# BiSparseFusion：面向AD诊断的sMRI-fMRI跨模态双向稀疏交互融合模型

刘金平[1,2*]，李兴旺[1]，刘家瑜[1]，刘芷娴[3]，刘亚琴[1,4]

（1. 湖南师范大学信息科学与工程学院，湖南长沙 410081；2. 湖南省智能康复机器人与辅具器械工程技术研究中心，湖南长沙 410004；3. 湖南师范大学商学院，湖南长沙 410081；4. 湖南师范大学数学与统计学院，湖南长沙 410081）

**摘 要：** 阿尔茨海默病（Alzheimer's Disease，AD）是一种隐匿且不可逆的进行性神经退行性疾病，早期精准识别对于延缓病程进展和辅助临床干预具有重要意义。结构磁共振成像（structural Magnetic Resonance Imaging，sMRI）能够反映脑萎缩、灰质退化等解剖结构异常，功能磁共振成像（functional Magnetic Resonance Imaging，fMRI）能够刻画脑区间功能活动及动态交互模式，二者从结构与功能层面为AD诊断提供互补信息。然而，面向sMRI与fMRI的多模态联合诊断仍面临三方面关键挑战：一方面，3D sMRI与4D fMRI在空间分辨率、时间维度和信号分布上存在明显差异，如何构建统一的端到端跨模态建模框架仍较困难；另一方面，基于Vision Transformer的sMRI特征提取方法虽然具有全局建模能力，但标准多头注意力机制存在注意力头冗余、头间协同不足以及局部结构细节表达不充分等问题，限制了模型对海马体、嗅皮质等AD相关关键脑区的敏感性；此外，现有多模态融合方法多采用特征拼接、单向注意力或稠密交互策略，难以在高维异构影像特征中有效筛选关键区域并建立结构—功能之间的细粒度双向关联。针对上述问题，本文提出一种面向AD诊断的sMRI-fMRI跨模态双向稀疏交互融合模型BiSparseFusion。BiSparseFusion采用双分支端到端架构，在sMRI分支中构建集成动态可组合多头注意力机制（Dynamic Composable Multi-Head Attention，DCMHA）与多层特征融合模块（Multi-level Feature Fusion Module，MFFM）的3D Vision Transformer，通过动态组合不同注意力头之间的关系降低冗余响应，并利用多层特征聚合增强局部病灶细节与全局语义信息的一致表达；在fMRI分支中引入SwiFT模型直接对原始4D fMRI进行时空依赖建模，以避免传统功能连接或感兴趣区分析法对动态信息的压缩损失；在跨模态融合阶段设计双向稀疏交叉注意力融合模块（Bidirectional Sparse Cross-Attention Fusion，BSCAF），通过模态内稀疏注意力抑制冗余特征，并通过双向交叉注意力实现sMRI结构表征与fMRI功能表征之间的深度互补交互。DCMHA与MFFM为结构分支提供更具判别性的多尺度解剖特征，SwiFT为功能分支提供时空动态特征，BSCAF则在统一特征空间内完成关键模态信息筛选、对齐与融合，从而形成结构与功能协同的AD诊断框架。本文基于ADNI数据集在AD/NC、MCI/NC和AD/MCI三个分类任务上对所提方法进行验证，实验结果表明，BiSparseFusion分别取得97.67%、93.27%和96.72%的分类准确率，优于各种单模态和多模态对比模型。可视化结果表明，模型能够聚焦海马体、嗅皮质、杏仁核等与AD病理相关的脑区，并形成类别区分度更高的融合特征空间，验证了BiSparseFusion在多模态神经影像特征建模、跨模态细粒度融合以及AD辅助诊断任务中的有效性。

**关键词：** 阿尔茨海默病；多模态融合；结构磁共振成像；功能磁共振成像；动态可组合多头注意力；双向稀疏交叉注意力融合

**基金项目：** 国家自然科学基金（No.62371187）；湖南省智能康复机器人与辅具器械工程技术研究中心开放课题（No.2024JS101）

**中图分类号：** TP181 **文献标识码：** A **文章编号：** 0372-2112(XXXX)XX-0001-13

**电子学报URL:**http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20251113

# BiSparseFusion: A Cross-Modal Bidirectional Sparse Interaction Fusion Model for sMRI – fMRI-based Alzheimer's Disease Diagnosis

LIU Jinping[1,2*], LI Xingwang[1], LIU Jiayu[1], LIU Zhixian[3], LIU Yaqin[1,4]

(1. *College of Information Science and Engineering, Hunan Normal University, Changsha, Hunan* 410081, *China*;
2. *Hunan Intelligent Rehabilitation Robot and Auxiliary Equipment Engineering Technology Research Center, Changsha, Hunan* 410004, *China*;
3. *Business School, Hunan Normal University, Changsha, Hunan* 410081, *China*;
4. *School of Mathematics and Statistics, Changsha, Hunan* 410081, *China*. )

**Abstract:** Alzheimer's Disease (AD) is a latent and irreversible progressive neurodegenerative disorder. Early pre-

收稿日期：2026-03-01；录用日期：2026-04-24；责任编委：刘爽

*通信作者：刘金平

cise identification is crucial for delaying disease progression and supporting clinical intervention. Structural Magnetic Resonance Imaging (sMRI) can reveal anatomical abnormalities such as brain atrophy and gray matter degeneration, while Functional Magnetic Resonance Imaging (fMRI) captures functional activities and dynamic interactions between brain regions. Both provide complementary information for AD diagnosis from structural and functional perspectives. Multimodal joint diagnosis for sMRI and fMRI faces three key challenges. First, 3D sMRI and 4D fMRI differ significantly in spatial resolution, temporal dimension, and signal distribution, making it difficult to construct a unified end-to-end cross-modal modeling framework. Second, although Vision Transformer-based sMRI feature extraction offers global modeling capabilities, the standard multi-head attention suffers from redundant heads, insufficient inter-head collaboration, and limited representation of local structural details, reducing sensitivity to AD-relevant regions such as the hippocampus and olfactory cortex. Third, most multimodal fusion methods rely on feature concatenation, unidirectional attention, or dense interaction strategies, which are insufficient to screen key regions and establish fine-grained bidirectional associations between structural and functional features in high-dimensional heterogeneous image data. To address these issues, this paper proposes BiSparseFusion, a cross-modal bidirectional sparse interaction fusion model. The sMRI branch employs a 3D Vision Transformer enhanced with a dynamic composable multi-head attention mechanism (DCMHA) and a multi-level feature fusion module (MFFM). DCMHA reduces redundant attention outputs by dynamically combining attention heads, and MFFM aggregates multi-level features to enhance local lesion details and global semantic representation. The fMRI branch uses SwiFT to directly model spatio-temporal dependencies of the original 4D fMRI, avoiding information loss caused by conventional region of interest-or connectivity-based methods. During cross-modal fusion, a bidirectional sparse cross-attention fusion module (BSCAF) suppresses redundant features within modalities and enables deep complementary interaction between sMRI structural and fMRI functional representations. The proposed method is validated on the ADNI dataset across three classification tasks: AD/NC, MCI/NC, and AD/MCI. BiSparseFusion achieves classification accuracies of 97.67%, 93.27%, and 96.72%, respectively, surpassing various single- and multi-modal comparison models. Visualization results indicate that the model effectively focuses on brain regions associated with AD pathology, including the hippocampus, olfactory cortex, and amygdala, forming a more discriminative fusion feature space. These results demonstrate the effectiveness of BiSparseFusion in multimodal neuroimaging feature modeling, cross-modal fine-grained fusion, and AD auxiliary diagnosis.

Keywords: Alzheimer's disease; multimodal fusion; Structural magnetic resonance imaging; Functional magnetic resonance imaging; Dynamic Composable Multi-Head Attention; Bidirectional Sparse Cross-Attention Fusion

Foundation Item(s): National Natural Science Foundation of China (No.62371187); The Open Program of Hunan Intelligent Rehabilitation Robot and Auxiliary Equipment Engineering Technology Research Center (No.2024JS101)

## 0 引言

阿尔茨海默症(Alzheimer′s Disease,AD)是一种常见的神经退行性疾病[1],以进行性认知功能减退、记忆受损以及脑部结构异常为主要特征。2021年,全球患有痴呆(其中60–70%的病例为AD或伴随其他病理的混合型AD)的人数估计超过5700万人,预计到2050年患病人数将增加约三倍[2]。阿尔茨海默病(AD)的病理进程复杂且隐匿,临床诊断往往需要综合影像学、神经心理学评估及生物标志物等多源证据。随着成像技术的迅猛发展,神经影像已成为非侵入性揭示脑结构与功能的关键手段:结构磁共振(structural Magnetic Resonance Imaging,sMRI)能够精确刻画脑区形态与体积的萎缩演变[3],而功能磁共振(functional Magnetic Resonance Imaging,fMRI)则可捕捉脑功能网络的动态交互特征[4]。两种模态分别从结构和功能层面对脑部病变进行描述,在AD诊断中发挥重要作用。

尽管sMRI和fMRI在AD诊断研究中均取得了一定进展,但单一模态影像信息显然难以全面反映AD的多层次病理机制[5]。比如,sMRI仅提供静态的解剖信息,难以捕捉脑区间的动态交互;fMRI虽然能够反映神经活动的时序特征,但存在信噪比低、时间序列复杂等问题,且当前大多数关于fMRI研究工作仍主要集中于功能连接、区域一致性等间接建模方法[6],直接利用原始4D fMRI数据进行端到端特征学习的研究仍然不足[7]。

