magnetic Resonance Imaging, MRI)^[3]已经成为研究阿尔茨海默病脑部病变的重要工具.尽管sMRI对于观察大脑结构变化具有非常高的灵敏度,但单纯的影像特征,如海马体萎缩程度等,通常难以直接作为诊断阿尔茨海默病的充分依据.因此,如何从复杂的脑部影像数据中提取有助于阿尔茨海默病早期诊断的有效特征,已成为当前研究的焦点之一.

另一方面,深度学习模型在进行疾病诊断^[4]和影像分析时存在一定的“黑箱”问题^[5].尽管这些模型能够以惊人的准确度进行疾病诊断和影像分析,但由于其复杂的结构和运作方式,往往很难理解模型是如何做出特定的诊断或分类决策的.这种不透明性限制了这些模型在临床实践中的应用,因为医疗行业对于诊断结果的可解释性和透明度有着极高的要求.

基于上述背景与需求,本文提出一种融合多模态信息与位置编码的阿尔茨海默病诊断方法,主要创新点如下:

(1) 三维位置编码与图像块融合.精准地定位

与阿尔茨海默病相关的病理性变化区域;

(2) 全卷积网络(Fully Convolutional Network, FCN)^[6]融合Transformer的自注意力机制,有助于网络更好地提取跨尺度的特征,更容易捕捉疾病在大脑中的整体分布规律.最终网络模型提供更多的解释层面;

(3) 影像与临床信息多模态融合.将代表大脑疾病信息的高风险体素点与临床信息相结合,为疾病诊断提供更全面的决策依据.

1 结合三维位置编码与多模态的阿尔茨海默病诊断模型

本研究提出基于三维图像分块的影像分析方法.融合多模态信息与位置编码的阿尔茨海默病诊断模型整体框架如图1所示.首先对结构磁共振成像(sMRI)数据进行网格采样,采用 $45 \times 45 \times 45$ 体素(voxel)的非重叠分块策略,对边缘区域实施零填充(zero-padding)处理保证全脑覆盖.模型训练阶段,通过随机采样策略从完整sMRI数据中获取训练样本块.特征提取网络由FCN与三维位置编码模块构成,通过迭代处理提取具有病理鉴别力的局部语义特征,并构建高维特征空间中的疾病概率映射.

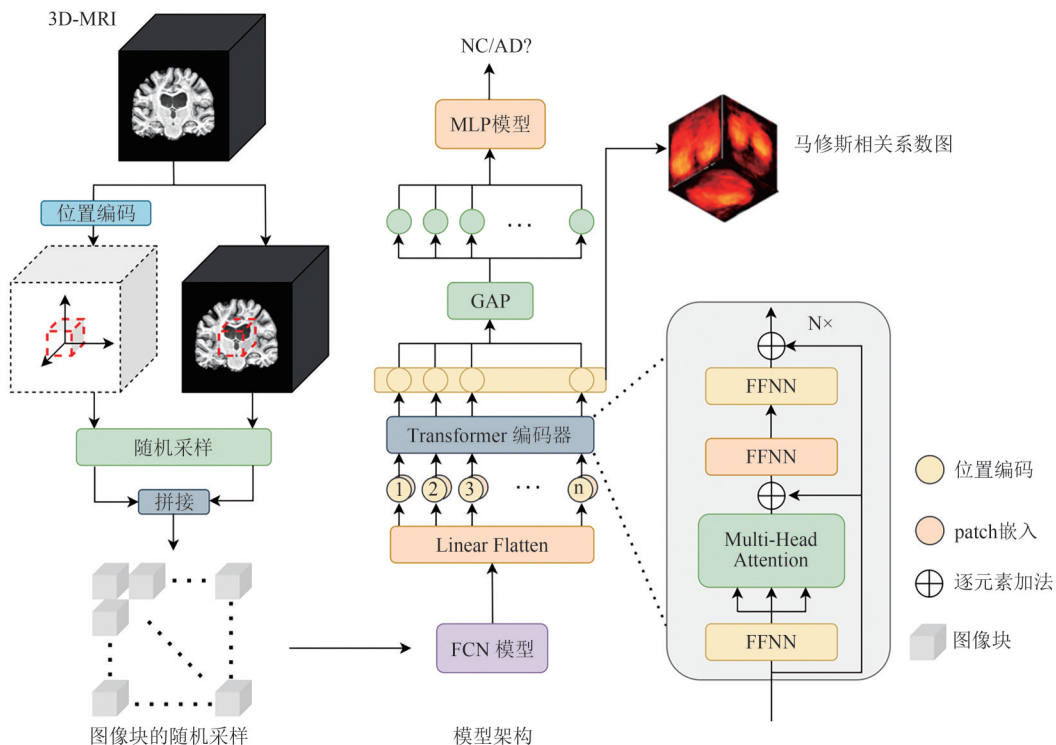


图1 融合多模态信息与位置编码的阿尔茨海默病诊断模型整体框架

Fig. 1 Holistic framework of Alzheimer's disease diagnosis model via fusion of multimodal information and position encoding

分类模块采用全局平均池化(Global Average Pooling, GAP)层与融合多模态注意力机制的多层感知器(MLP),实现阿尔茨海默病与正常对照(Normal Contrast, NC)的二元分类.在推理阶段,固定网络参数后,通过逆分块操作将Transformer编码器输出的特征块按解剖位置重组,生成保持原始空间拓扑关系的三维疾病概率分布图,实现病理区域的可视化定位.

