

黎舰泽, 刘立程, 吴慧东. 基于多尺度通道空间感知Mamba的阿尔茨海默症PET影像分类方法[J]. 广东工业大学学报, 2026, 43(1): 31-39. doi: 10.12052/gdutxb.250037.  
Li Jianze, Liu Licheng, Wu Huidong. Multi-scale channel-spatial perception Mamba for Alzheimer's disease PET image classification method[J]. Journal of Guangdong University of Technology, 2026, 43(1): 31-39. doi: 10.12052/gdutxb.250037.

# 基于多尺度通道空间感知Mamba的阿尔茨海默症 PET影像分类方法

黎舰泽, 刘立程, 吴慧东

(广东工业大学 信息工程学院, 广东 广州 510006)

**摘要:** 针对现有基于卷积神经网络和Transformer的医学影像诊断方法存在长程依赖建模能力不足及二次计算复杂度问题, 本文提出一种基于多尺度通道空间感知Mamba模型的三维正电子发射断层扫描(Positron Emission Tomography, PET)影像分类框架SSHCM(State Space Hybrid Convolutional Model)。该模型融合线性状态空间模型与多尺度特征交互机制, 通过堆叠LMamba块对三维体素序列进行长程动态建模; 设计逐层跨尺度通道注意力融合模块实现全局上下文语义自适应融合; 构建通道空间感知模块, 结合大核卷积与倒置式瓶颈结构优化空间特征融合, 提升病灶定位精度。在阿尔茨海默病神经影像学计划1187名受试者数据集上的结果表明: 本方法的准确率和AUC较ResNet、ViT及Mamba变种模型提升明显。在AD分类任务和MCI转化预测任务中的准确率分别达到97.03%和83.33%。

**关键词:** 阿尔茨海默病; 正电子发射断层扫描; Mamba; 状态空间模型; 图像分类

中图分类号: TN911.73;

文献标志码: A

文章编号: 1007-7162(2026)01-0031-09

## Multi-scale Channel-spatial Perception Mamba for Alzheimer's Disease PET Image Classification Method

Li Jianze, Liu Licheng, Wu Huidong

(School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China)

**Abstract:** To address the limitations of existing medical image diagnosis methods based on convolutional neural networks and Transformers, such as insufficient long-range dependency modeling and quadratic computational complexity, this paper proposes a 3D positron emission tomography (PET) image classification framework named state space hybrid convolutional model (SSHCM). The framework is built upon a multi-scale channel space perception Mamba architecture, which integrates a linear state-space model with a multi-scale feature interaction mechanism, using stacked LMamba blocks to capture long-range dependencies in 3D voxel sequences dynamically. A layer-wise cross-scale channel attention fusion module is designed to achieve adaptive fusion of global contextual semantic. Additionally, a channel-spatial perception module is constructed by combining large kernel convolutions with an inverted bottleneck structure, enhancing spatial feature fusion and improving lesion localization accuracy. Experimental results on the Alzheimer's Disease Neuroimaging Initiative dataset with 1187 subjects show that the proposed model significantly outperforms ResNet, ViT, and Mamba variant models in terms of both accuracy and AUC. Specifically, the model achieves accuracy rates of 97.03% for AD classification and 83.33% for MCI conversion prediction tasks.

**Key words:** Alzheimer's disease; positron emission tomography; Mamba; state space models; image classification

收稿日期: 2025-02-17 录用日期: 2025-04-28 网络首发日期: 2025-10-31

基金项目: 国家自然科学基金资助项目(61976058, 61300107); 广东省自然科学基金资助项目(S2012010010212); 广州市科技计划项目(202206010007)

作者简介: 黎舰泽(2000-), 男, 硕士研究生, 主要研究方向为深度学习在阿尔茨海默病诊断的应用, E-mail: jzli\_cn@163.com

通信作者: 刘立程(1972-), 男, 副教授, 博士, 主要研究方向为信号处理、模式识别、人工智能在视频图像中的应用, E-mail: celcliu@gdut.edu.cn

阿尔茨海默症(Alzheimer's Disease, AD)是最常见的神经退行性痴呆类型,表现为记忆衰退和认知功能进行性下降,严重威胁老年人健康并加重社会负担<sup>[1]</sup>。AD的病理机制与淀粉样蛋白斑块和TAU蛋白缠结的异常积累密切相关<sup>[2]</sup>,这些病变可以通过正电子发射断层扫描(Positron Emission Tomography, PET)在分子水平上可视化<sup>[3]</sup>。其中18F-FDG PET能够检测脑葡萄糖代谢异常,成为识别AD的重要生物标志物<sup>[4]</sup>。然而,PET图像分析依赖专家的先验知识,且缺乏明确的轻度认知障碍(Mild Cognitive Impairment, MCI)向AD转化的客观标志,因此亟需高效自动化诊断方法。

近年来,深度学习为医学影像分析提供了新的方向。相较于依赖人工特征提取的传统机器学习方法,卷积神经网络(Convolutional Neural Network, CNN)通过端到端学习显著提升了特征泛化能力<sup>[5]</sup>,然而,CNN在处理3D医学影像时面临较高的计算开销<sup>[6]</sup>且受限于局部感受野,导致其难以捕捉跨脑区代谢关联。此外CNN需要额外的计算资源以及精巧的结构设计来处理3D高维数据,这将导致训练效率低下和训练时间增加<sup>[7]</sup>。徐萍萍等<sup>[8]</sup>提出的AFEM-Transformer模型通过改进Swin Transformer实现3D块划分,并创新性地引入自适应特征提取模块(Adaptive Feature Extraction Module, AFEM),有效解决了传统3D卷积神经网络在sMRI影像全局特征捕获和病灶区域精准定位方面的局限性。但其自注意力机制的二次复杂度限制了其将3D数据完整输入进网络进行分析<sup>[9]</sup>。