通过多模态融合,同时利用sMRI的结构信息与fMRI的功能性信息[8],实现对脑部异常的全局-局部、结构-功能一体化表征,是实现AD早期精准诊断的有效途径。然而,现有研究仍面临以下三方面挑战:第一,模态间特征分布差异显著,缺乏统一的跨模态端到端高维建模框架。3D sMRI与4D fMRI在空间分辨率、时间维度及信号特性上存在本质差异[9-10]。第二,经典的基于单模态脑部影像特征提取的方法Vision Transformer(ViT),其多头注意力存在冗余与低效性[11-12],注意力头之间缺乏动态协同,限制了对与AD

相关关键脑区依赖关系的精准捕获。此外，现有ViT在处理3D医学影像时普遍依赖深层输出的全局表征作为模态表示，降低了与AD相关的局部脑部结构变化（如嗅皮质和海马萎缩）的敏感性[13-14]。其三，现有的跨模态融合机制多为单向化且粒度粗[15]，难以实现结构－功能层面的深度协同[16]。多数方法仅支持从一个模态向另一模态的信息传递，缺乏双向、细粒度的特征交互建模[17]。

针对上述问题，本文提出一种名为BiSparseFusion的端到端AD诊断模型，该模型基于3D sMRI与4D fMRI进行跨模态双向稀疏交互融合，以高效整合脑结构与功能特征信息并发挥其互补性。模型采用双分支架构，在sMRI分支中，构建了集成动态可组合多头注意力机制（Dynamic Composable Multi-Head Attention，DCMHA）[18]与多层特征融合模块（Multi-level Feature Fusion Module，MFFM）的3D ViT。其中，DCMHA模块用于减少多头注意力中的冗余性并增强头间交互，从而提升结构特征建模的精细性与判别力；MFFM通过卷积聚焦细粒度特征并跨层聚合多尺度语义，强化局部病灶与全局语义的一致性表达。fMRI分支采用SwiFT[7]模型直接对原始4D高维体数据进行时空依赖建模，充分挖掘神经功能动态模式。此外，为实现跨模态深度协同，设计了双向稀疏交叉注意力融合（Bidirectional Sparse Cross-Attention Fusion，BSCAF）模块，在模态内进行稀疏关键特征筛选，在模态间建立双向细粒度交互。最终，融合后的多模态特征被输入分类器以实现AD、轻度认知障碍（MCI）与正常对照（NC）的精准判别。本文主要贡献可总结如下：

（1）构建了集成DCMHA机制与MFFM的3D ViT编码器，可高效建模sMRI体数据中的空间结构特征。该结构能够在保证全局建模能力的同时减少注意力计算的冗余，并通过多层特征交互机制增强不同层次特征之间的信息传递，协同整合局部病灶与全局语义特征，从而提高特征表达的判别性。

（2）设计了一种新颖的BSCAF机制，在模态内部通过稀疏注意力实现特征选择与冗余抑制，在模态之间通过双向交互机制实现结构与功能特征的协同建模，有效提升了多模态特征融合的相关性与稳健性，为后续分类任务提供了更具判别力的融合表征。

（3）基于上述组件，构建了一种端到端的3D sMRI－4D fMRI跨模态联合分析的框架，实现了脑结构与功能特征的同步深度建模。在ADNI数据集上取得了优于主流单模态与多模态基线模型的AD/MCI/NC的分类性能，为多模态神经影像智能诊断提供了一种可扩展的新型融合范式。

## 1 相关工作

### 1.1 基于sMRI的AD诊断方法

sMRI为AD的诊断提供了关键的脑结构变化信息，目前大量研究主要依赖于3D卷积神经网络（Convolutional Neural Network，CNN）进行端到端的特征学习，比如，Abbas提出一种新型3D雅可比域卷积神经网络（Jacobian Domain Convolutional Neural Network，JD-CNN）[19]；ResNet[20]利用残差连接有效缓解深层网络中的梯度消失问题，因此也被广泛扩展至3D结构，在AD诊断中常作为sMRI特征提取的主干网络。CNN模型在捕获局部脑区结构变化方面虽然表现出较高精度，但由于卷积操作本质上难以建模远距离依赖关系[21-22]，因此对AD病灶的全局特征表达能力也存在局限性。

为突破卷积模型的局部感受野限制，ViT被引入计算机视觉领域，并迅速扩展至3D医学影像分析[23]。例如，研究人员提出Brain Informer[24]模型，通过高效的自注意力机制与特征提取策略，从sMRI中挖掘判别性特征，在AD诊断任务中取得优异性能；Lu等人[25]提出了一种名为RanCom-ViT的新型计算机辅助诊断方法，旨在自动且精确地解读大脑sMRI切片，以实现AD诊断。

基于ViT的模型中基本都采用标准的多头注意力机制，但研究表明多头注意力机制中的各个注意力头往往存在冗余[11-12]，同时多头注意力机制也会引发低秩瓶颈[26]，降低模型的表达能力。此外，ViT模型的深层特征相较于强调组织边界与局部结构模式的浅层特征更加抽象、偏全局化，而与AD密切相关的部分脑区结构是需要模型能够准确识别的局部结构，如果缺乏合适的聚合机制，深度网络难以同时保持结构细节与高级语义的一致表征[27-28]，进而无法准确呈现从海马体与内嗅皮质等局部萎缩到皮层普遍变薄、脑室体积扩大等全脑结构改变的多尺度病理模式。

### 1.2 基于fMRI的AD诊断方法

fMRI可非侵入性地捕捉脑部神经活动的时空动态特征，早期的fMRI研究主要基于功能连接分析（Functional Connectivity，FC），通过计算脑区间信号的统计相关性构建功能连接矩阵。例如，研究者采用区域相关性（ROI-based correlation）分析AD患者的脑功能连接模式[29]。随着深度学习技术的发展，深度神经网络被广泛引入AD诊断任务。Duc等人[30]利用3D CNN直接从fMRI体数据中学习空间特征，实现了较高的分类精度；Li等人[31]提出C3d-LSTM模型进一步将3D CNN与长短期记忆网络（Long Short-Term Memory，LSTM）相结合，以同时捕获空间和时间依赖关系；Hu等人[32]基于图卷积网络（Graph Convolu-

tional Network，GCN）的新型成本敏感加权对比学习方法 CSWCL-GCNs 进行不平衡的 AD 分期。当 Transformer 架构近年来被引入 fMRI 分析任务后，许多研究人员开始将卷积神经网络用于提取空间特征，而将处理时间信息的结构替换为 Transformer 模型[33-35]。

上述研究总体可分为两大类。第一类通常通过对高维 fMRI 数据进行脑区划分或计算功能连接来获得特征表示，此类方法能够在一定程度上降低数据维度、提升计算效率，但其强烈依赖预定义的脑区模板，易丢失对刻画个体脑功能差异至关重要的细粒度信息[36]。第二类研究则直接利用原始 fMRI 体数据进行建模，但模型在空间域与时间域分别进行特征提取，导致时空信息解耦，难以实现全时空依赖关系的联合学习。

### 1.3 多模态特征融合诊断方法

多模态医学影像融合旨在综合不同成像模态所包含的互补信息，以获得更全面、更鲁棒的病理特征表示，对于疾病的精准诊断与表征具有重要意义。研究表明多模态信息联合建模能够提升 AD 诊断的性能。例如，Choudhury 等人[37]提出一种新的端到端耦合 GAN 模型，该模型将 MRI 和 PET 图像编码到共享的潜在空间中对 AD 的特定阶段进行分类；Hojjati 等人[38]结合 sMRI 与 fMRI 的互补结构－功能信息提高了 MCI 到 AD 转归预测的敏感性；Shukla 等人[39]基于 PET 与 MRI 的融合特征并采用集成分类器，同样获得了稳定的性能提升；Abdelaziz 等人[40]提出的多尺度多模态深度学习框架则利用多头注意力与交叉注意力增强模态间的特征交互能力；Zhou 等人[41]提出 HybridCA-Net 新诊断网络，整合多模态对比学习和跨模态注意力机制进行 AD 诊断。尽管上述多模态医学影像融合方法在 AD 诊断中展现出了显著潜力，但在高维异构模态的深度交互建模方面仍存在明显不足。大部分研究的特征融合机制过于简单或是基于单向注意力的浅层交互[16-17]，难以充分捕获不同影像数据之间复杂的耦合关系。此外，注意力机制往往呈现过度稠密的响应模式，缺乏对模态内关键区域的有效筛选，限制了模型在高维时空特征中的判别能力。

## 2 BiSparseFusion 模型

### 2.1 整体框架

本文提出了一种基于 sMRI 与 fMRI 融合的 AD 诊断框架，即 BiSparseFusion，其整体结构如图 1 所示。在输入层面，sMRI 使用 T1 加权（T1w）结构像，fMRI 使用 4D 原始高维数据；在特征提取阶段，sMRI 通道采用 3D Vision Transformer 作为基础编码器，并引入 DCMHA 机制和 MFFM，以增强对关键结构区域的建模能力；fMRI 通道则采用 SwiFT 模型，直接作用于 4D 高维数据输入以学习全脑功能网络的时空表征。最后两种模态的特征通过设计的稀疏交叉注意力融合模块进行交互，融合后的特征最终输入至全连接分类器，得到 AD 早期诊断结果。