1.1 三维位置编码

虽然FCN在提取局部特征时具备良好的鲁棒性和泛化能力,但它在生成图像块级别的响应时没有考虑每个图像块的具体空间位置信息,导致一些含有关键病变信息的图像块未能得到充分重视,从而可能造成重要诊断线索的遗漏.为此,文中进一步引入了基于元素空间位置的三维位置编码机制,以弥补FCN在处理时忽略绝对位置信息的不足.

本文构建了一个能在三维笛卡尔空间中实现完全平移不变性的表示方法,在这个表示中三维笛卡尔空间的坐标被组织成一个四维张量 $I \in L^{3 \times W \times H \times D}$,其中包含了3个坐标通道:冠状面、矢状面和轴位面.例如, $I^{\text{Coronal}} \in L^{1 \times W' \times H' \times D'}$ 其在第一冠状平面上的所有像素值设置为1,第二冠状平面上的像素值为2,第三冠状平面上的像素值为3,依此类推,直至覆盖所有冠状层面.同样,矢状面和轴位面的坐标信息也在各自对应的通道中以同样的方式进行填充和归一化处理.通过联合这3个坐标通道,就形成了一个全面表述三维笛卡尔空间中每一个体素位置信息的四维张量 I ,它可以用来精确指示大脑中任意一个三维图像块在整体空间中的确切位置,使模型能够更高效地利用空间位置信息去指导和优化特征提取.

设 V 为整个三维sMRI图像体素空间,对于每一个训练样本,定义输入图像 X 的维度为 $181 \times 217 \times 181$ 的体素空间,从 V 中随机选取 N 个体积图像块 P_i ,其中 $i=1, 2, \dots, N, N=125$,每个图像块的尺寸定义为 $45 \times 55 \times 45$,图像块随机采样策略可以用以下公式表示:

$$P(V) = \{P_1, P_2, \dots, P_{125} | P_i \subseteq V \wedge \forall i \neq j, P_i \cap P_j = \emptyset\}, \quad (1)$$

其中, $P(V)$ 是从sMRI图像体素空间 V 中生成的图像块集合, $P_i \subseteq V$ 表示每个图像块 P_i 是体素空间 V 的一个子集, $P_i \cap P_j = \emptyset; \bigcup_{i=1}^{125} P_i = V$ 表明所有的图像

块覆盖了整个图像体素空间 V .

对于每个训练样本,设 P_i 是提取的图像块, L_i 是相应的三维位置编码,将图像块和位置编码相拼接后得到 T_i ,表示为:

$$T_i = \text{Concat}(P_i, L_i). \quad (2)$$

由于所有的图像块 T_i 都是独立的并以同样的方式在FCN模型上处理,整个三维sMRI图像体素空间的输出可以表示为特征映射集合 $\{F_1, F_2, \dots, F_{125}\}$,每个 F_i 都是一个分辨率为 $9 \times 11 \times 9$ 的特征映射,再经过线性扁平化(linear flatten)操作,将FCN模型输出的特征映射 F_i 转换为一维特征向量 F'_i :

$$F'_i = \bigcup_{i=1}^{125} L(F_i), \quad (3)$$

其中, L 代表扁平化处理, F' 是扁平化后的特征向量,其维度是 125×891 .

1.2 特征提取网络

虽然FCN在提取局部特征方面表现优异,但缺乏对空间位置关系的建模能力.受启发于视觉Transformer在处理自然图像任务中展现的卓越的全局依赖捕捉能力,本文设计了一种融合了FCN和Transformer两大框架优势的新颖网络架构,最大化地利用FCN在进行局部特征提取方面的优势,以及Transformer的自注意力机制可以捕捉图像块之间长距离依赖关系的能力.这种融合有助于网络更好地提取跨尺度的特征,从而更容易捕捉疾病在大脑中的整体分布规律.最终期望网络模型提供更多的解释层面,解决传统模型在可解释性方面的不足,在阿尔茨海默病的诊断中实现更高的精准度和可信度.

在输入Transformer编码器之前,为每个深层特征 x_p 添加位置嵌入.最终的嵌入 $z_0 = x_p + E_{\text{pos}}$,其中 E_{pos} 是位置编码,确保网络能充分利用空间上下文信息.接下来这些嵌入 z_0 被送入Transformer编码器.每个编码器包含两部分:多头注意力(Multi-Head Attention, MHA)模块和前馈神经网络(Feed-Forward Neural Network, FFNN).

