

DOI:10.13232/j.cnki.jnju.2026.01.005

基于空间与频率融合及图神经网络的 帕金森病 MRI 数据分类

胡国庆,施振全*,黄嘉爽,丁卫平
(南通大学人工智能与计算机学院,南通,226019)

摘要:帕金森病(Parkinson's Disease, PD)是一种常见的神经系统退行性疾病,其早期诊断对于延缓疾病进展至关重要。磁共振成像(Magnetic Resonance Imaging, MRI)凭借无创性和高分辨率的特点,在对 PD 的诊断中得到了广泛应用。然而,现有方法往往仅依赖单一域信息,存在信息建模不足的问题,同时,PD 的病理变化不是孤立的,现有方法对图像块之间的区域关联性往往考虑不多,忽略了脑区之间的相互影响。因此,提出一种融合空间与频率信息的双分支深度学习框架,空间分支采用视觉 Transformer 来捕获 MRI 图像中的全局空间关系,频域分支则利用全局滤波网络来提取频域特征,通过高斯加权欧氏距离来构建邻接矩阵,并引入图卷积网络来建模图像块间的拓扑关系。在模型训练中,选取轴向 2D 切片并结合 ImageNet 的预训练权重进行迁移学习,通过多数投票策略比较单个受试者的多切片结果,最终生成受试者级别的分类结果。在包含患者与健康对照的 PD 数据集上进行评估,实验结果显示,所提方法的准确率、特异性和 F1-score 等指标均优于多个现有方法,证实其在临床应用中的有效性。

关键词:帕金森病,磁共振成像,深度学习,图卷积网络,Transformer

中图分类号:R742.5, TP183, TP391.72

文献标志码:A

Parkinson's disease MRI data classification based on spatial-frequency fusion and graph neural network

Hu Guoqing, Shi Zhenquan*, Huang Jiashuang, Ding Weiping
(School of Artificial Intelligence and Computer Science, Nantong University, Nantong, 226019, China)

Abstract: Parkinson's disease (PD) is a common neurodegenerative disorder, and early diagnosis is crucial for slowing disease progression. Magnetic resonance imaging (MRI) has been widely used in the diagnosis of PD due to its non-invasive nature and high-resolution capabilities. However, existing methods often rely on information from a single domain, resulting in insufficient information modeling. Furthermore, the pathological changes in PD are not isolated, existing approaches frequently fail to account for regional correlations between image patches, thereby neglecting the functional interactions among brain regions. To address these limitations, we propose a two-branch deep learning framework that integrates spatial and frequency-domain information. The spatial branch employs a Vision Transformer to capture global spatial relationships in MRI images, while the frequency branch utilizes GFNet (Global Filter Network) to extract frequency-domain features. An adjacency matrix is constructed using Gaussian-weighted Euclidean distance, and a graph convolutional network (GCN) is introduced to model the topological relationships between image patches. During model training, axial 2D slices are selected and fine-tuned using pre-trained weights from ImageNet through transfer learning. A majority voting strategy is then applied to aggregate predictions from multiple slices of a single subject to produce a subject-level classification result. The proposed

基金项目:国家自然科学基金(62176131)

收稿日期:2025-08-22

* 通信联系人, E-mail: szq@ntu.edu.cn

method was evaluated on a PD dataset comprising both patients and healthy controls. Experimental results demonstrate that our approach outperforms several state-of-the-art methods in terms of key metrics including accuracy, specificity, and $F1$ -score, thereby confirming its potential for effective clinical application.