### 2.2 sMRI 特征提取

为了充分捕获 3D sMRI 影像中蕴含的多层次空间结构与全局语义特征，本研究引入 DCMHA 和 MFFM，设计了一种基于 3D ViT 的 sMRI 特征提取模型。

输入的 3D sMRI 数据 $\boldsymbol{X} \in \mathbb{R}^{1\times H\times W\times D}$ 通过一个 3D

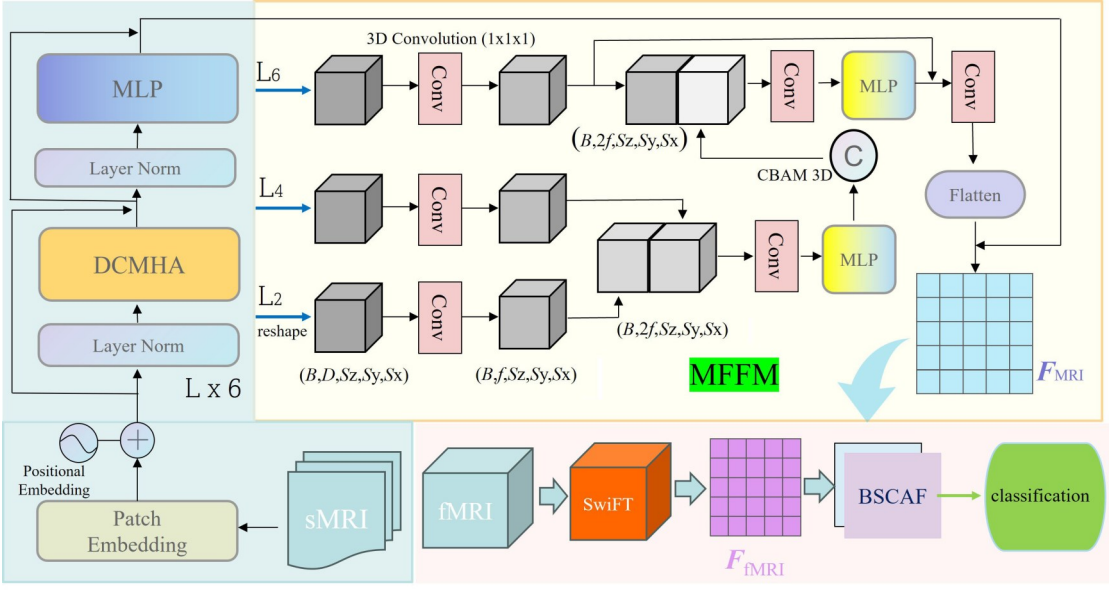


图 1 BiSparseFusion:基于 sMRI 与 fMRI 的 AD 融合诊断框架图

Figure 1 Architecture of BiSparseFusion based on sMRI and fMRI for AD diagnosis

卷积嵌入层划分为体积patch并映射到高维特征空间，得到初始的token序列，再加入可学习的位置编码后输入堆叠的Transformer编码块进行特征提取，每个编码块由层归一化、DCMHA注意力子层和多层感知机（MLP）子层组成，并包含残差连接，其计算形式为

$$\boldsymbol{Z}^{(l)'}=\boldsymbol{Z}^{(l-1)}+\mathrm{DCMHA}\left(\mathrm{LN}\left(\boldsymbol{Z}^{(l-1)}\right)\right) \tag{1}$$

$$\boldsymbol{Z}^{(l)}=\boldsymbol{Z}^{(l)'}+\mathrm{MLP}\left(\mathrm{LN}\left(\boldsymbol{Z}^{(l)'}\right)\right) \tag{2}$$

其中，$\boldsymbol{Z}^{(l)}$表示第$l$层的输出特征，$\boldsymbol{Z}^{(l)'}$表示第$l$层中生成的中间特征，$\mathrm{LN}(\cdot)$为层归一化操作，$\mathrm{MLP}(\cdot)$为多层感知机模块。

其中，DCMHA机制具体示意图如图2所示。DCMHA的核心在于引入一个输入依赖的动态组合机制，该机制能够在注意力矩阵层面实现不同注意力头之间的灵活协作，从而动态地组合出更具表现力的新注意力头。

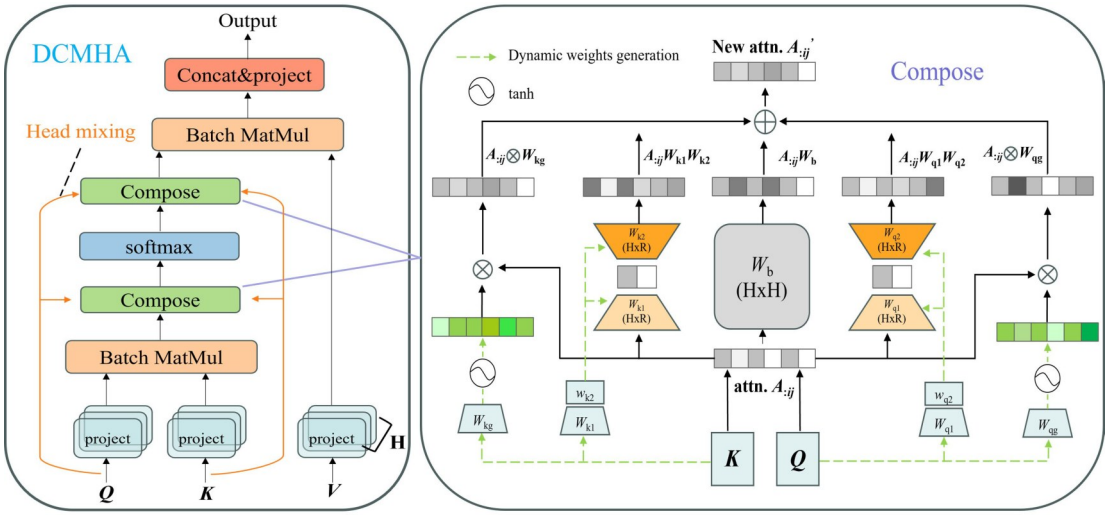


图2 DCMHA机制详解图

Figure 2 Detailed Diagram of the DCMHA Mechanism

给定输入序列，首先通过线性投影得到查询（$\boldsymbol{Q}$）、键（$\boldsymbol{K}$）和值（$\boldsymbol{V}$）矩阵并计算其第$i$个头的注意力得分矩阵$\boldsymbol{A}_i^S$，然后DCMHA通过一个Compose()函数在注意力中Softmax计算的前后阶段对注意力得分矩阵中堆叠的注意力张量进行变换。Compose()函数的核心运算是针对每一对查询向量$\boldsymbol{Q}_i$和键向量$\boldsymbol{K}_j$所对应的注意力向量$\boldsymbol{A}_{:ij}\in\mathbb{R}^H$，H表示多头数目，将其动态地映射为一个新的注意力向量$\boldsymbol{A}'_{:ij}$。此变换通过五个并行分支的加权和实现：

$$\begin{aligned}\boldsymbol{A}'_{:ij}=&\boldsymbol{A}_{:ij}\boldsymbol{W}_b+\boldsymbol{A}_{:ij}\boldsymbol{W}_{q1}\boldsymbol{W}_{q2}+\boldsymbol{A}_{:ij}\otimes\boldsymbol{W}_{qg}\\&+\boldsymbol{A}_{:ij}\boldsymbol{W}_{k1}\boldsymbol{W}_{k2}+\boldsymbol{A}_{:ij}\otimes\boldsymbol{W}_{kg}\end{aligned} \tag{3}$$

其中，$\boldsymbol{W}_{\mathrm{b}}\in\mathbb{R}^{H\times H}$是一个与输入无关的静态基础投影矩阵，作为组合的偏置项，第二项与第四项构成了低秩动态投影分支，$\boldsymbol{W}_{\mathrm{q1}}\in\mathbb{R}^{H\times R}$与$\boldsymbol{W}_{\mathrm{q2}}\in\mathbb{R}^{R\times H}$是由查询$\boldsymbol{Q}_i$生成的权重矩阵，$\boldsymbol{W}_{\mathrm{k1}}$与$\boldsymbol{W}_{\mathrm{k2}}$则是由键$\boldsymbol{K}_j$对称生成的权重矩阵，这两个分支通过低秩（$R\ll H$）投影建模头部间复杂的权重共享模式，第三项与第五项构成了动态门控分支，$\boldsymbol{W}_{\mathrm{qg}}\in\mathbb{R}^H$与$\boldsymbol{W}_{\mathrm{kg}}\in\mathbb{R}^H$分别由查询与键生成。

最终DCMHA模块的输出按如下顺序计算：

$$\boldsymbol{A}^S=\mathrm{Stack}\left(\left\{\boldsymbol{A}_i^S\right\}_{i=1}^H\right)$$

$$\boldsymbol{A}^S\leftarrow\mathrm{Compose}\left(\boldsymbol{A}^S,\boldsymbol{Q},\boldsymbol{K};\ \theta_{\mathrm{pre}}\right)$$

$$\boldsymbol{A}^W=\mathrm{Softmax}\left(\boldsymbol{A}^S,\mathrm{dim}=-1\right)$$

$$\boldsymbol{A}^W\leftarrow\mathrm{Compose}\left(\boldsymbol{A}^W,\boldsymbol{Q},\boldsymbol{K};\ \theta_{\mathrm{post}}\right)$$

$$\boldsymbol{O}_i=\boldsymbol{A}_i^W\boldsymbol{V}\boldsymbol{W}_i^V$$

$$\boldsymbol{O}=\mathrm{Concat}\left(\boldsymbol{O}_1,\boldsymbol{O}_2,\ldots,\boldsymbol{O}_H\right)\boldsymbol{W}^O \tag{4}$$