为了解决Transformer在长序列处理中的计算效率问题,Gu等<sup>[10]</sup>提出了Mamba模型,通过引入选择性状态空间的机制,能够在保持线性时间复杂度的同时获得媲美Transformer的性能,为医学影像分析提供了新方向。例如,MedMamba<sup>[11]</sup>引入了SS-Conv SSM的混合块,将卷积层的局部特征提取能力与SSM捕获长程依赖性的能力相结合。VMamba<sup>[12]</sup>则通过双向状态空间建模强化了视觉上下文理解。Han等<sup>[13]</sup>揭示了门控机制和模块设计是Mamba模型在视觉任务中取得成功的核心因素。这些Mamba变体虽通过状态空间模型实现了线性复杂度,但未针对3D PET影像特征的高维度、低信噪比及病灶区域代谢弱特点进行优化。

为解决这一问题,本文将上述门控机制和模块化设计融入到线性注意力模型中,提高PET影像分类任务中的效果。本文设计了一种状态空间混合卷积模型(State Space Hybrid Convolutional Model,

SSHCM)。SSHCM结合了线性时间复杂度的优势和门控机制的噪声鲁棒性,融合多分辨率代谢特征,对微弱代谢变化病灶进行检测,能够有效地应对3D PET影像的挑战。该模型通过动态转换输入特征,能够有效地捕获和保留3D PET影像中的重要空间信息,从而提升PET图像的分类性能。模型的主干部分由多个LMamba块(Linear Mamba Block)构成,用于建模不同尺度的全局信息。逐层跨尺度通道注意力融合(Cross Scale Channel Attention Fusion, CSCAF)模块促进不同尺度间的特征信息交互,以增强海马区等微小病灶的定位精度。此外,为了优化高维语义特征融合,设计了通道空间感知机制(Channel and Spatial Perception Mechanism, CSPM)进行特征融合,使得融合后的特征兼具全局信息与局部特征敏感性。并通过倒置式瓶颈块(Inverted Bottleneck Block, IBB)平衡特征表达与计算效率。实验表明,SSHCM在AD分类和MCI转化预测中展现出显著优势,为高精度、低成本的智能辅助诊断提供了新的范式。

## 1 SSHCM模型

### 1.1 模型整体架构

图1展示了本文提出的SSHCM模型整体架构,模型输入三维PET影像(尺寸 $B \times C \times W \times H \times D$ ),采用多阶段渐进式特征提取策略。主干部分使用Kernel = 7的大核卷积扩大感受野,保留更多的空间细节。核心结构由LMamba块构成,每个LMamba块包含空间特征融合模块(Space Convergent Block, SCB)和状态空间混合卷积模块(State Space Hybrid Convolutional Block, SSHCB)。SSHCB中设计了类Mamba的门控结构,集成了线性注意力机制,构成了类Mamba线性注意力(Mamba-Type Linear Attention, MTLA),MTLA核心部分加入了额外的深度卷积过程以捕获更多的局部特征,提高对微型病灶的定位精度和分类效率。随后将主干提取到的特征 $F_0$ 与跨尺度支路提取到的特征 $F_1$ 通过CSCAF块交互后充分融合以增强多尺度特征表达能力,这确保了模型能捕获单一尺度可能遗漏的细节或上下文信息,从而提高模型的鲁棒性。阶段之间使用卷积核大小为2的卷积层进行空间下采样,最终通过CSPM对融合特征进行维度压缩与空间关系建模,并经由IBB进行高维语义特征提取,得到脑部细节特征,输出至分类器完成AD诊断决策。