Keywords: Parkinson's disease, MRI, deep learning, graph convolutional network, Transformer

帕金森病(Parkinson's Disease, PD)是一种渐进性神经退行性疾病,主要影响中枢神经系统^[1],主要是由于黑质区多巴胺神经元的退化,从而影响运动功能。由于缺乏可靠的生物标志物,且疾病前期脑部结构的改变较为细微,PD的早期精准诊断仍是临床面临的重大挑战^[2]。近年来,结构磁共振成像(Structural Magnetic Resonance Imaging, sMRI)已成为一种极具价值的无创成像手段,能够揭示PD患者与健康对照者之间的神经解剖学差异,为计算机辅助诊断及疾病监测提供了可能^[3]。

早期基于sMRI的PD分类方法,通常先通过感兴趣区域分析或基于体素的形态学分析(Voxel-Based Morphometry, VBM)提取手工设计的形态学特征(如皮质下体积和皮质厚度),并结合支持向量机、随机森林等传统机器学习分类器^[4-5]来进行分类。这类方法虽然能达到中等的诊断准确率,但有明显的局限,一方面依赖人工设计的特征表示,另一方面对高维特征的处理能力有限,过多无关特征会干扰分类决策,导致“维度灾难”。

随着深度学习的兴起,卷积神经网络(Convolutional Neural Network, CNN)成为从MRI中自动提取特征的主流框架,并在神经系统疾病辅助诊断领域取得了显著进展^[6]。例如, Sivaranjini and Sujatha^[7]采用 AlexNet,通过迁移学习对PPMI数据库中的MRI图像进行PD检测。Zhang et al^[8]开发了基于ResNet的框架,以灰质sMRI切片为输入,将特定区域特征提取与注意力机制相结合,增强灰质特征的表征能力。然而,由于固有的局部感受野限制, CNN在建模大脑区域之间的全局关系方面存在不足^[9]。此外, CNN通常仅在空间域进行建模,往往忽略磁共振成像采集过程中固有的频率信息,可能遗漏具有互补性的判别线索。

近年来,深度学习领域涌现出众多基于

Transformer架构的方法,如视觉Transformer(Vision Transformer, ViT)^[10]。ViT通过自注意力机制可有效捕捉远距离特征间的联系,提升全脑范围内的表示能力。Swin Transformer^[11]通过窗口滑动机制引入局部感受野,并构建层次化结构以实现局部到全局的特征聚合。DeiT^[12]通过引入知识蒸馏、简化位置嵌入、优化训练策略的方式,使其在小规模数据集上仍能取得良好的性能。然而,纯Transformer模型在局部细节特征提取方面比CNN弱,且对大规模数据依赖较高,训练稳定性较差。同时,PD的病理变化不是孤立的,现有方法对图像块之间的区域关联性往往考虑不多,忽略了脑区之间的相互影响。因此,在识别早期PD患者时仍面临以下关键挑战:一是从复杂结构中有效提取局部细微病变特征;二是在全脑范围内建模多区域之间的空间拓扑关系;三是充分融合空间和频率信息,以提升对病理区域的敏感性和诊断的准确性。

针对以上问题,本文提出了一种融合空间与频域特征的双分支模型PDFormer,结合空间与频率双域特征、局部与全局联合建模以及图神经网络融合机制,显著提升了模型对复杂脑部结构异常的识别能力。

本文的具体贡献如下。

(1)联合建模空间域与频率域特征:采用视觉Transformer提取MRI图像中的全局空间特征,并结合全局滤波网络捕捉频率域中的局部高频细节,通过特征拼接获得融合表示,有效整合多尺度信息。