其中，Compose()函数的可训练参数集合$\theta$（例如用于前置组合的$\theta_{\mathrm{pre}}$）定义为$\theta=\left\{\boldsymbol{W}_{\mathrm{q1}},\boldsymbol{W}_{\mathrm{q2}},\boldsymbol{W}_{\mathrm{qg}},\boldsymbol{W}_{\mathrm{k1}},\boldsymbol{W}_{\mathrm{k2}},\boldsymbol{W}_{\mathrm{kg}},\boldsymbol{W}_{\mathrm{b}}\right\}$，$\boldsymbol{W}^O$为输出投影矩阵，$\boldsymbol{O}$为每一层Transformer块整合多头关系后输出的特征表示。

每层的输出重新映射回三维特征图后经过1×1×1卷积降维得到待融合特征$\boldsymbol{M}^{(l_i)}$；随后，多层特征融合模块先融合中低层特征，通过3D卷积注意力模块CBAM3D对通道与空间维度进行加权得到注意力增强后的特征：

$$\boldsymbol{F}_1=\mathrm{CBAM3D}\left(\mathrm{MLP}\left(\mathrm{Conv3D}\left(\left[\boldsymbol{M}^{(l_1)},\boldsymbol{M}^{(l_2)}\right]\right)\right)\right) \tag{5}$$

其中，$[\cdot,\cdot]$表示特征拼接，MLP层中包含批归一化以及ReLU激活函数，然后将中层特征$\boldsymbol{F}_1$与深层特征$\boldsymbol{M}^{(l_3)}$融合并通过1×1×1卷积映射回Transformer嵌入维度：

$$\boldsymbol{F}_{\mathrm{MRI}}=\mathrm{Conv3D}_{1\times1\times1}\left(\mathrm{MLP}\left(\mathrm{Conv3D}\left(\left[\mathrm{F}_1,\boldsymbol{M}^{(l_3)}\right]\right)\right)\right)+\boldsymbol{Z}^{(l_3)} \tag{6}$$

再通过元素级相加与原始深层特征 $\boldsymbol{Z}^{(l_3)}$ 实现残差连接，用以保留深层语义的稳定性，最终将得到的融合特征记为 $\boldsymbol{F}_{\mathrm{MRI}}$。

### 2.3 4D fMRI特征提取

为从4D fMRI数据中有效提取时空特征，本文采用SwiFT[7]模型作为特征提取器。SwiFT是一种基于Swin Transformer的先进架构，其核心创新在于扩展了4D局部窗口多头自注意力机制，能够以端到端的方式直接处理原始的fMRI数据，避免了传统基于感兴趣区域(ROI)或分步式方法可能造成的信息丢失。在输入层，fMRI数据形式化表示为 $\boldsymbol{X}_{\mathrm{fMRI}}\in\mathbb{R}^{B\times C\times D\times H\times W\times T}$，其中 $B$ 为批次大小，$C$ 为通道数，$H,W,D$ 为空间维度，模型输入时的空间维度设置为[96,96,96]，$T$ 为时间帧数，输入模型时 $T$ 被固定为20。为将原始时间帧数大于20的fMRI数据适配模型输入，研究中采用步长为20的滑动窗口对时间序列进行分割：

$$S=\left\{X_{t:t+20}\middle|t=0,20,40,\cdots,T-20\right\} \tag{7}$$

从而将每个被试的扫描转化为[$T$/20]个子序列，保证模型能覆盖全程的神经动态。

### 2.4 双向稀疏交叉注意力融合模块

为实现sMRI结构特征与fMRI动态功能的深度协同，本研究设计了双向稀疏交叉注意力融合模块(BSCAF)。该模块通过稀疏机制抑制模态内冗余，并利用双向交互强化互补信息，图3详细描述了BSCAF的具体结构。

具体而言，首先将fMRI特征 $\boldsymbol{F}_{\mathrm{fMRI}}\in\mathbb{R}^{B\times N_f\times d_f}$ 和sMRI特征 $\boldsymbol{F}_{\mathrm{MRI}}\in\mathbb{R}^{B\times N_m\times d_m}$ 分别通过可学习的投影矩阵进行线性投影将其映射到相同的融合空间维度，然后在模态内表征阶段利用稀疏注意力机制通过自适应地筛选关键区域，突出病变敏感区域的表征，从而提升sMRI特征和fMRI特征的判别性。然后，将标准的稠密注意力分布与通过ReLU²激活获得的稀疏分布通过可学习的稀疏性权重 $\alpha$ 进行自适应加权融合：

$$A=\alpha\cdot\mathrm{Softmax}\left(\frac{\boldsymbol{QK}^{\top}}{\sqrt{d}}\right)+(1-\alpha)\cdot ReLU\left(\frac{\boldsymbol{QK}^{\top}}{\sqrt{d}}\right)^2 \tag{8}$$

在模态间交互阶段，使用双向交叉注意力机制以实现结构与功能信息的深度融合。fMRI的动态功能特征从sMRI解剖表征中获取补充信息，同时sMRI的结构特征也可受到fMRI功能活动模式的调制。计算公式为

$$\hat{\boldsymbol{Z}}^{\mathrm{fMRI}}=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}^{\mathrm{fMRI}}(\boldsymbol{K}^{\mathrm{MRI}})^{\top}}{\sqrt{d}}\right)\boldsymbol{V}^{\mathrm{MRI}} \tag{9}$$

$$\hat{\boldsymbol{Z}}^{\mathrm{MRI}}=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}^{\mathrm{MRI}}(\boldsymbol{K}^{\mathrm{fMRI}})^{\top}}{\sqrt{d}}\right)\boldsymbol{V}^{\mathrm{fMRI}} \tag{10}$$

最后，模型分别对 $\hat{\boldsymbol{Z}}^{\mathrm{fMRI}}$ 和 $\hat{\boldsymbol{Z}}^{\mathrm{MRI}}$ 进行池化，并将两者拼接后得到双模态融合特征h，最后将融合特征输入分类层后计算分类概率，得到AD诊断结果。

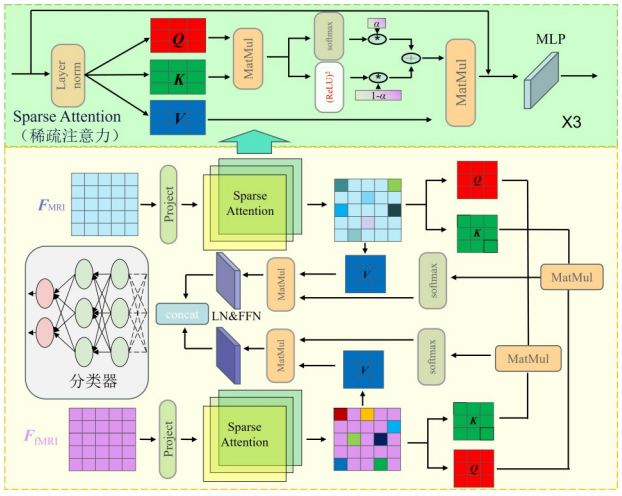


图3 BSCAF详细结构图

Figure 3 Detailed structure diagram of BSCAF

## 3 实验验证与结果分析

### 3.1 数据集和预处理

研究中使用的数据来自ADNI数据集，fMRI数据统一采用静息态功能性磁共振成像(resting-state functional Magnetic Resonance Imaging，rs-fMRI)，研究中选取了89名AD患者、142名MCI患者和220名正常对照(NC)作为实验样本。对于sMRI和fMRI两种模态的数据，研究中统一采用DPABI工具进行数据预处理，其中sMRI数据处理流程为：去除头皮结构，将T1结构像配准到功能像空间，分割结构像得到灰质图像，标准化到MNI空间，3D体积大小重采样为128×128×128。fMRI数据处理流程为：去除数据前10帧，时间片校正，头动校正，头动参数回归，空间标准化和空间平滑，将空间维度体积重采样为96×96×96。

### 3.2 实验设置和评价指标

在本研究中，所有实验均在基于PyTorch2.0.0+CUDA 11.8的深度学习框架下实现，GPU选择NVIDIA A800GPU。模型训练采用AdamW优化器，以$5\times10^{-6}$作为初始学习率，训练批次大小(batch size)设置为32，并使用二元交叉熵(Binary Cross-Entropy，BCE)作为损失函数。MRI分支的Vit模型的基本参数如下：patch大小为16，嵌入维度为384，模型深度为6层，多头数为6个。在数据划分方面，按照被试者编号随机划分训练集与测试集，比例为8:2，确保数据

间不存在数据泄漏。模型性能评估采用多项指标进行综合分析，包括分类准确率（Classification Accuracy，ACC）、受试者工作特征曲线下面积（Area Under the Receiver Operating Characteristic Curve，AUC）、特异度（Specificity，SPE）以及敏感度（Sensitivity，SEN）。