(1)多头注意力(MHA)是一种扩展自注意力(Self-Attention)的技术,其目的是让模型能够同时关注不同表示子空间中的信息,在多头注意力中,有多个独立的自注意力模块(称为“头”),每个头有自己独立的参数,可以并行地进行自注意力计算.多头注意力的流程可以表述如下:

首先,对输入数据进行线性变换,得到每个头的查询(Query)、键(Key)和值(Value):

$$Q_i = XW_{Q_i}, \quad (4)$$

$$K_i = XW_{K_i}, \quad (5)$$

$$V_i = XW_{V_i}, \quad (6)$$

其中, X 是输入, $W_{Q_i}, W_{K_i}, W_{V_i}$ 是每个头的查询、键、值的权重矩阵.

使用 Q_i, K_i, V_i 计算自注意力:

$$\text{Attention}(Q_i, K_i, V_i) = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i, \quad (7)$$

其中, d_k 是键的维度, softmax 函数使得所有的权重之和为 1, 并且较大的权重对应的位置会被放大.

最后, 将所有头的注意力输出拼接在一起, 然后进行一次线性变换, 得到最终的输出:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) W_o, \quad (8)$$

其中, $\text{head}_i = \text{Attention}(Q_i, K_i, V_i)$, W_o 是输出的权重矩阵, h 是头的数量. 经过多头注意力模块, 模型可以在不同的表示子空间中学习和关注输入数据的不同部分.

(2) 前馈神经网络 (FFNN): 每个 MHA 的输出后通过一个 FFNN, 这是另一种非线性变换, 它通常包含两个线性层和一个激活函数.

MHA 和 FFNN 每个部分之后都有残差连接, 并跟随一个层归一化 (Layer Normalization), 表示为:

$$z' = \text{LayerNorm}(z + \text{Sublayer}(z)), \quad (9)$$

其中 z 是子层的输入, Sublayer 是 MHA 或 FFNN.

Transformer 编码器^[7]堆栈中的第 1 个块在经过层归一化、多头自注意力和 MLP 操作后, Transformer 编码器输出特征块并按照原始取块位置重新组合成表示脑影像数据的空间结构信息的 3D 立体图, 准确生成反映当前样本的大脑疾病分布的疾病概率图, 这种像素级别的疾病概率图能够提供最细致的病变分布信息, 极大提高了模型的可解释性; 同时根据每个单样本的疾病概率图可以构建代表本数据集全样本的 Matthews 相关系数 (Matthews Correlation Coefficient, MCC) 图^[8], 直观地显示模型在哪个脑区对阿尔茨海默病鉴别能力的强弱, 从而帮助理解哪些脑区对于识别疾病状态至关重要, 同时揭示模型在不同数据集上的稳健性和泛化能力. 在 Transformer 模型的末端, 本文放弃了传统的 class token^[9]的设计, 转而采用全局平均池化层 (Global Average Pooling, GAP). GAP 在维持对输入图像平移不变性的同时, 能够在某些任务中取得优于 class token 的成效.

1.3 融合多模态注意力机制的 MLP 分类模型

通过 FCN 与 Transformer 模型构建了高分辨率的大脑疾病风险概率图和 MCC 图后, 将这些高 MCC 值的特征位置 (即大脑中被认为最有可能出现 AD 病理变化的区域) 作为重点关注对象. 将这些位置对应的特征信息输入到多层感知器 (MLP) 模型中, 以进一步细化和优化对 AD 状态的预测. 首先从 MCC 图中获取相应数量具有高 Matthews 相关系数的体素位置, 然后使用这些体素位置对应每个样本相应位置的疾病概率体素值作为分类模型的输入, MLP 模型接收这些概率值作为输入后, 进一步使用结合年龄、性别和 MMSE 得分等非影像临床特征进行 AD 状态预测, 以期通过综合多种来源的信息, 更精确地对阿尔茨海默病状态进行诊断.

为整合影像学特征和非影像学特征 (年龄、性别和 MMSE 评分), 本文创新地采用了一种多模态注意力机制, 将其与 MLP 模型结合. 多模态注意力模型如图 2 所示.

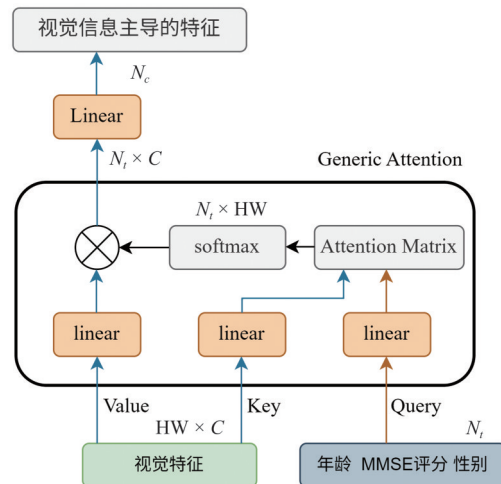


图 2 多模态注意力模型结构

Fig. 2 Structure of the multimodal attention model

通过这种框架不仅能够从复杂的神经影像数据中捕捉到与 AD 相关的微妙变化, 还能够将易于获得的临床数据融入模型决策过程, 从而提升模型诊断的准确性和可解释性. 将年龄、性别、MMSE 评分非影像数据转为语言模态的特征 N_i , 将其作为 Q 与 $HW \times C$ 的视觉特征交互, 将两者分别经过 Linear 层后相乘得到 $N_i \times HW$ 的特征矩阵, 再经过 softmax 与经过 Linear 的视觉特征 Value 相乘得到 $N_i \times C$ 的特征, 再经过 Linear 层后得到视觉信息主导的特征 (Vison-dominated Feature).