### 1.2 空间特征融合模块SCB

受遗忘门机制启发,本文设计了一个空间信息

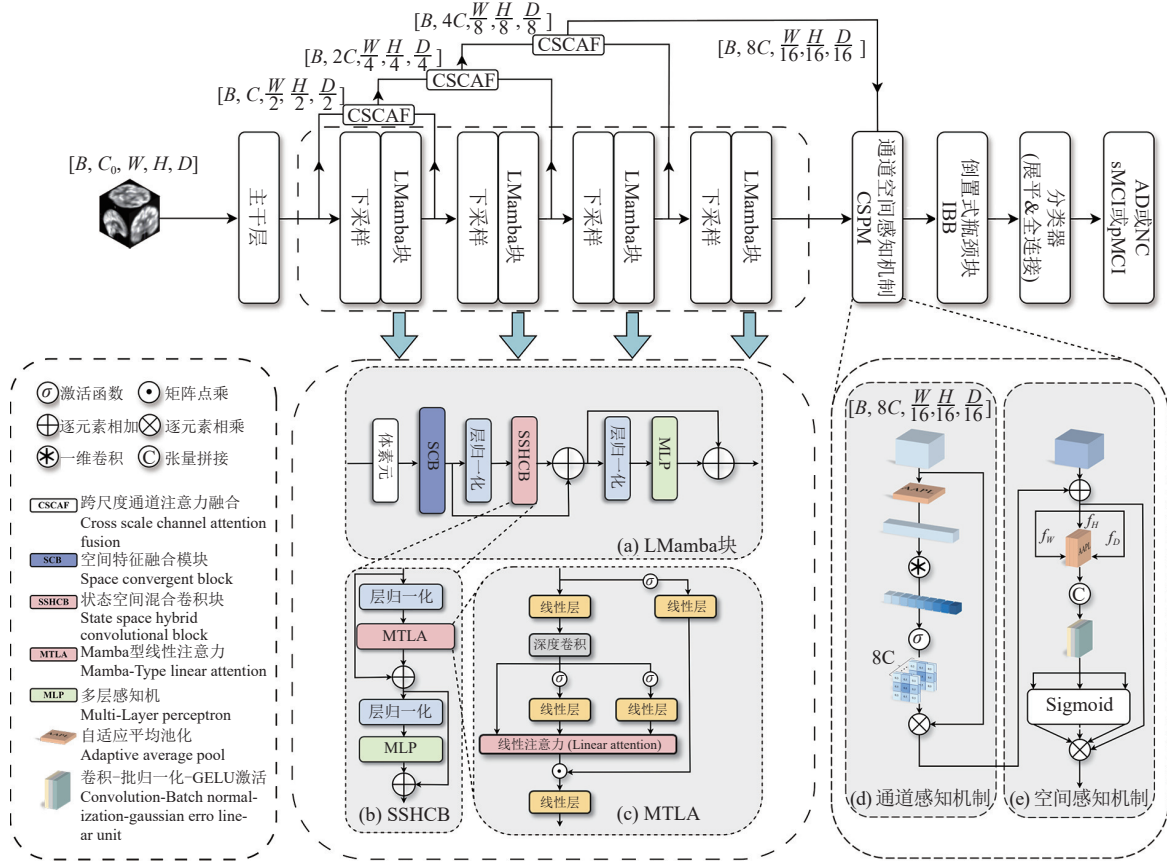


图 1 SSHCM整体框架

Fig.1 The SSHCM overall architecture

聚合块SCB,用在序列特征之前,以有效地感知和保留空间信息,从而实现对特征间长程依赖关系的高效建模。其整体结构如图2所示。

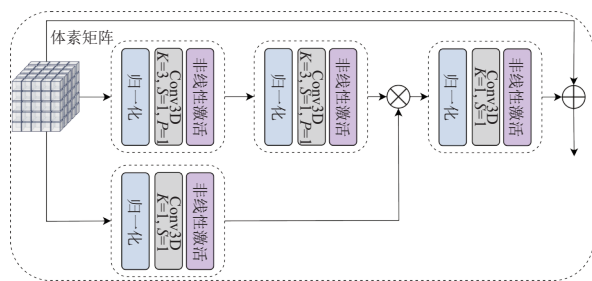


图 2 空间特征融合模块

Fig.2 Space convergent block

SCB采用双分支结构对输入特征进行多尺度空间建模。输入的3D特征先经过两层包含归一化、卷积和非线性激活的卷积块进行处理,卷积核大小及步长和填充参数如图2所示。Kernel = 3的卷积操作能捕捉全局性的复杂空间特征,Kernel = 1的卷积对这些全局空间特征进行调整与融合,同时也能在通道维度上进行信息的压缩或扩展。最后通过Kernel = 1的卷积块调整信息的维度,将两个分支的特征进行

融合,以增强特征的空间表达能力。输入的残差连接有效地避免了梯度消失问题同时帮助模型保留原始信息,使得网络学习到不同层次的特征并优化,保证信息的有效流动。用公式表示为

$$SCB(f) = f + Conv_1((Conv_3(Conv_3(f))) \cdot Conv_1(f)) \quad (1)$$

式中:  $f$  表示输入3D特征,  $Conv_i$  表示卷积核大小为  $i$  的卷积块。

### 1.3 状态空间混合卷积块SSHCB

SSHCB是一个深度学习网络层,其核心功能是对空间-通道解耦建模,将输入的高维特征展平为序列形式,分离空间与通道维度,以适配轻量化特征变换。它通过卷积位置编码注入位置信息,增强模型对局部结构的敏感性。采用线性注意力替代自注意力,显著降低了计算复杂度,使其适合长序列输入。结合残差多路径融合,通过残差连接与多分支特征乘法交互,以提升梯度流动并捕获跨尺度特征依赖。其主要的运算步骤是通过MTLA实现的。

线性注意力<sup>[14]</sup>不同于单头注意力计算每个查询键对之间的相似性,从而导致  $O(L^2)$  复杂度在长序列

建模任务中产生无法承担的计算开销。使用线性归一化代替了非线性的Softmax函数并在 $\mathbf{Q}$ 和 $\mathbf{K}$ 中采用额外的核函数 $\phi$ ,这使得基于矩阵乘法的结合特性将计算顺序从 $(\mathbf{Q}\mathbf{K}^T)\mathbf{V}$ 重排为 $\mathbf{Q}(\mathbf{K}^T\mathbf{V})$ ,从而将复杂度降低到 $O(L)$ 。为了将具有全局感受野的线性注意力嵌入到自回归模型中,将第 $i$ 个词元的感受野限制为前面(索引小于等于 $j$ )的词元,即 $j \leq i$ 。该因果线性注意力的公式为

$$\mathbf{Y}_i = \frac{\mathbf{Q}_i(\sum_{j=1}^i \mathbf{K}_j^T \mathbf{V}_j)}{\mathbf{Q}_i(\sum_{j=1}^i \mathbf{K}_j^T)} \triangleq \frac{\mathbf{Q}_i \mathbf{S}_i}{\mathbf{Q}_i \mathbf{Z}_i} \quad (2)$$

$$\mathbf{S}_i = \sum_{j=1}^i \mathbf{K}_j^T \mathbf{V}_j \quad (3)$$

$$\mathbf{Z}_i = \sum_{j=1}^i \mathbf{K}_j^T \quad (4)$$

式中:  $\mathbf{Q} = x\mathbf{W}_Q$ ,  $\mathbf{K} = x\mathbf{W}_K$ ,  $\mathbf{V} = x\mathbf{W}_V$ 。进一步将式(2~4)改写为循环的线性注意力形式。

$$\mathbf{S}_i = \mathbf{S}_{i-1} + \mathbf{K}_i^T \mathbf{V}_i, \mathbf{Z}_i = \mathbf{Z}_{i-1} + \mathbf{K}_i^T, \mathbf{Y}_i = \frac{\mathbf{Q}_i \mathbf{S}_i}{\mathbf{Q}_i \mathbf{Z}_i} \quad (5)$$

式中:  $\mathbf{S}_i \in \mathbb{R}^{d \times C}$ ,  $\mathbf{K}_i^T \in \mathbb{R}^{d \times 1}$ ,  $\mathbf{V}_i \in \mathbb{R}^{1 \times d}$ ,  $\mathbf{Q}_i \in \mathbb{R}^{1 \times d}$ 。

经典状态空间模型是一个连续系统,通过隐藏状态 $h_t \in \mathbb{R}^{d \times 1}$ 将输入 $x(t) \in \mathbb{R}$ 映射到输出 $y(t) \in \mathbb{R}$ ,公式表达为

$$\begin{cases} h'(t) = \mathbf{A}h(t) + \mathbf{B}x(t) \\ y(t) = \mathbf{C}h(t) + \mathbf{D}x(t) \end{cases} \quad (6)$$