### 3.3 对比实验

如表1所示，为验证所提出模型在AD诊断任务中的有效性，本文选取多种具有代表性的模型进行对比实验。所选模型涵盖了3D卷积神经网络、Transformer结构、图神经网络及多模态融合方法等不同类型，既包括经典的单模态特征提取网络，也包含近年来在多模态脑影像分析中表现突出的先进模型，包括常用的基线图模型BrainGNN[42]；经典3D卷积神经网络3D ResNet[20]；基于Swin Transformer的3D医学图像分割模型Swin UNETR[43]；基于正交潜空间学习的多模态融合AD诊断模型OLFG[44]；利用视觉变换器和交叉注意力机制，有效融合sMRI与功能性网络连接矩阵的模型MultiVit[45]；在大规模的3D医学图像数据集上训练的模型SAM-Med3D[46]；构建包含病灶脑区与风险基因的交互图实现AD诊断的模型BF[47]；基于sMRI和fMRI双模态数据，采用多模态对比学习机制并结合跨模态注意力的融合诊断模型HybridCA-Net[41]；基于多重特征融合与疾病诱导学习的多模态AD诊断模型MDL-Net[48]和基于去相关约束与多模态特征交互的AD诊断模型D-MAFF[49]。

**表1 提出的AD诊断框架与其他方法的对比结果**

Table 1 Diagnosis results of the proposed AD diagnostic framework with comparative methods

| 方法 | AD/NC | | | | MCI/NC | | | | AD/MCI | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ACC | AUC | SPE | SEN | ACC | AUC | SPE | SEN | ACC | AUC | SPE | SEN |
| BrainGNN | 67.44*** | 73.56*** | 72.55 | 60.00 | 60.58*** | 63.91*** | 68.97 | 50.00 | 65.57*** | 76.51*** | 78.12 | 51.72 |
| 3D ResNet | 82.56*** | 86.78** | 82.35 | 82.86 | 74.04*** | 83.51** | 67.24 | 82.61 | 72.13*** | 79.53** | 81.25 | 62.07 |
| Swin UNETR | 84.88** | 88.68** | 82.35 | 88.57 | 66.35*** | 68.89*** | 84.48 | 43.48 | 78.69** | 82.44** | 62.50 | **96.55** |
| OLFG | 88.03 | 85.58 | 90.72 | 85.58 | 67.04 | 70.93 | 70.90 | 58.11 | 71.22 | 77.12 | 76.89 | 79.07 |
| Multivit | 88.37* | 90.81* | 86.27 | 91.43 | 85.58 | 87.63* | 94.83 | 73.91 | 75.41*** | 74.03*** | **96.88** | 51.72 |
| SAM-Med3D | 88.37* | 93.84* | 90.20 | 88.89 | 87.50 | 90.33* | 87.93 | **86.96** | 83.61* | 93.10 | 84.37 | 82.76 |
| BF | 91.87 | 94.14 | 92.86 | 89.74 | 88.18 | 91.55 | 90.71 | 86.15 | 87.09 | 89.61 | 87.95 | 86.15 |
| HybridCA-Net | 94.19 | 95.49 | 96.08 | 91.43 | 87.50 | 87.97* | 84.48 | 91.30 | 90.16 | 95.91 | 90.62 | 89.66 |
| MDL-Net | 96.37 | 98.48 | 95.38 | 97.40 | 73.61 | 76.08 | 73.02 | 73.01 | 85.29 | 88.16 | 82.00 | 93.09 |
| D-MAFF | 96.77 | 92.69 | **97.63** | 93.15 | 90.32 | 91.59 | 87.63 | 88.15 | 88.23 | 92.75 | 91.86 | 93.54 |
| BiSparseFusion | **97.67** | **99.16** | 97.14 | **98.04** | **93.27** | **97.41** | **96.55** | 89.13 | **96.72** | **98.28** | **96.88** | **96.55** |

注：*表示$p$值小于0.05；**表示$p$值小于0.01；***表示$p$值小于0.001；加粗字体表示指标最优值。

从对比结果可以看出，本文提出的双模态融合框架在AD/NC、MCI/NC、AD/MCI三个分类任务上均取得了最优性能，整体表现明显优于现有的单模态和多模态先进方法。与传统3D CNN（如3D ResNet）和基于Transformer深度模型（如Swin UNETR、SAM-Med3D）的单模态方法相比，多模态方法在高维影像数据的建模能力更强，刻画AD脑部病症信息更全面，诊断结果更加精准。在多模态融合模型方面，OLFG、MDL-Net与D-MAFF等方法通过潜空间约束、跨模态注意力或多尺度融合等机制虽取得一定提升，但其融合结构大多仍依赖单向交互或稠密注意力模式，对模态内冗余特征抑制不足，且对结构-功能间细粒度对应关系的建模仍不够充分。为全面评估模型的复杂性，在相同条件下对比了各方法的运算复杂度（见表2）。

如表2结果显示，基于图网络或特征降维的方法虽然显存与延迟较低，但代价是预先将高维数据压缩为节点序列，其中，BrainGNN、OLFG在计算开销上展现出极低的数值，这主要归因于这些方法在非端到端的预处理阶段将高维影像转化为低维矩阵；而直接处理高维医学影像的3D网络（如3D ResNet、Swin UNETR）虽保留了丰富空间信息，却面临极高的计算挑战。相比之下，本文提出的BiSparseFusion在性能与效率间取得了一定的平衡，在维持双分支端到端高分辨率时空建模的前提下，其参数量为$2.22\times10^7$，FLOPs显著降低至$3.29\times10^{10}$，单样本推理时间仅22.90 ms。为了进一步明确本模型的贡献，本文与表现优异的HybridCA-Net进行了底层机制的深入对比。首先，在特征表征上，HybridCA-Net的fMRI分支依赖预定义的脑区图谱提取节点特征，而本模型直接对原始4D fMRI进行端到端的建模。其次，HybridCA-Net采用单向交叉注意力进行跨模态融合，而BSCAF创新性地引入了稀疏性约束的同时实现了结构与功能的双向交互。尽管本模型因保留高维时空特征导致

表2 不同方法运算复杂度比较

Table 2 Comparison of computational complexity of different methods

| 方法 | 参数量 | FLOPs | 单样本显存占用/MB | 单样本推理时间/ms |
|---|---|---|---|---|
| BrainGNN | $6.70\times10^4$ | $4.60\times10^6$ | 14.06 | 5.14 |
| 3D ResNet | $3.32\times10^7$ | $2.52\times10^{11}$ | 1 178.27 | 16.32 |
| Swin UNETR | $1.67\times10^7$ | $5.92\times10^{10}$ | 3 823.74 | 74.11 |
| OLFG | $2.17\times10^5$ | | 12.14 | 0.91 |
| MultiVit | $2.40\times10^7$ | $6.59\times10^9$ | 123.78 | 3.55 |
| SAM-Med3D | $9.26\times10^7$ | $8.95\times10^{10}$ | 1 477.21 | 54.20 |
| BF | $6.66\times10^7$ | $3.00\times10^8$ | 264.73 | 1.67 |
| HybridCA-Net | $3.42\times10^7$ | $6.46\times10^{10}$ | 333.52 | 4.26 |
| MDL-Net | $2.83\times10^6$ | $2.74\times10^{10}$ | 197.44 | 7.92 |
| D-MAFF | $4.11\times10^6$ | $2.80\times10^8$ | 31.25 | 2.77 |
| BiSparseFusion | $2..22\times10^7$ | $3.29\times10^{10}$ | 843.19 | 22.90 |

推理速度略慢于依赖图谱降维的方法(如HybridCA-Net),但在参数量控制上优势明显。得益于DCMHA机制对多头冗余信息的有效抑制,本模型减少了多头数量,从而大幅降低了参数量。相比之下,以标准3D ResNet作为骨干提取sMRI特征的HybridCA-Net,其总参数量反而更高。综上所述,本文所提出的模型在三个任务上的分类性能表现突出:ACC分别达到97.67%、93.27%、96.72%,AUC分别达到99.16%、97.41%、98.28%,均领先于现有经典方法和最新的研究方法,同时在复杂度和计算效率上也取得了一定的平衡。为验证模型性能的提升具有严格的统计学意义,本文还使用McNemar和deLong分别对ACC和AUC进行了统计显著性检验,结果中***表示$p$值小于0.001,**表示$p$值小于0.01,*表示$p$值小于0.05。鉴于数据模态设置存在的差异性,未与OLFG、BF、MDL-Net、D-MAFF进行统计显著性检验,而其余可比方法的对照实验中,BiSparseFusion都能达到统计显著性水准,表明其性能提升不仅体现在点估计指标上,更具有稳健且可重复的统计证据支撑。