本文设计了两类分类模型架构, 如图 3 所示, 其一为多模态融合分类模型, 该模型整合了影像数据

特征与临床信息特征,其核心在于多模态注意力机制层,即前文所述的多模态注意力模型;其二为单模态分类模型,仅以影像数据为输入,即仅依据从sMRI图像中提取的疾病概率值特征进行诊断,未融合任何非影像临床信息。

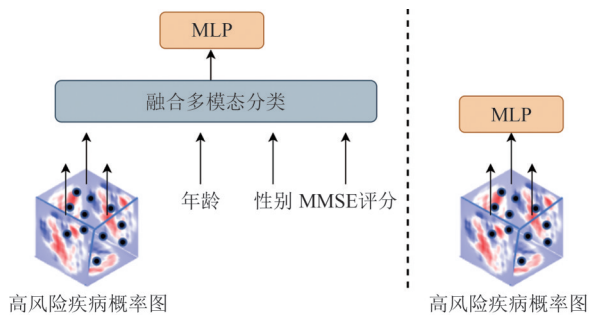


图3 多模态与单模态分类模型

Fig. 3 Multimodal and unimodal classification models

2 实验结果与分析

2.1 数据集及预处理

2.1.1 实验所用数据集介绍

为了深入探究阿尔茨海默病的神经影像特征,本研究挑选了3个公认且广泛使用的数据集:阿尔茨海默病神经影像学计划(ADNI)^[10];澳大利亚老龄化影像、生物标志物和生活方式研究(AIBL)^[11];美国国家阿尔茨海默病协调中心(NACC)数据库^[12],并根据ADNI基线招募标准,从ADNI、AIBL和NACC数据集中筛选出符合条件的55岁及以上参与者,分为阿尔茨海默病患者组和认知正常对照组。在排除混合型痴呆、其他类型痴呆及可能影响认知评估的中枢神经系统疾病或损伤,综合考虑结构磁共振数据及年龄、性别、MMSE评分等临床信息后,最终从ADNI筛选出417名、AIBL筛选出382名、NACC筛选出292名参与者,构建完整数据平台。实验数据信息如表1所示。

表1 实验数据集信息

Tab. 1 Experimental dataset information

Dataset	Category	Total (Male/ Female)	Age, median [range]	MMSE, median [range]
ADNI	NC	229(119/110)	76[60,90]	29[25,30]
	AD	188(101/87)	77[55,91]	23[18,28]
AIBL	NC	320(144/176)	72[60,92]	29[25,30]
	AD	62(24/38)	66[55,93]	21[6,28]
NACC	NC	151(49/102)	72[56,94]	29[21,30]
	AD	141(64/77)	76[55,95]	22[0,30]

2.1.2 数据集预处理

结构磁共振成像是研究大脑结构和神经变化的重要手段,然而sMRI数据的收集存在不少变数,如设备差异、操作标准不一,各个研究中心在成像协议上的差异以及个体大脑形态的天然差异等,因此本文对sMRI数据进行了一系列严格的标准化预处理操作,旨在通过标准流程消除或最小化设备与个体差异带来的影响,以确保数据的可比较性,包括:(1)使用SPM工具包对所有数据统一进行前连合(Anterior Commissure, AC)和后连合(Posterior Commissure, PC)校正,确保不同患者之间大脑解剖结构的一致性;(2)使用HD-BET工具进行颅骨分离,从sMRI图像中移除非脑组织,如颅骨、皮肤、脂肪和颅外软组织^[13],只保留脑组织的图像信息;(3)使用FSL^[14]软件进行仿射配准将sMRI图像与MNI152模板^[15]线性对齐,减低图像数据的整体性差异,同时对所有数据进行重采样,处理后的数据具有相同的空间分辨率即 $1 \times 1 \times 1 \text{ mm}^3$;4)将配准后的图像数据通过ANTs^[16]软件包中的N4BiasFieldCorrection工具进行偏置场校正来消除图像中由于设备硬件限制或磁场不均导致的非均匀亮度变化^[17]。

本研究中使用的模型均是以ADNI数据库为基础搭建而成,将ADNI数据库按照3:1:1的比例随机分为训练组(250例)、验证组(83例)和测试组(84例)。同时,为了验证模型的泛化能力,还采用了两个独立的数据库(AIBL和NACC)进行了测试,样本数分别为382例和292例。在网络训练方面,采用了Adam优化器和交叉熵与熵损失作为混合损失函数。为确保结果的准确性,在每次训练之前都会对训练集和验证集进行重新分配,这一过程重复了5次。模型性能的评估则是通过计算平均值及其标准偏差来呈现。此外,网络模型是从随机权重开始训练的,每批次处理10个样本,学习率设置为0.0001,总共迭代训练3000次。每进行20次迭代,就对验证集进行一次评估,并在取得最佳效果时保存模型的权重。最终通过测试集评估得出的网络模型在分类精度和疾病概率图特性方面的性能优劣。