对于 $N$ 个词元 $x \in \mathbb{R}^{N \times C}$ 的输入序列,循环线性注意力式(5)与状态空间模型具有相似之处,根据隐状态 $\mathbf{S}_i$ 可以预测未来的输出 $\mathbf{Y}_i$ 。将其改写为状态空间模型类似的形式,对于 $N$ 个词元的输入序列 $x$ ,选择性线性注意力状态空间模型定义为

$$\mathbf{S}_i = 1 \odot \mathbf{S}_{i-1} + \mathbf{K}_i^T (1 \odot \mathbf{V}_i) \quad (7)$$

$$\mathbf{Y}_i = \frac{\mathbf{Q}_i \mathbf{S}_i}{\mathbf{Q}_i \mathbf{Z}_i} + 0 \odot x_i \quad (8)$$

将状态空间模型的 $\mathbf{A}$ 矩阵简化为1, $\mathbf{D}$ 矩阵简化为0。线性注意力输出除以 $\mathbf{Q}_i \mathbf{Z}_i$ ,以保持注意力权重之和为1。

总的来说,MTLA是对Mamba块进行的改进,加入了额外的深度卷积和残差连接,通过线性投影以 $O(L)$ 的复杂度替代自注意力 $O(L^2)$ 进行计算,解决Transformer二次计算复杂度的开销难题。通过门控交互,实现特征激活与注意力权重的动态调制。模块内置随机丢弃路径以提升泛化性。

#### 1.4 跨尺度通道注意力融合模组CSCAF

为了全局融合不同尺度的特征信息,使得提取到的抽象的高级特征图更加符合人类视觉系统对高级语义特征的理解,本文设计了CSCAF来解决这个问题,如图3所示。

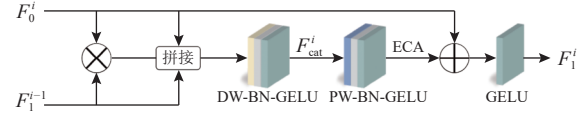


图3 跨尺度通道注意力融合模组

Fig.3 Cross scale channel attention fusion module

首先将主干的输出 $F_0^i$ (其中 $i$ 为阶段数, $i=1,2,3,4$ )和融合的特征图 $F_1^{i-1}$ 进行逐元素相乘,将元素相乘结果与原始特征图在通道维度上进行级联,紧接着在通道维度上将级联结果与原始特征映射进行整合。拼接后的特征图经由一个组合层处理,该层依次通过深度可分离卷积提取局部特征并降低计算量,经批归一化加速训练并稳定网络,最后通过GELU激活函数引入非线性以平滑梯度流,从而增强模型对复杂模式的建模能力。使用点卷积的组合层改变通道数实现特征图的重映射,以增加模型的非线性表达能力。采用高效通道注意力(Efficient Channel Attention, ECA<sup>[15]</sup>)机制实现全局特征感知,促进特征图的进一步融合。最后将 $F_0^i$ 与感知特征图相加,经过GELU激活后得到融合的特征图 $F_1^i$ 。这一过程表达式为

$$F_{\text{cat}}^i = \sigma(\text{BN}(\text{DW}(\text{cat}(F_0^i, F_1^{i-1}, F_0^i \otimes F_1^{i-1})))) \quad (9)$$

$$F_1^i = \sigma(\text{ECA}(\sigma(\text{BN}(\text{PW}(F_{\text{cat}}^i)))) + F_0^i) \quad (10)$$

式中:  $\text{cat}(\cdot)$ 表示将特征图沿通道维度串联, $\sigma(\cdot)$ 表示GELU激活函数,  $\text{BN}$ (Batch Normalization)、 $\text{DW}$ (Depthwise Convolution)、 $\text{PW}$ (Pointwise Convolution)分别表示批归一化、深度可分离卷积和点卷积。

#### 1.5 通道空间感知机制CSPM

CSPM整合了来自主干和CSCAF的详细多尺度空间特征信息,CSPM的架构如图1(d)、(e)所示。它由一个接受来自主干输出 $F_0$ 作为输入的通道感知网络,以及一个接受来自通道感知网络输出和CSCAF输出的全局特征信息 $F_1$ 叠加作为输入的空间感知网络组成。传统的通道注意力感知<sup>[16]</sup>会降低维度以控制模型的复杂度,但是降低维度会给通道注意力的预测造成阻碍,且捕获所有通道之间的注意力依赖关系是低效的。为了更有效地捕捉跨通道交互,本文设计了一种没有降维的通道感知机制。具

体地,  $F_0$ 在经过没有降维的自适应全局平均池化后,通过核大小为 $P$ 的一维卷积(Conv1d)处理每个通道及其 $P$ 个最近邻,以有效捕捉局部跨通道交互。通过式(11)确定 $P$ 的大小以获得最佳覆盖率:

$$P = \left\lfloor \frac{\log_2(C) + b}{\gamma} \right\rfloor_{\text{odd}} \quad (11)$$

式中: $\lfloor \cdot \rfloor_{\text{odd}}$ 表示最近邻的奇数, $\gamma = 2$ ,  $b = 1$ 。根据上述非线性映射关系可知,高价值通道的特征交互时间更长。因此通道感知机制可以对多个通道特征进行更深层次的挖掘,上述过程表示为

$$F_{\text{cp}} = \sigma_2(\text{Conv1d}(\text{Avgpool}(F_0^4))) * F_0^4 \quad (12)$$

式中: $F_{\text{cp}}$ 表示通道感知的输出,通过一维卷积(Conv1d)进行轻量化交互, $\sigma_2$ 表示sigmoid激活函数,Avgpool表示平均池化,\*为逐元素乘法。

本文将 $F_{\text{cp}}$ 与 $F_0^4$ 相结合作为互补的信息,记作 $F_{\text{in}}$ ,用于空间感知网络的输入,从而提取关键的空间位置信息。空间注意力机制从宽度、高度和深度( $W, H, D$ )3个维度分别出发,将输入分割为3个部分( $f_w, f_h, f_d$ )分别处理,进行特征编码,并进行池化以保留特征图的空间结构信息:

$$f_w = \frac{1}{W} \sum_{0 \leq j, k \leq W} F_{\text{in}}(i, j, k) \quad (13)$$

$$f_h = \frac{1}{H} \sum_{0 \leq i, k \leq H} F_{\text{in}}(i, j, k) \quad (14)$$

$$f_d = \frac{1}{D} \sum_{0 \leq i, j \leq D} F_{\text{in}}(i, j, k) \quad (15)$$