### 3.4 消融实验

为系统评估本文提出的DCMHA、MFFM以及BSCAF三个核心模块在整体框架中的有效性,本文进一步开展了消融实验,实验结果如表3所示。

表3 消融实验结果

Table 3 The results of the ablation experiment

| DCMHA | MFFM | BSCAF | AD/NC | | | | MCI/NC | | | | AD/MCI | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | ACC | AUC | SPE | SEN | ACC | AUC | SPE | SEN | ACC | AUC | SPE | SEN |
| | | | 81.40*** | 83.59** | 94.12 | 62.86 | 79.81** | 84.11** | 91.38 | 65.22 | 81.97* | 82.97** | **96.88** | 65.52 |
| √ | | | 83.72*** | 86.27** | 82.35 | 85.71 | 81.73* | 85.91** | 87.93 | 73.91 | 83.61* | 85.13** | 93.75 | 72.41 |
| | √ | | 82.56*** | 87.17* | 82.35 | 82.86 | 80.77* | 86.24* | 77.59 | 84.78 | 85.25* | 86.64* | **96.88** | 72.41 |
| | | √ | 84.88** | 93.56* | 92.16 | 74.29 | 81.73* | 87.56* | 68.97 | **97.83** | 83.61* | 85.99* | 75.00 | 93.10 |
| √ | √ | | 95.35 | 96.64 | 94.12 | 97.14 | 91.35 | 96.85 | **96.55** | 84.78 | 90.16 | 94.29 | 87.50 | 93.10 |
| √ | | √ | 89.53* | 93.05* | 84.31 | 97.14 | 88.46 | 91.45 | 87.93 | 89.13 | 88.52 | 94.07 | 90.62 | 86.21 |
| | √ | √ | 94.19 | 96.92 | 96.08 | 91.43 | 89.32 | 95.08 | 93.33 | 86.21 | 93.44 | 96.66 | 93.75 | 93.10 |
| √ | √ | √ | **97.67** | **99.16** | **97.14** | **98.04** | **93.27** | **97.41** | **96.55** | 89.13 | **96.72** | **98.28** | **96.88** | **96.55** |

注:*表示$p$值小于0.05;**表示$p$值小于0.01;***表示$p$值小于0.001;加粗字体表示指标最优值;√表示消融实验的模型中使用的模块。

为了提供明确的参考基准,首先构建了基线模型。基线模型移除了上述三大模块,仅使用标准3D ViT和SwiFT模型分别提取sMRI与fMRI特征,并在末端通过特征拼接后进行多模态分类。

从结果可见,基线模型在缺乏深层特征交互与稀疏约束的情况下,分类表现相对有限,难以充分挖掘高维多模态数据中的判别性信息。相比之下,三个核心模块单独引入时,均能在基线上取得一定提升。在双模块组合实验中,DCMHA和MFFM的组合在AD/NC任务上获得显著提升(ACC达到96.51%),说明注意力冗余抑制与多尺度结构聚合的协同作用能够更加敏感地捕获典型AD萎缩模式,DCMHA与BSCAF和MFFM与BSCAF的组合均在MCI/NC与AD/MCI任务中也表现出更高的增益,这一趋势表明跨模态双向

注意力对于AD病症中较弱、分布更分散的结构-功能异常具有更强的辨识能力。此外,MFFM和BSCAF相结合后表现整体优于DCMHA和BSCAF的组合,反映出层级结构表达的稳定强化更适合作为有效的功能调制基底,使BSCAF能在一致的结构语义空间中执行跨模态对齐与关键特征强化。在整合全部三个模块后,模型在三类任务上均达到最优性能,表明三个核心模块之间存在明确的互补性,DCMHA、MFFM与BSCAF三个模块的组合也展现出了协同效应。这种非线性的性能激增源于各模块在特征处理链路上的深度互补,DCMHA有效抑制了sMRI的空间冗余,提取出纯净的解剖结构特征,MFFM对其进行多尺度聚合,确保了局部微小病灶与全局语义的一致性。这一系列前置处理为最终的跨模态交互奠定了极高的特征质量基础。而当三大模块联合时,纯净且多尺度的结构表征极大提升了BSCAF的双向融合精度,三者层层递进、相互激发,从而最大化了模型对多模态AD高维影像的判别潜力。

### 3.5 模块可视化

为验证本方法在特征提取中的有效性,在预处理后的sMRI图像上通过热力图可视化技术,展示了本方法学习到的特征可视化结果,结果如图4所示。

其中在切片位置选择上,统一选择了三个方向上的中间切片。颜色深浅直观呈现了脑部区域的重要性,横截面的可视化结果显示模型在sMRI上重点关注脑区下方区域,这个区域主要是与短期记忆相关的颞叶所在位置,其中包含嗅皮质、海马体、杏仁核等与AD有重要关联的脑区部位;冠状模型也重点关注脑部下方位置,这对应横断面中颞叶位置,而更聚焦中部也是由于与AD相关的关键结构位于脑部下方中心位置;矢状面模型关注分布覆盖内侧颞叶及周边高相关区域,表明模型在3D结构空间中均能稳定定位信息量丰富且具有生物学意义的部位。

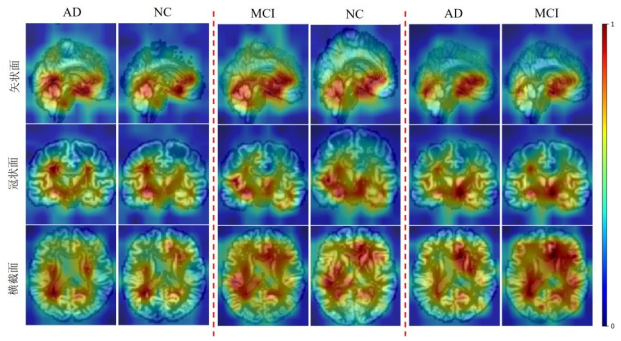


图4 在三个分类任务中热力图可视化结果

Figure 4 The visualization results of the heat maps in the three classification tasks

为深入理解多模态融合模型在AD诊断任务中的特征判别能力,本研究采用t-SNE降维技术在测试集上对模型不同层次的特征表示进行可视化分析,可视化结果如图5所示。

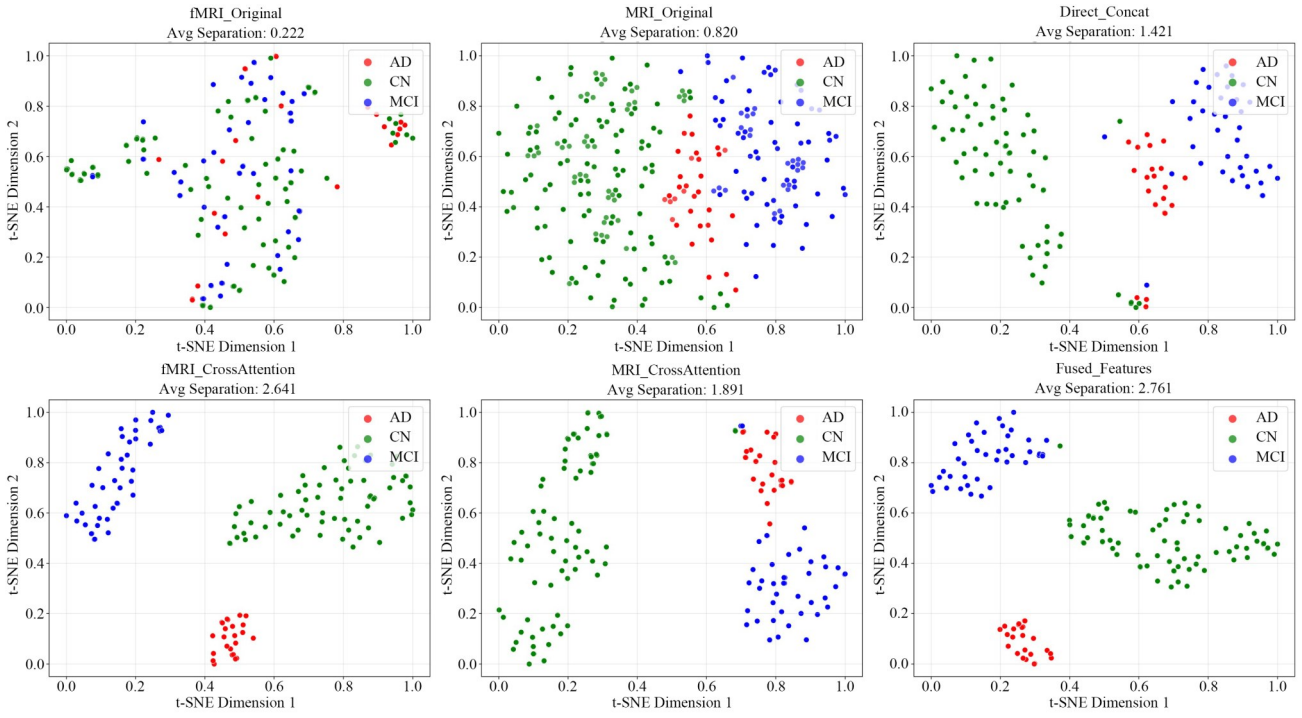


图5 t-SNE可视化结果

Figure 5 The visualization result of t-SNE

本研究系统性地提取了fMRI分支提取的原始特征（fMRI_Original），sMRI分支提取的原始特征（sMRI_Original），由原始特征拼接的融合特征（Direct_Concat），经过稀疏交叉注意力处理的fMRI特征（fMRI_CrossAttention），经过稀疏交叉注意力处理的sMRI特征（sMRI_CrossAttention）以及经过稀疏交叉注意力处理的特征经过BSCAF融合后的特征（Fused_Features）这六个关键特征进行对比分析，t-SNE可视化揭示了不同特征空间中AD、NC、MCI三类样本的分布特性，除可视化观察外，引入分离度得分（Separation Score）作为量化指标来衡量类别间相对距离，具体定义为降维后2D空间中平均类间距离与平均类内离散度的比值，计算公式如下：

$$S=\frac{\dfrac{2}{C(C-1)}\displaystyle\sum_{i=1}^{C-1}\sum_{j=i+1}^{C}\left\|\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\right\|_2}{\dfrac{1}{C}\displaystyle\sum_{i=1}^{C}\frac{1}{n_i}\sum_{\boldsymbol{x}\in\boldsymbol{X}_i}\left\|\boldsymbol{x}-\boldsymbol{\mu}_i\right\|_2} \tag{11}$$