2.2 评价指标

模型的性能通过模型预测的准确率(Accuracy, ACC)、敏感度(Sensitivity, SEN)、特异度(Specificity, SPE)、F1分数(F1-Score)和MCC来评估。F1-Score综合反映了模型的精确度和召回率,后者是衡量模型正确标识所有实际正例的关键指标。马修斯相关系数是一个用于衡量二分类问题中模型预测质量的

指标,其值介于-1和1之间.MCC考虑了真正例、假
 正例、真负例和假负例这4种情况,常用于衡量不平
 衡分类问题的性能,各指标的公式如下:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

$$SEN = \frac{TP}{TP + FN} \quad (11)$$

$$SPE = \frac{TN}{TN + FP} \quad (12)$$

$$F1 - Score = \frac{2TP}{2TP + FN + FP} \quad (13)$$

MCC =

$$\frac{TP \times TN - FP \times FN}{[(TP + FP)(TP + FN)(TN + FP)(TN + FN)]^{\frac{1}{2}}}, \quad (14)$$

其中,真正例(True Positives, TP)表示模型正确将正
 例实例预测为正例的数量,真负例(True Negatives,
 TN)表示模型正确地预测为负例的实例数量,假负
 例(False Negatives, FN)表示模型错误地预测为负例
 的实例数量,假正例(False Positives, FP)表示模型

错误地预测为正例的实例数量.

2.3 实验结果与分析

2.3.1 病理区域检测模型实验结果与分析

本实验采用 patch-wise 训练策略来训练局部
 FCN模型,并将其迁移到整体的sMRI影像模型中,
 以获取相应大脑区域中的疾病概率,进一步生成全
 脑AD病理概率分布图.本文将3D病理概率图按照
 冠状面、矢状面和轴位面进行展示.图4展示了临床
 确诊为阿尔茨海默症受试者的疾病概率图,红色区
 域表示阿尔茨海默症的高风险区,其推断的阿尔茨
 海默症概率>0.5,而蓝色区域表示低风险的病理区
 域,其推断的阿尔茨海默症概率<0.5.图5展示了临
 床诊断为正常受试者的疾病概率图.图6展示了
 AD/NC对照的4组对样本,前两个样本被临床诊
 断为阿尔茨海默症,而后两个样本被临床证实为
 认知正常的受试者.从AD与NC的对比可以看出FCN
 产生的高响应分布在具有解剖学意义的区域,例如
 海马、颞叶和顶叶区域.表明生成的病理概率图可
 以辨别出与AD相关的病理特征.

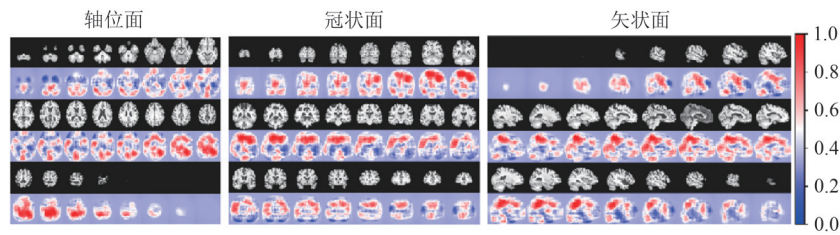


图4 阿尔茨海默病受试者的疾病概率图

Fig. 4 Disease probability map of subjects with Alzheimer's disease

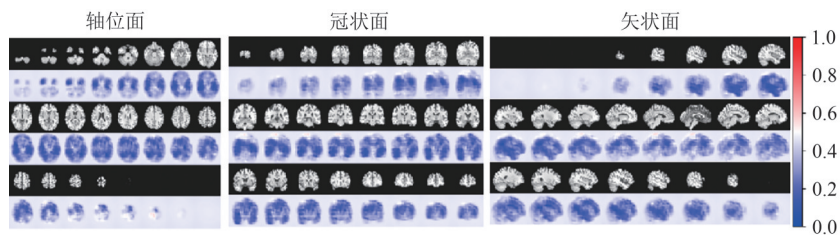


图5 正常受试者的疾病概率图

Fig. 5 Disease probability map of normal subjects

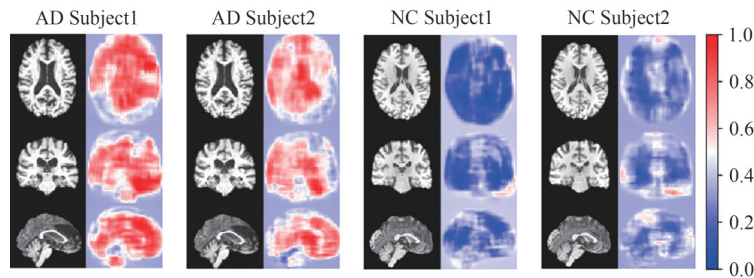


图6 AD/NC受试者的病理概率图

Fig. 6 Pathological probability map of AD/NC subjects

为了评估 FCN 生成病理概率分布图的阿尔茨海默症病理区域与神经解剖学的一致性,本文构建了 Matthews 相关系数(MCC)的全样本图,该样本图能够展示出每个数据集病理区域预测相较集中的