并使用卷积块进行空间特征提取,生成各维度的特征权重:

$$(f'_w, f'_h, f'_d) = \text{split}[\sigma[\text{BN}[\text{Conv1d}[\text{cat}(f_w, f_h, f_d)]]]] \quad (16)$$

重复记号不赘述,split表示沿高度、宽度、深度维度进行分割。将权重矩阵使用激活函数 $\sigma_2$ 处理后生成掩码为[0~1]范围的权重矩阵,与 $F_{\text{in}}$ 特征加权处理:

$$\mathbf{M}_{\text{CSPM}} = \sigma_2(f'_w, f'_h, f'_d), F_{\text{CSPM}} = \mathbf{M}_{\text{CSPM}} \times F_{\text{in}} \quad (17)$$

式中: $\mathbf{M}_{\text{CSPM}}$ 为通过 $\sigma_2$ 生成的通道空间注意力的权重系数矩阵, $F_{\text{CSPM}}$ 为CSPM输出特征图,表示全局和局部信息融合的高维语义特征。

## 1.6 倒置式瓶颈块IBB

为了充分混合远距空间和位置信息,高效地提取全局上下文信息,本文设计了大卷积核的倒置式瓶颈块。该结构如图4所示。

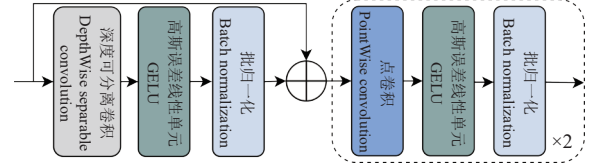


图4 倒置式瓶颈块  
Fig.4 Inverted bottleneck block

IBB旨在使模型在保持较低的计算复杂度的同时提高对细节特征的表达能。本文使用卷积核 $\text{Kernel} = 3$ 的深度可分离卷积来提取每个通道的全局信息,并使用残差结构进行连接以避免梯度消失。点卷积堆叠形成倒置式瓶颈块的核心操作,用于充分混合空间和通道信息。具体来说,在输入特征向量为高维的情况下,深度可分离卷积保持通道数不变,两个点卷积之间的隐藏维度设置为输入维度的4倍,扩展的隐藏维度能够全面充分地混合深度可分离卷积提取的全局空间维度信息。IBB的公式定义为

$$F''_{\text{out}} = \text{BN}(\sigma(\text{DW}(F_{\text{CSPM}}))) + F_{\text{CSPM}} \quad (18)$$

$$F'_{\text{out}} = \text{BN}(\sigma(\text{PW}(F''_{\text{out}}))) \quad (19)$$

$$F_{\text{out}} = \text{BN}(\sigma(\text{PW}(F'_{\text{out}}))) \quad (20)$$

式(18)表示IBB通过DW扩展输入 $F_{\text{CSPM}}$ 的通道维度,同时和输入形成残差连接,保留基底特征,输出记为 $F''_{\text{out}}$ 。式(19)表示通过PW进行特征提取,同时将通道维度扩展为输入的4倍,输出记为 $F'_{\text{out}}$ 。式(20)表示再次通过点卷积还原通道维度为输入维度,提取抽象的隐藏层高维特征,最终输出记为 $F_{\text{out}}$ 。这一设计在保持高维特征表现力的同时,增强了模型对细节特征的捕捉能力。

## 2 实验结果及讨论

### 2.1 数据来源

#### 2.1.1 ADNI数据库

本文的研究基于阿尔茨海默病神经影像学计划(Alzheimer's Disease Neuroimaging Initiative, ADNI)数据进行分析。从ADNI-1和ADNI-2阶段中选择了在质量评估方面最好的18F-FDG-PET扫描的1187名受试者组成数据集。数据集中受试者的分布如表1所示。其中NC为正常对照组(Normal Control)。

统计参数包括性别、年龄、简易精神状态检查(Mini-mental state examination, MMSE)、临床痴呆评分(Clinical dementia rating, CDR)。该数据集总体数据的66.7%和16.7%分别作为ADNI1的模型训练集和验

表1 受试者人口统计信息  
Table 1 Demographic information for subjects

数据集	类别	男/女	年龄/岁	MMSE/分	CDR/分
ADNI-1	AD	83/119	75.08±7.81	20.92±3.23	1.82±0.22
	NC	163/142	74.39±6.14	29.41±0.89	0.02±0.09
	MCI	199/284	73.66±7.47	26.76±1.52	0.79±0.71
ADNI-2	AD	21/19	74.42±7.19	22.94±2.43	0.85±0.25
	NC	30/31	74.95±5.96	28.27±1.30	0.03±0.00
	MCI	62/34	72.82±7.41	27.67±1.99	0.88±0.46

证集,剩余部分为ADNI2的泛化性验证测试集。训练集、验证集、测试集的整体比例为4:1:1。

### 2.1.2 数据预处理

使用基于SPM(Statistical Parametric Mapping)平台的CAT12(Computational Anatomy Toolbox)软件工具包进行脑影像预处理。主要包括蒙特利尔神经学研究所(Montreal Neurological Institute, MNI)空间配准,去除颅骨和8 mm半峰全宽高斯平滑,处理后得到图像维度为 $121 \times 145 \times 121(W \times H \times D)$ ,每个体素的空间分辨率为 $1.5 \text{ mm} \times 1.5 \text{ mm} \times 1.5 \text{ mm}$ 。对每张图像进行基于体素的强度归一化,即每个体素的值被归一化为原始值除以其所在图像的最大值。图5总结了主要操作流程。