其中，$C$为类别总数，$\boldsymbol{\mu}_i$和$n_i$分别表示第$i$类样本的中心坐标与样本数量。$S$值越大，表明不同类别间的界限越清晰且类内分布越紧凑。可视化结果表明原始的fMRI和sMRI特征并不能明显分离三个类别，其中fMRI特征存在明显重叠区域，直接进行特征拼接优于单模态但逊于注意力融合，说明简单的特征组合不足以充分使模型学习到足够的类别判别性，经过稀疏交叉注意力后的单模态特征优于原始的单模态特征，多模态融合特征则展现出最佳的分离效果，验证了融合策略的有效性。

### 3.6 BrainNet Viewer可视化分析

为了进一步理解本文所提出的模型在AD诊断中的脑区贡献，本文利用BrainNet Viewer进行了可视化分析。具体而言，在BSCAF融合过程中计算了各脑区的平均特征强度分数，归一化后映射到AAL模板定义的标准脑区上，随后从全部脑区中选取平均特征强度排名前20的区域，并利用BrainNet Viewer生成空间分布可视化网络图，以展现模型重点关注脑区的拓扑结构及其在3D空间中的位置分布（见图6），图中每个脑区名称后标明了平均特征强度分数。从结果中可以看出，模型在多个与AD早期病理密切相关的脑区上呈现较高的特征权重，其中包括嗅皮质、杏仁核、海马以及海马旁回等典型记忆与情绪调控相关结构，此外，脑岛、颞横回、颞极和梭状回等区域也表现出较高贡献，进一步体现了模型对认知、感知及情绪加工相关脑区的敏感性。整体而言，该可视化结果表明模型能够在多模态信息融合过程中同时捕捉与记忆功能、情绪调控以及高级认知加工相关的关键区域，为模型决策提供更具神经生物学依据的解释。

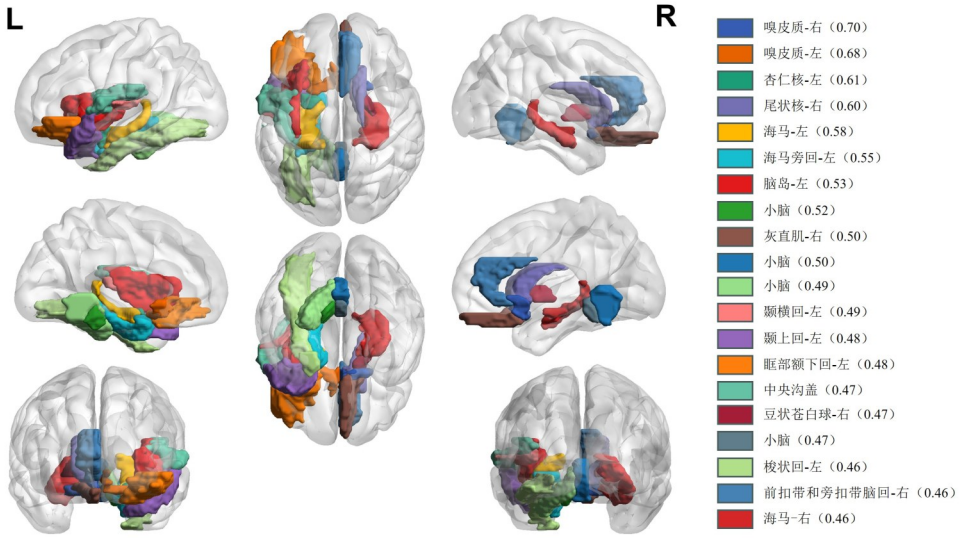


图6 BrainNet Viewer脑区映射结果

Figure 6 BrainNet Viewer brain region mapping results

## 4 结论

本文提出了一种面向AD早期诊断的3D sMRI与4D fMRI多模态融合框架BiSparseFusion。通过在sMRI分支引入DCMHA与MFFM增强结构特征表达，在fMRI分支利用SwiFT捕获全脑时空动态信息，并通过BSCAF实现模态内关键区域筛选与模态间深度互补融合。基于ADNI数据集的实验表明，所提方法在AD/NC、MCI/NC与AD/MCI任务中均取得显著优于现有方法的性能，验证了三大模块在结构建模与跨模态对齐中的协同作用。可视化分析进一步显示模型能

够聚焦病灶相关脑区并生成区分度更高的特征空间，证明了双模态联合建模的有效性。尽管BiSparseFusion在多模态特征协同建模与AD诊断性能上取得了显著成效，但仍有若干方面值得进一步探索。第一，面向临床部署的模型轻量化研究。通过结构化剪枝剔除稀疏注意力中持续处于低激活状态的冗余计算通路，在保证高精度的前提下实现模型压缩。第二，基于跨模态影像重建技术的模态缺失处理。在真实临床环境中，常面临fMRI或sMRI影像缺失的困境，未来工作可引入先进的跨模态影像合成技术。在此过程中，可充分利用BSCAF模块已学习到的“结构-功能双向稀疏映射先验”作为生成条件，通过已有单模态影像重建缺失模态的伪影像。最后，从静态分类迈向疾病进展的连续预测。现有的分类范式仅能刻画离散的病理阶段，未来可将现有的双模态架构拓展为纵向追踪模型，利用脑结构与功能的联合时空表征来预测患者的疾病发展，例如MCI转化为AD的概率。总体而言，本文为多模态神经影像融合提供了一种高效且具有临床潜力的解决方案，为未来在更大规模、多中心数据上的推广奠定了基础。

**参考文献**


[1] Zheng Qiuyang, Wang Xin. Alzheimer’s disease: Insights into pathology, molecular mechanisms, and therapy[J]. Protein & Cell, 2025, 16(2): 83-120.

[2] Frisoni G B, Hansson O, Nichols E, et al. New landscape of the diagnosis of Alzheimer’s disease[J]. The Lancet, 2025, 406(10510): 1389-1407.

[3] Sebenius I, Dorfschmidt L, Seidlitz J, et al. Structural MRI of brain similarity networks[J]. Nature Reviews Neuroscience, 2025, 26(1): 42-59.

[4] Arpanahi S K, Hamidpour S, Jahromi K G. Mapping Alzheimer´s disease stages toward it’s progression: A comprehensive cross-sectional and longitudinal study using resting-state fMRI and graph theory[J]. Ageing Research Reviews, 2025, 103: 102590.

[5] Ning Zhenyuan, Xiao Qing, Feng Qianjin, et al. Relation-induced multi-modal shared representation learning for Alzheimer’s disease diagnosis[J]. IEEE Transactions on Medical Imaging, 2021, 40(6): 1632-1645.

[6] Kan Xuan, Dai Wei, Cui Hejie, et al. Brain network transformer[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2022: 1855.

[7] Kim P Y, Kwon J, Joo S, et al. SwiFT: Swin 4D FMRI transformer[C]//Proceedings of the 37th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2023: 1820.

[8] Qiu Shangran, Miller M I, Joshi P S, et al. Multimodal deep learning for Alzheimer’s disease dementia assessment [J]. Nature Communications, 2022, 13(1): 3404.

[9] Bhattacharya S, Prusty S, Pande S P, et al. Integration of multimodal imaging data with machine learning for improved diagnosis and prognosis in neuroimaging[J]. Frontiers in Human Neuroscience, 2025, 19: 1552178.

[10] Qu Gang, Zhou Ziyu, Calhoun V D, et al. Integrated brain connectivity analysis with fMRI, DTI, and sMRI powered by interpretable graph neural networks[J]. Medical Image Analysis, 2025, 103: 103570.

[11] Michel P, Levy O, Neubig G. Are sixteen heads really better than one [C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2019: 1257.

[12] Voita E, Talbot D, Moiseev F, et al. Analyzing multi-head self-attention: Specialized heads do the heavy lifting, the rest can be pruned[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. New York: ACL, 2019: 5797-5808.

[13] Bravo-Ortiz M A, Holguin-Garcia S A, Quiñones-Arredondo S, et al. A systematic review of vision transformers and convolutional neural networks for Alzheimer’s disease classification using 3D MRI images[J]. Neural Computing and Applications, 2024, 36(35): 21985-22012.

[14] 胡杰, 昌敏杰, 徐博远, 等. ConvFormer: 基于Transformer的视觉主干网络[J]. 电子学报, 2024, 52(1): 46-57.
Hu Jie, Chang Minjie, Xu Boyuan, et al. ConvFormer: Vision backbone network based on Transformer[J]. Acta Electronica Sinica, 2024, 52(1): 46-57. (in Chinese)

[15] 励志勇, 姜伟, 刘浩杰. 可见光-红外跨模态行人重识别研究综述[J]. 电子学报, 2025, 53(12): 4811-4832.
Li Zhiyong, Jiang Wei, Liu Haojie. A survey on visible-infrared cross-modality person re-identification[J]. Acta Electronica Sinica, 2025, 53(12): 4811-4832. (in Chinese)

[16] Lu Peixin, Hu Lianting, Mitelpunkt A, et al. A hierarchical attention-based multimodal fusion framework for predicting the progression of Alzheimer’s disease[J]. Biomedical Signal Processing and Control, 2024, 88: 105669.