区域,可以作为一种辅助手段来直观地展示对阿尔茨海默病诊断影响最大的结构区域.图 7 显示了来自 ADNI 的训练集、验证集、测试集和 AIBL 及 NACC 数据集的 MCC 图.

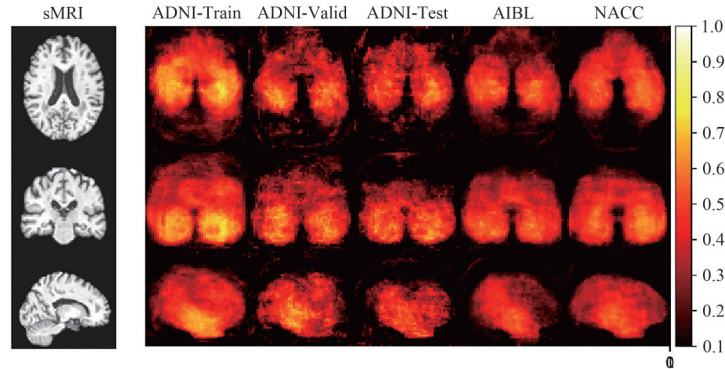


图 7 不同数据集的 MCC 图对比

Fig. 7 Comparison of MCC plots of different datasets

2.3.2 分类模型实验结果和分析

实验共采用了 3 种不同的分类诊断方式:首先是直接使用病理概率图提取相应的影像体素值作为分类模型的输入;其次是将非影像数据,包括年龄、性别和 MMSE 值作为分类模型的输入;最后是采用多模态的方式,将疾病概率值和临床信息共同作为分类模型的输入.

在使用疾病概率图作为输入时,本文特别关注疾病概率值的选取.首先从 MCC 图中获取相应数量具有高 Matthew's 相关系数的体素位置,然后使用这些体素位置对应每个样本相应位置的疾病概率体素值作为分类模型的输入.如图 8 所示,在不同体素数量输入的条件下,分类模型的诊断性能存在显著差异.经过综合评估,最终确定选取 200 个体素点作为分类模型的输入,此时模型的分性能达到最优.

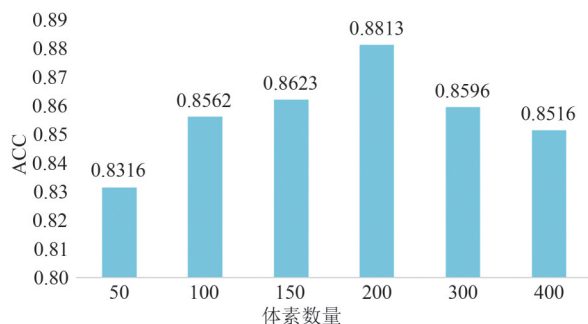


图 8 不同体素数量的模型准确率

Fig. 8 Accuracy of the model with different voxel counts

首先,在仅利用 sMRI 影像数据生成的疾病概率图中,本研究以提取的疾病概率值作为分类模型的输入,对分类模型进行了测试,并对其在 ADNI 测试集、AIBL 和 NACC 数据集上的分类性能进行了统计

分析.如表 2 所示,相较于基准 FCN 和三维位置编码的 FCN 模型,Transformer-FCN 融合模型在 ACC、SEN、SPE、F1 值以及 MCC 等指标上都有小幅提升,说明融合后的 FCN 模型有助于从疾病概率图中提取更具区分能力的特征,提高了分类性能.值得注意的是,三维位置编码增强了模型对解剖结构空间关系的感知,但在 AIBL 数据集中,因部分病例病理变化未引发显著结构位移,位置编码可能过度关注正常解剖约束,反而弱化了局部纹理特征(如灰质细微萎缩)的敏感性,致使 SPE 和 F1 值略有降低.

进一步将疾病概率值与临床信息(MMSE 评分、年龄和性别)一起作为多模态输入,输入到融合多模态注意力机制的 MLP 模型中,得出融合多模态数据的性能效果.如图 9 所示,多模态输入使模型的灵敏度特异性曲线(Area Under the Curve, AUC)和准确率-召回曲线(Area Under the Precision-Recall Curve, AUC-PR)均有提升,证明模型具备临床异构数据的泛化融合能力.

最终,将融合了多模态注意力机制的 MLP 模型与未添加多模态注意力机制的常规 MLP 模型,在多模态输入条件下展开对比研究,以探究多模态注意力机制对多模态融合所起到的促进作用.表 3 呈现了这两种模型在 3 个数据集上的定量分类指标对比情况.经对比发现,融合了多模态注意力机制的 MLP 模型,能够更为有效地捕捉影像信息与临床信息之间的高阶关联,推动两种模态信息实现相互补充,进而极大地提升了模型的分性能.