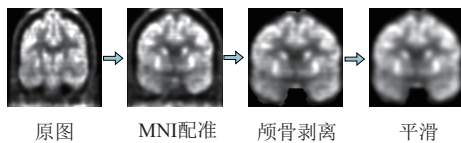


图5 18FFDG-PET预处理流程图

Fig.5 18FFDG-PET preprocessing flowchart

### 2.1.3 数据增强

为了提升模型的泛化能力和增强训练数据的多样性,在预处理之后,训练数据加载时通过对50%比例的原始数据进行数据增强操作生成多样化的数据,从而缓解小规模数据集数据不足导致的过拟合问题。在配准后的图像上继续进行数据增强操作,不会破坏配准到标准空间的效果。数据增强操作主要包括随机旋转、伽马调整、随机遮掩。其中随机旋转是在 $(-30^\circ \sim +30^\circ)$ 角度范围内旋转,模拟不同的视角,增强模型对输入数据角度变化的适应性;伽马调整通过调整图像的伽马值非线性地调整图像的亮度分布,提升模型适应不同对比度下的图像;随机遮掩模拟部分丢失和遮挡的情况,促使模型学习更多全局特征提高鲁棒性。

## 2.2 实验参数设定及评价指标

本文实验基于配备GeForce RTX 3090 GPU、

Intel Core i9-14900K CPU的Linux服务器上完成,使用PyTorch2.0.1、Python3.11.9框架实现。操作系统为Ubuntu24.04.1,CUDA版本为11.7。为避免数据泄露,ADNI2测试集数据已提前按比例划分。实验使用Adam优化器,批次大小为6,初始学习率为0.00001。使用双余弦退火策略灵活调整学习率,震荡周期为75,总训练轮次为150。

准确度(Accuracy, ACC)、灵敏度(Sensitivity, SEN)、特异度(Specificity, SPE)、受试者工作特征曲线(Receiver Operating Characteristic, ROC)的曲线下面积(Area Under the Curve, AUC)等评价指标通常用于评估深度学习模型在阿尔茨海默病图像分类中的性能。这些指标验证了模型从神经影像中正确识别AD的能力,有助于算法的优化与诊断准确性的提高。通过这些评估指标,本文可以预估模型在临床环境中的可靠性和有效性。对应的公式定义为

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (21)$$

$$SEN = \frac{TP}{TP + FN} \quad (22)$$

$$SPE = \frac{TN}{TN + FP} \quad (23)$$

式中:T代表正例,P代表阳性。反之F为反例,N为阴性。例如TP代表真阳性。SEN反映了模型对正类的识别能力,适用于漏判代价高的场景。SPE反映了模型对负类的识别能力,适用于误报代价高的场景。ROC曲线越接近左上角,模型的分类能力越强。

## 2.3 实验结果分析

### 2.3.1 对比实验

本研究采用五折交叉验证策略综合评估SSHCM性能,并通过对比分析ResNet、ViT及Mamba变种模型验证架构有效性。由于现有研究大多聚焦于AD与NC的二分类任务,而同时涵盖AD/NC分类与MCI转化预测的多任务研究较为有限,因此本文主要针对主流的AD与NC分类任务进

行方法验证与性能对比,以确保实验结果的可比性与可靠性。MCI转化预测任务将作为补充实验,用于进一步验证模型的泛化能力。基线方法的对比结果如表2及图6所示。

表 2 多种基线方法的分类结果比较  
Table 2 Comparison of classification results obtained by different baseline methods

方法	参数量/MB	FLOPs/GB	ACC/%	
			AD vs NC	sMCI vs pMCI
ResNet <sup>[17]</sup>	33.1	85.7	72.90	67.70
ViT <sup>[18]</sup>	85.3	136.5	82.77	66.04
VMamba <sup>[12]</sup>	48.1	78.4	88.71	73.95
MedMamba <sup>[11]</sup>	<b>21.7</b>	<b>42.8</b>	92.48	75.12
SSHCM	36.7	53.6	<b>97.03</b>	<b>83.33</b>

SSHCM模型在AD分类任务中准确率达97.03%(AUC98.23%),显著优于ResNet和ViT模型,且相较

于MedMamba模型(ACC92.48%)提升4.55个百分点。在MCI转化预测任务中,SSHCM模型准确率达83.33%(AUC83.57%),准确率较MedMamba(75.12%)和VMamba(73.95%)分别提升8.21和9.38个百分点。

在计算效率方面,Mamba变种模型参数量优于ResNet和ViT模型,这也证明了Mamba模型的线性复杂度能够有效降低参数量,提高计算效率。SSHCM实现了性能与效率的平衡。FLOPs较ResNet降低37.5%,较ViT降低60.7%。相较于Mamba变种模型来说参数量仅为VMamba的76.3%,略逊于MedMamba,但准确率优于MedMamba,特别是较难的MCI转化预测任务中SSHCM较MedMamba提升了8.21个百分点,验证了SSHCM对复杂病灶的适应能力。结果表明,所提出的多尺度特征融合机制与状态空间建模策略有效平衡了全局依赖捕获与局部特征细化,为AD早期诊断提供了可靠的影像分析框架。

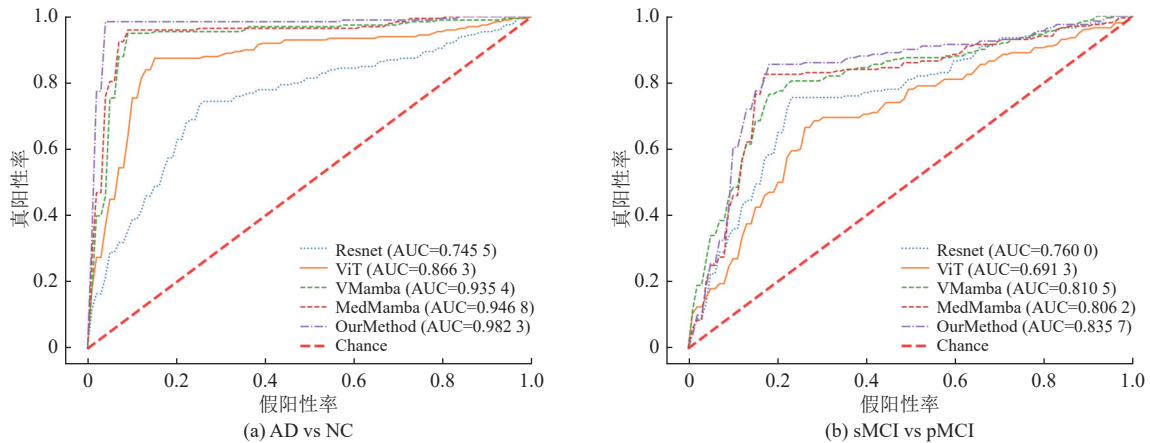


图 6 多种基线方法的ROC曲线  
Fig.6 ROC curves of multiple baseline methods

为了进一步验证模型的有效性,本文与现有研究成果进行AD/NC分类任务对比实验,实验结果如表3所示。

表 3 与现有研究成果对比的模型性能评估  
Table 3 Comparative performance evaluation against existing research %