[17] Tang Chaosheng, Wei Mingyang, Sun Junding, et al. CsAGP: Detecting Alzheimer’s disease from multimodal images via dual-transformer with cross-attention and graph pooling[J]. Journal of King Saud University-Computer and Information Sciences, 2023, 35(7): 101618.

[18] Xiao Da, Meng Qingye, Li Shengping, et al. Improving transformers with dynamically composable multi-head attention[C/OL]//Proceedings of the 41st International Conference on Machine Learning, 2024: 2231. https://www.emergentmind.com/papers/2405.08553.

[19] Abbas S Q, Chi Lianhua, Chen Y P P. Transformed domain convolutional neural network for Alzheimer's disease diagnosis using structural MRI[J]. Pattern Recognition, 2023, 133: 109031.

[20] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.

[21] 张晓华, 练秋生. 基于小波域的复数卷积和复数Transformer的轻量级MR图像重建方法[J]. 电子学报, 2025, 53(4): 1221-1231.
Zhang Xiaohua, Lian Qiusheng. Lightweight MR image reconstruction network based on wavelet domain complex convolution and complex transformer[J]. Acta Electronica Sinica, 2025, 53(4): 1221-1231. (in Chinese)

[22] 刘诗怡, 刘金平, 黄丽娟, 等. 基于多尺度协调卷积与自适应加权的红外与可见光图像融合[J]. 智能系统学报, 2026, 21(1): 95-108.
Liu Shiyi, Liu Jinping, Huang Lijuan, et al. Infrared and visible image fusion based on multi-scale coordinated convolution and adaptive weighting[J]. CAAI Transactions on Intelligent Systems, 2026, 21(1): 95-108.(in Chinese)

[23] Hatamizadeh A, Tang Yucheng, Nath V, et al. UNETR: Transformers for 3D medical image segmentation[C]// 2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2022: 1748-1758.

[24] Zhu Jiayi, Tan Ying, Lin Rude, et al. Efficient self-attention mechanism and structural distilling model for Alzheimer´s disease diagnosis[J]. Computers in Biology and Medicine, 2022, 147: 105737.

[25] Lu Siyuan, Zhang Yudong, Yao Yudong. An efficient vision transformer for Alzheimer's disease classification using magnetic resonance images[J]. Biomedical Signal Processing and Control, 2025, 101: 107263.

[26] Bhojanapalli S, Yun C, Rawat A S, et al. Low-rank bottleneck in multi-head attention models[C]//Proceedings of the 37th International Conference on Machine Learning. MA:PMLR, 2020: 864-873.

[27] Shamshad F, Khan S, Zamir S W, et al. Transformers in medical imaging: A survey[J]. Medical Image Analysis, 2023, 88: 102802.

[28] 王芳芳, 刘明华, 渠连恩, 等. 基于图卷积与自适应Transformer的行人轨迹预测[J]. 电子学报, 2025, 53(12): 4507-4517.
Wang Fangfang, Liu Minghua, Qu Lian'en, et al. Pedestrian trajectory prediction based on graph convolution and adaptive transformer[J]. Acta Electronica Sinica, 2025, 53(12): 4507-4517.(in Chinese)

[29] Warren S L, Moustafa A A. Functional magnetic resonance imaging, deep learning, and Alzheimer's disease: A systematic review[J]. Journal of Neuroimaging, 2023, 33(1): 5-18.

[30] Duc N T, Ryu S, Qureshi M N I, et al. 3D-deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI[J]. Neuroinformatics, 2020, 18(1): 71-86.

[31] Li Wei, Lin Xuefeng, Chen Xi. Detecting Alzheimer's disease based on 4D fMRI: An exploration under deep learning framework[J]. Neurocomputing, 2020, 388: 280-287.

[32] Hu Yan, Wang Jun, Zhu Hao, et al. Cost-sensitive weighted contrastive learning based on graph convolutional networks for imbalanced Alzheimer's disease staging[J]. IEEE Transactions on Medical Imaging, 2024, 43(9): 3126-3136.

[33] Bhojanapalli S, Chakrabarti A, Jain H, et al. Eigen analysis of self-attention and its reconstruction from partial computation[PP/OL]. V1. arXiv (2021-06-16) [2025-11-20]. https://arxiv.org/abs/2106.08823.

[34] Malkiel I, Rosenman G, Wolf L, et al. Self-supervised transformers for fMRI representation[C]//Proceedings of the 5th International Conference on Medical Imaging with Deep Learning.MA: JMLR.org, 2022: 895-913.

[35] Nguyen S, Ng B, Kaplan A D, et al. Attend and decode: 4D fMRI task state decoding using attention models[C]// Proceedings of the Machine Learning for Health. MA: JMLR.org, 2020: 267-279.

[36] Rahman M M, Mahmood U, Lewis N, et al. Interpreting models interpreting brain dynamics[J]. Scientific Reports, 2022, 12(1): 12023.

[37] Choudhury C, Goel T, Tanveer M. A coupled-GAN architecture to fuse MRI and PET image features for multi-stage classification of Alzheimer's disease[J]. Information Fusion, 2024, 109: 102415.

[38] Hojjati S H, Ebrahimzadeh A, Khazaee A, et al. Predicting conversion from MCI to AD by integrating rs-fMRI and structural MRI[J]. Computers in Biology and Medi-

cine, 2018, 102: 30-39.
[39] Shukla A, Tiwari R, Tiwari S. Alzheimer's disease detection from fused PET and MRI modalities using an ensemble classifier[J]. Machine Learning and Knowledge Extraction, 2023, 5(2): 512-538.
[40] Abdelaziz M, Wang Tianfu, Anwaar W, et al. Multi-scale multimodal deep learning framework for Alzheimer's disease diagnosis[J]. Computers in Biology and Medicine, 2025, 184: 109438.
[41] Zhou Wenjun, Luo Weicheng, Gong Liang, et al. Enhanced early diagnosis of Alzheimer's disease with HybridCA-Net: A multimodal fusion approach[J]. Expert Systems with Applications, 2025, 292: 128580.
[42] Li Xiaoxiao, Zhou Yuan, Dvornek N, et al. BrainGNN: Interpretable brain graph neural network for fMRI analysis[J]. Medical Image Analysis, 2021, 74: 102233.
[43] Hatamizadeh A, Nath V, Tang Yucheng, et al. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in mri images[C]//Proceedings of the 7th International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Heidelberg: Springer, 2022: 272-284.
[44] Chen Zhi, Liu Yongguo, Zhang Yun, et al. Orthogonal latent space learning with feature weighting and graph learning for multimodal Alzheimer's disease diagnosis[J]. Medical Image Analysis, 2023, 84: 102698.
[45] Bi Yuda, Abrol A, Fu Zening, et al. A multimodal vision transformer for interpretable fusion of functional and structural neuroimaging data[J]. Human Brain Mapping, 2024, 45(17): e26783.
[46] Wang Haoyu, Guo Sizheng, Ye Jin, et al. SAM-Med3D: A vision foundation model for general-purpose segmentation on volumetric medical images[J]. IEEE Transactions on Neural Networks and Learning Systems, 2025, 36(10): 17599-17612.
[47] Zou Qi, Shang Junliang, Liu Jinxing, et al. BIGFormer: A graph transformer with local structure awareness for diagnosis and pathogenesis identification of Alzheimer's disease using imaging genetic data[J]. IEEE Journal of Biomedical and Health Informatics, 2025, 29(1): 495-506.
[48] Qiu Zifeng, Yang Peng, Xiao Chunlun, 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.
[49] Cheng Jiayuan, Wang Huabin, Wei Shicheng, et al. Alzheimer's disease prediction algorithm based on de-correlation constraint and multi-modal feature interaction[J]. Computers in Biology and Medicine, 2024, 170: 108000.

## 作者简介

**刘金平** 男，1983年9月出生于湖南省洞口县。现为湖南师范大学信息科学与工程学院教授、博士生导师。在国内外期刊发表学术论文80余篇，授权发明专利26项。主要从事大数据分析、工业故障诊断、智能监控等研究。
E-mail: ljp202518@163.com

**李兴旺** 男，2002年5月出生于湖南省娄底市。现为湖南师范大学软件工程专业硕士研究生。研究方向为多模态医学图像分析与智能医学诊断相关研究。
E-mail: lxw@hunnu.edu.cn

**刘家瑜** 男，2001年10月出生于湖南省洞口县。现为湖南师范大学信息科学与工程学院硕士研究生。研究方向为机器学习与智能图像处理。
E-mail: 202570294510@hunnu.edu.cn

**刘芷娴** 女，2007年6月出生于湖南省洞口县。现为湖南师范大学商学院本科生，研究兴趣为大数据智能决策。
E-mail: 202430072074@hunnu.edu.cn

**刘亚琴** 女，1981年12月出生于湖南省邵阳县。现为湖南师范大学数学与统计学院副教授、硕士生导师。主要从事医学图像分析、大数据智能决策相关研究。
E-mail: 17873801607@163.com