为了更清晰地体现本文提出的阿尔茨海默病诊断模型的优势,将本文实验结果与近几年关于阿

表 2 不同模型方法的性能对比

Tab. 2 Comparison of the performance of different model methods

Method	Dataset	ACC	SEN	SPE	F1-Score	MCC
FCN	ADNI-test	0.8042	0.7407	0.8561	0.7729	0.6031
	AIBL	0.8543	0.6290	0.8979	0.5838	0.4982
	NACC	0.8151	0.8012	0.7947	0.8137	0.6316
三维位置编码的 FCN	ADNI-test	0.8333	0.7870	0.8612	0.8095	0.6626
	AIBL	0.8421	0.6667	0.8760	0.5778	0.4896
	NACC	0.8291	0.7967	0.8322	0.8162	0.6296
三维位置编码的 Transformer-FCN	ADNI-test	0.8813	0.8241	0.8712	0.8317	0.6968
	AIBL	0.8770	0.6935	0.9125	0.6467	0.5748
	NACC	0.8368	0.8109	0.8609	0.8274	0.6735

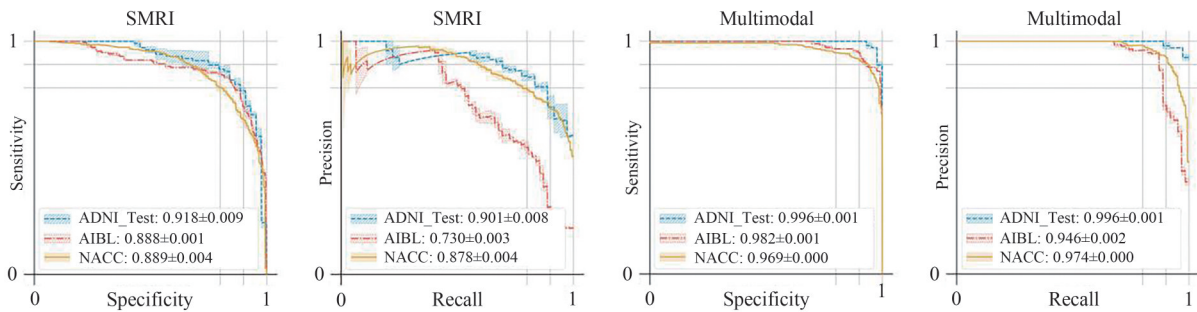


图 9 单模态输入和多模态输入的性能曲线对比

Fig. 9 Comparison of performance curves for unimodal inputs and multimodal inputs

表 3 融合了多模态注意力机制的 MLP 模型和普通 MLP 模型的性能对比

Tab. 3 Comparison of the performance of the MLP model integrated with the multimodal attention mechanism and the common MLP model

Method	Dataset	ACC	SEN	SPE	F1-Score	MCC
普通 MLP	ADNI-test	0.9427	0.8676	0.9014	0.8861	0.8037
	AIBL	0.8839	0.7861	0.9031	0.7267	0.6847
	NACC	0.8519	0.8231	0.8935	0.8434	0.7231
融合了多模态注意力机制的 MLP	ADNI-test	0.9758	0.9596	0.9655	0.9606	0.9326
	AIBL	0.9328	0.8462	0.9496	0.8048	0.7973
	NACC	0.9009	0.8586	0.9404	0.8931	0.8236

尔茨海默病识别和分类的文献结果进行了对比,见表 4. 本文提出的方法在 ACC、SEN、SPE 都达到了最优,印证了本文方法在阿尔茨海默病诊断领域的先进性与有效性.

表 4 阿尔茨海默病分类结果不同方法的比较

Tab. 4 Comparison of different methods for classification of Alzheimer's disease

Method	ACC	SEN	SPE
FENG et. al ^[18]	0.9658	0.9244	0.9322
GAO et. al ^[19]	0.9170	0.9350	0.9143
ODUSAMI et. al ^[20]	0.9183	0.8387	0.8943
BANGYAL et. al ^[21]	0.9463	0.9214	0.9416
SETHURAMAN et. al ^[22]	0.9661	0.9434	0.9486
SHOJAEI et. al ^[23]	0.9660	0.9517	0.9372
ZHOU et. al ^[24]	0.9432	0.9343	0.9653
Ours	0.9785	0.9596	0.9655

3 结语

为提升阿尔茨海默病诊断模型的精确性与可解释性,本文创新性地提出了一种融合多模态信息与位置编码的阿尔茨海默病诊断模型,相较于其他方法,在诊断准确率和可解释性方面均取得了较优的结果.但本研究仍有进一步完善和拓展的空间:例如,可以拓展到其他神经退行性疾病的诊断,如帕金森病等,通过适当调整和迁移模型,实现智能辅助诊断.此外,可以对数据特征进行队列追踪分析,长期追踪实验数据中变量的变化,比较和分析模型在时间轴上的稳定性和鲁棒性,探究模型的预测结果是否会随时间变化而波动等.