方法	AD vs NC			
	ACC↑	SEN	SPE	AUC↑
文献[19]	93.13	90.32	95.49	97.11
文献[20]	91.83	90.04	93.54	96.84
文献[21]	89.80	86.30	91.90	-
文献[22]	92.31	95.50	90.21	96.00
文献[23]	84.80	76.10	88.70	-
文献[24]	96.90	96.10	<b>97.50</b>	96.90
文献[25]	92.00	<b>100</b>	84.00	96.00
本文方法	<b>97.03</b>	97.49	96.72	<b>98.23</b>

选取近年应用最先进技术的论文模型与本文提出的模型进行性能对比,实验结果表明本文提出的方法在AD与NC的分类任务中表现出显著优势。本文方法的准确率达到97.03%,较Rehman等提出的最优方法提升了0.13个百分点;灵敏度为97.49%,略高于Rehman的96.10%;特异性为96.72%,虽略低于Rehman的97.50%,但仍优于其他对比方法。此外,本文方法的AUC值达到98.23%,显著高于文献[19]方法的97.11%和Rehman的96.90%,表明模型在区分AD与NC时具有更高的判别能力。

### 2.3.2 消融实验

为了检验各设计模块对使用本文提出模型进行AD分类的影响,通过一系列的消融实验验证不同模块对模型性能的影响。为确保公平比较,使用不含

SCB的LMamba块作为基线主干网络,所有模型都使用2.2节的实验参数配置进行训练,结果如表4所示。消融实验结果分析表明,本文提出的各模块对模型性能均有显著贡献。在AD与NC分类任务中,单独引入SCB块使准确率从79.56%提升至86.13%,AUC值从86.34%提升至90.86%,表明SCB有效增强了空间

特征提取能力。而IBB模块使准确率提升8.16个百分点,SCB使准确率提升6.57个百分点,提升幅度相近,但IBB更好地平衡了敏感性与特异性,实现了更高的AUC(94.46%)。CSCAF和CSPM的联合应用展现出了更强的性能(ACC94.85%/AUC97.24%),揭示了跨尺度特征融合对阿尔茨海默病早期诊断具有关键作用。

表4 SSHCM的消融实验结果  
Table 4 Ablation study results of SSHCM

方法	SCB	IBB	CSCAF &CSPM	AD vs NC				sMCI vs pMCI			
				ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
Baseline				79.56	78.84	80.04	86.34	69.16	72.25	67.69	71.10
+SCB	√			86.13	72.49	95.08	90.86	71.66	<b>81.29</b>	67.07	78.15
+IBB		√		87.72	89.50	86.56	94.46	71.04	69.03	72.00	73.51
+CSCAF,CSPM			√	94.85	94.50	95.08	97.24	74.59	74.84	74.46	77.56
本文方法	√	√		<b>97.03</b>	<b>97.49</b>	<b>96.72</b>	<b>98.23</b>	<b>83.33</b>	70.96	<b>89.23</b>	<b>83.57</b>

在更具挑战的MCI转化预测任务中,单独加入模块相较于基线结果都有提升。但模块组合的协同效应尤为显著,本文方法在保持高敏感性的同时将特异性提升至89.23%。AUC达到83.57%,较基线提升12.47个百分点。值得注意的是,CSCAF与CSPM的组合在两类任务中均展现出超越单模块的性能增益(ACC74.59%/AUC77.56%),验证了跨尺度通道注意力特征融合和通道空间感知的普适有效性。最终融合SCB、IBB和CSCAF-CSPM的完整模型在两类任务中取得最好性能,证实各模块联合共同优化了模型对神经退行性病变的表征能力。

为验证LMamba块的堆叠对模型性能的影响,本文固定其他模块的结构。并逐步增加LMamba块的数量,进而在ADNI测试集上评估。

实验结果如表5所示,随着LMamba块数量的增加,模型性能的提升效果逐级降低。实验数据表明,当LMamba块的数量从2增加至3时,模型处于性能提

升阶段,在AD分类任务中准确率提升了3.77个百分点,AUC提升了2.04个百分点,表明增加多尺度建模可显著增强长程依赖建模能力,提升性能表现。随着LMamba块的数量进一步增加,性能趋近饱和,AD分类任务准确率提升了2.97个百分点,AUC提升0.59个百分点,呈现出边际效益降低趋势。这一趋势在更具挑战的MCI转化预测任务中更加明显,当LMamba块数量从2增加至3时,准确率显著提升了9.38个百分点,AUC显著提升了5.19个百分点。当进一步提升LMamba块数量时,准确率提升了6.66个百分点,AUC提升了3.04个百分点,同样呈现边际效益降低趋势。这一现象表明,AD任务接近性能上限,MCI转换预测任务仍然保持稳定的性能增益,说明复杂病灶预测需更多尺度、更深层次的特征抽象。当LMamba块数为4时,准确率和AUC的标准差均有所降低,表明模型稳定性增强。

表5 LMamba块数量对性能影响结果  
Table 5 The impact of LMamba block count on model performance

Lmamba块数量	AD vs NC				sMCI vs pMCI			
	ACC (Mean±Std)	SEN (Mean±Std)	SPE (Mean±Std)	AUC (Mean±Std)	ACC (Mean±Std)	SEN (Mean±Std)	SPE (Mean±Std)	AUC (Mean±Std)
2	90.29±2.3	90.50±3.3	90.16±4.3	95.60±0.9	67.29±6.5	<b>84.52±9.9</b>	59.08±12.7	75.34±4.1
3	94.06±2.1	92.00±1.9	95.41±4.3	97.64±0.6	76.67±4.1	81.29±3.8	74.46±6.6	80.53±2.8
4	<b>97.03±2.0</b>	<b>97.49±2.5</b>	<b>96.72±3.2</b>	<b>98.23±0.5</b>	<b>83.33±2.3</b>	70.96±5.7	<b>89.23±5.9</b>	<b>83.57±0.8</b>