参 考 文 献

[1] HARDY J, SELKOE D J. The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics[J]. Science, 2002, 297(5580): 353-356.
 [2] SALEH H, ELRASHIDY N, ELAZIZ MABD, et al. Genetic algorithm-based hybrid deep learning model for explainable Alzheimer's disease prediction using temporal multimodal cognitive data [J]. International Journal of Data Science and Analytics, 2025, 20(2): 1073-1103.
 [3] SPERLING R A, AISEN P S, BECKETT L A, et al.

- Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease [J]. *Alzheimer's & Dementia*, 2011, 7(3): 280-292.
- [4] TU Y, LIN S, QIAO J, et al. Multimodal fusion diagnosis of Alzheimer's disease based on FDG-PET generation [J]. *Biomedical Signal Processing and Control*, 2024, 89: 105709.
- [5] JACK C R, KNOPMAN D S, JAGUST W J, et al. Tracking pathophysiological processes in Alzheimer's disease: An updated hypothetical model of dynamic biomarkers [J]. *The Lancet Neurology*, 2013, 12(2): 207-216.
- [6] 刘子仪, 唐奇伶, 蔡玉. 任务引导的径向基网络对乳腺病理图像有丝分裂检测[J]. *电子学报*, 2023, 51(11): 3146-3154.
- [7] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[J]. *arXiv: 2020.2010.11929*.
- [8] CHICCO D, TÖTSCH N, JURMAN G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation [J]. *BioData Mining*, 2021, 14(1): 13.
- [9] XU L, OUYANG W, BENNAMOUN M, et al. Multi-class token transformer for weakly supervised semantic segmentation [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022: 4300-4309.
- [10] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015: 3431-3440.
- [11] ADAMSON C, BEARE R, BALL G, et al. Callosal thickness profiles for prognosticating conversion from mild cognitive impairment to Alzheimer's disease: A classification approach[J]. *Brain and Behavior*, 2018, 8(12): e01142.
- [12] WEINER M W, VEITCH D P, AISEN P S, et al. The Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception[J]. *Alzheimer's & Dementia*, 2013, 9(5): e111-e194.
- [13] ELLIS K A, BUSH A I, DARBY D, et al. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease [J]. *International Psychogeriatrics*, 2009, 21(4): 672-687.
- [14] BEEKLY D L, RAMOS E M, LEE W W, et al. The national Alzheimer's coordinating center (NACC) database: The uniform data set [J]. *Alzheimer Disease & Associated Disorders*, 2007, 21(3): 249-258.
- [15] IGLESIAS J E, LIU C Y, THOMPSON P M, et al. Robust brain extraction across datasets and comparison with publicly available methods [J]. *IEEE Transactions on Medical Imaging*, 2011, 30(9): 1617-1634.
- [16] SMITH S M, JENKINSON M, WOOLRICH M W, et al. Advances in functional and structural MR image analysis and implementation as FSL [J]. *NeuroImage*, 2004, 23: S208-S219.
- [17] SONG J, ZHANG Z. Brain tissue segmentation and bias field correction of MR image based on spatially coherent FCM with nonlocal constraints [J]. *Computational and Mathematical Methods in Medicine*, 2019, 2019(1): 4762490.
- [18] FENG J, ZHANG S W, CHEN L, et al. Alzheimer's disease classification using features extracted from nonsubsampling contourlet subband-based individual networks [J]. *Neurocomputing*, 2021, 421: 260-272.
- [19] GAO X, SHI F, SHEN D, et al. Task-induced pyramid and attention GAN for multimodal brain image imputation and classification in Alzheimer's disease [J]. *IEEE Journal of Biomedical and Health Informatics*, 2022, 26(1): 36-43.
- [20] ODUSAMI M, MASKELIŪNAS R, DAMAŠEVIČIUS R, et al. Analysis of features of Alzheimer's disease: Detection of early stage from functional brain changes in magnetic resonance images using a finetuned ResNet18 network [J]. *Diagnostics*, 2021, 11(6): 1071.
- [21] BANGYAL W H, REHMAN N U, NAWAZ A, et al. Constructing domain ontology for Alzheimer disease using deep learning based approach [J]. *Electronics*, 2022, 11(12): 1890.
- [22] SETHURAMAN S K, MALAIYAPPAN N, RAMALINGAM R, et al. Predicting Alzheimer's disease using deep neuro-functional networks with resting-state fMRI [J]. *Electronics*, 2023, 12(4): 1031.
- [23] SHOJAEI S, SANIEE ABADDEH M, MOMENI Z. An evolutionary explainable deep learning approach for Alzheimer's MRI classification [J]. *Expert Systems with Applications*, 2023, 220: 119709.
- [24] ZHOU Z, WANG Q, AN X, et al. A novel graph neural network method for Alzheimer's disease classification [J]. *Computers in Biology and Medicine*, 2024, 180: 108869.