### 2.3.3 可解释性研究

为了更好地验证模型的学习效果,本节应用3D梯度加权类激活映射(3D Gradient-weighted Class Activation Mapping, 3D GCAM)实现可视化分析,如图7所示。

图7模型的可视化热图分析表明,其决策机制与AD影像学生物标志物的空间分布规律高度吻合,显著覆盖颞叶内侧(海马、内嗅皮层)、顶叶(楔前叶/后扣带回)、前额叶及默认模式网络(DMN)相关脑区。这些区域与淀粉样蛋白(A $\beta$ )沉积的高发区及葡萄糖

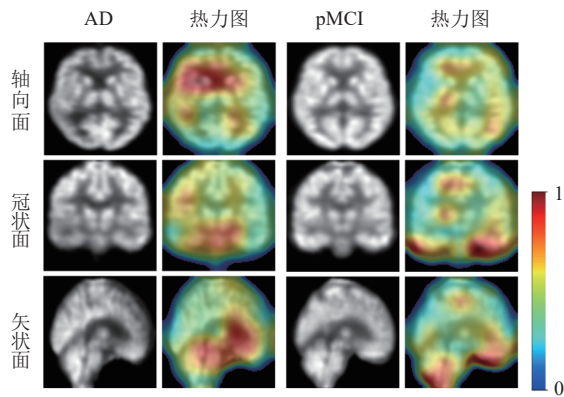


图 7 可视化注意力热力图  
Fig.7 Visual attention heatmap

代谢(18F-FDG)降低区高度重合。尤其对海马-顶叶轴代谢异常敏感,能够有效区分MCI亚型的转化风险,为AD早期诊断提供了可解释的影像学依据。

### 3 结论

本文提出了一种基于线性状态空间模型的三维PET影像分类框架,用于阿尔茨海默病(AD)的早期诊断与轻度认知障碍(MCI)转化预测。针对现有CNN和Transformer方法在长程依赖建模和计算效率上的局限性,本文创新性地设计了基于LMamba架构的多尺度特征融合模型。通过堆叠式LMamba块实现长序列依赖的动态建模,结合逐层跨尺度通道注意力融合模块(CSCAF)和通道空间感知模块(CSPM),有效融合全局上下文信息与局部空间特征,显著提升了病灶区域的定位精度与分类性能。在ADNI数据集上的实验结果表明,本文方法在AD分类任务中准确率达到97.03%,在MCI转化预测任务中准确率为83.33%,均优于现有方法。未来研究将进一步优化模型在多模态数据集上的泛化能力,结合其他生物标志物和基因数据多层次联合分析,为AD早期诊断提供更高效的解决方案。

#### 参考文献:

- [1] HEBERT L E, WEUVE J, SCHERR P A, *et al.* Alzheimer disease in the United States (2010-2050) estimated using the 2010 census[J]. *Neurology*, 2013, 80: 1778-1783.
- [2] TUAN T A, PHAM T B, KIM J Y, *et al.* Alzheimer's diagnosis using deep learning in segmenting and classifying 3D brain MR images[J]. *International Journal of Neuroscience*, 2022, 132(7): 689-698.
- [3] GRAY K R, WOLZ R, HECKEMANN R A, *et al.* Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease[J]. *NeuroImage*, 2012, 60(1): 221-229.
- [4] NOBILI F, ARBIZU J, BOUWMAN F, *et al.* European Association of Nuclear Medicine and European Academy of Neurology recommendations for the use of brain 18F-fluorodeoxyglucose positron emission tomography in neurodegenerative cognitive impairment and dementia: Delphi consensus[J]. *European Journal of Neurology*, 2018, 25(10): 1201-1217.
- [5] O'SHEA K, NASH R. An introduction to convolutional neural networks[EB/OL]. arXiv: 1511.08458v2(2015-12-02) [2025-01-20]. <https://doi.org/10.48550/arXiv.1511.08458>.
- [6] SAYED M, GIBSON J, WATSON J, *et al.* 3D reconstruction without 3D convolutions: U S 18/144, 099[P]. 2023-11-09.
- [7] RAO C, LIU Y. Three-dimensional convolutional neural network (3D-CNN) for heterogeneous material homogenization[J]. *Computational Materials Science*, 2020, 184: 109850.
- [8] 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.
- [9] 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.
- [10] JIA X, LIU Y, YANG Z, *et al.* Multi-modality self-attention aware deep network for 3D biomedical segmentation[J]. *BMC Medical Informatics and Decision Making*, 2020, 20: 1-7.
- [11] GU A, DAO T. Mamba: linear-time sequence modeling with selective state spaces[EB/OL]. arXiv: 2312.00752v2 (2024-05-31) [2025-01-20]. <https://doi.org/10.48550/arXiv.2312.00752>.
- [12] YUE Y, LI Z. Medmamba: vision mamba for medical image classification[EB/OL]. arXiv: 2403.03849(2024-03-06) [2025-01-20]. <https://doi.org/10.48550/arXiv.2403.03849>.
- [13] ZHU L, LIAO B, ZHANG Q, *et al.* Vision mamba: efficient visual representation learning with bidirectional state space model[EB/OL]. arXiv: 2401.09417(2024-01-17) [2025-01-20]. <https://doi.org/10.48550/arXiv.2401.09417>.
- [14] HAN D, WANG Z, XIA Z, *et al.* Demystify mamba in vision: a linear attention perspective[EB/OL]. arXiv: 2405.16605(2024-05-26) [2025-01-20]. <https://doi.org/10.48550/arXiv.2405.16605>.
- [15] KATHAROPOULOS A, VYAS A, PAPPAS N, *et al.* Transformers are rnns: fast autoregressive transformers with linear attention[C]//International conference on machine learning. Online: PMLR, 2020: 5156-5165.

(下转第95页)