(2)基于欧氏距离构建拓扑邻接图:提出一种高斯加权的patch中心距离图建模方法,用于刻画MRI图像中patch之间的空间关系,增强结构信息表达能力。

(3)引入图卷积神经网络(Graph Convolution Network, GCN)进行高阶特征融合:利用

GCN 对 patch 间拓扑结构进行建模,提升模型对复杂结构变化的感知能力,并通过切片级多数投票策略提升诊断稳定性。

1 相关工作

1.1 基于传统机器学习的 PD 分类方法 传统的基于机器学习的方法多是分两步进行^[13],即先手动或者自动地从影像数据中提取与疾病相关的特征或者感兴趣区域,然后将提取到的内容送入机器学习模型进行分类。例如,Abós et al^[14]基于 Brainnetome 脑图谱从静息态 fMRI (Functional Magnetic Resonance Imaging) 重建功能连接组,利用逻辑回归进行特征选择,结合支持向量机进行训练和分类。Anita^[15]提出一种基于三维 SPECT (Single-Photon Emission Computed Tomography) 的早期 PD 诊断方法,从三组不同的体绘制图像 (Volume Rendering Image, VRI) 切片中提取形状特征和表面拟合特征,使用遗传算法从提取的特征中选择最优特征集,并使用径向基函数 ELM (Extreme Learning Machine) 分类器和径向基函数支持向量机分类器进行分类。Babu et al^[16]提出一种带有递归特征消除的元认知径向基函数网络 (McRBFN),使用基于体素的形态测量特征并采用基于投影的学习算法。然而,由于 PD 相关脑区的结构复杂,手动定义最优特征集往往十分困难。再加上医学影像数据的高维特性,限制了传统机器学习模型的有效性。

1.2 基于深度学习的空间域与频域建模 深度学习利用受人脑启发的神经网络结构来实现特征学习与数据分析^[17],各种深度学习网络框架也已经被开发出来,并广泛应用于计算机视觉、语音识别与数据挖掘等领域。

大多数基于 MRI 的 PD 预测研究使用 CNN 进行预测。例如,Mary and Suganthi^[18]用 MFEA (Multifactorial Evolutionary Algorithm) 提取多种特征,再利用主成分分析进行特征降维,将处理后的特征输入带有 YOLO 应用风格的 DNetCNN 模型进行 PD 的分类。Chakraborty et al^[19]构建了一个 3D CNN,用于从整个脑部 MRI 图像中学习复杂的三维结构模式,以区分 PD 病例与健康对照。Zhu^[20]提出一种混合深度学习的模型,结合来自

MRI 脑部扫描和患者特征(症状数据)的多模态信息将患者划分为五个阶段,实现更细致的严重程度评估。

与此同时,基于 Transformer 的架构也作为一种新的范式出现,利用自注意力机制捕捉全局上下文关系。最初,该类方法应用于自然语言处理任务,在 BERT^[21]与 GPT-3^[22]等模型中凭借强大的特征表征能力取得了显著成功。受此启发,研究者将 Transformer 编码器拓展至计算机视觉领域,提出如 ViT 等具有影响力的模型用于图像分类。与空间建模互补的是,频域建模近年来也成为一种有效方法,用于挖掘 MRI 中固有的频谱信息。例如,GFNet^[23]通过快速傅里叶变换构建全局滤波器,从而增强频域表征能力。

1.3 基于图神经网络的拓扑特征学习 近年来,图神经网络 (Graph Neural Networks, GNNs) 因其能够建模非欧几里得结构并捕捉区域或图像块之间的拓扑关系,而逐渐受到关注。随着图卷积网络的发展,研究者提出多种复杂的消息传递机制。例如,Atwood and Towsley^[24]提出的扩散卷积神经网络 (DCNN) 通过在图上定义基于随机游走的扩散过程来实现卷积操作,从而在图结构数据中有效捕捉局部与多步邻域信息。Tran et al^[25]引入基于最短路径距离的多阶卷积方法,将不同距离的邻居信息按不同权重聚合,让模型在训练中学习“远邻居”的贡献。Gilmer et al^[26]提出消息传递神经网络 (Message Passing Neural Network, MPNN),通过迭代的消息传递与节点更新机制来学习图结构表示。这些方法在分子建模、社交网络等任务中已展现出优越的性能。

在结构性 MRI 分析中,基于图的方法的独特优势在于其能够显式编码空间关系,而这类关系在 CNN 与 Transformer 架构中通常是隐式捕捉的。同时,虽然 GCN 在拓扑特征学习方面发展迅速,但与图像分类任务的结合研究仍较为有限,这也为本研究提供了新的切入点。

2 提出的方法

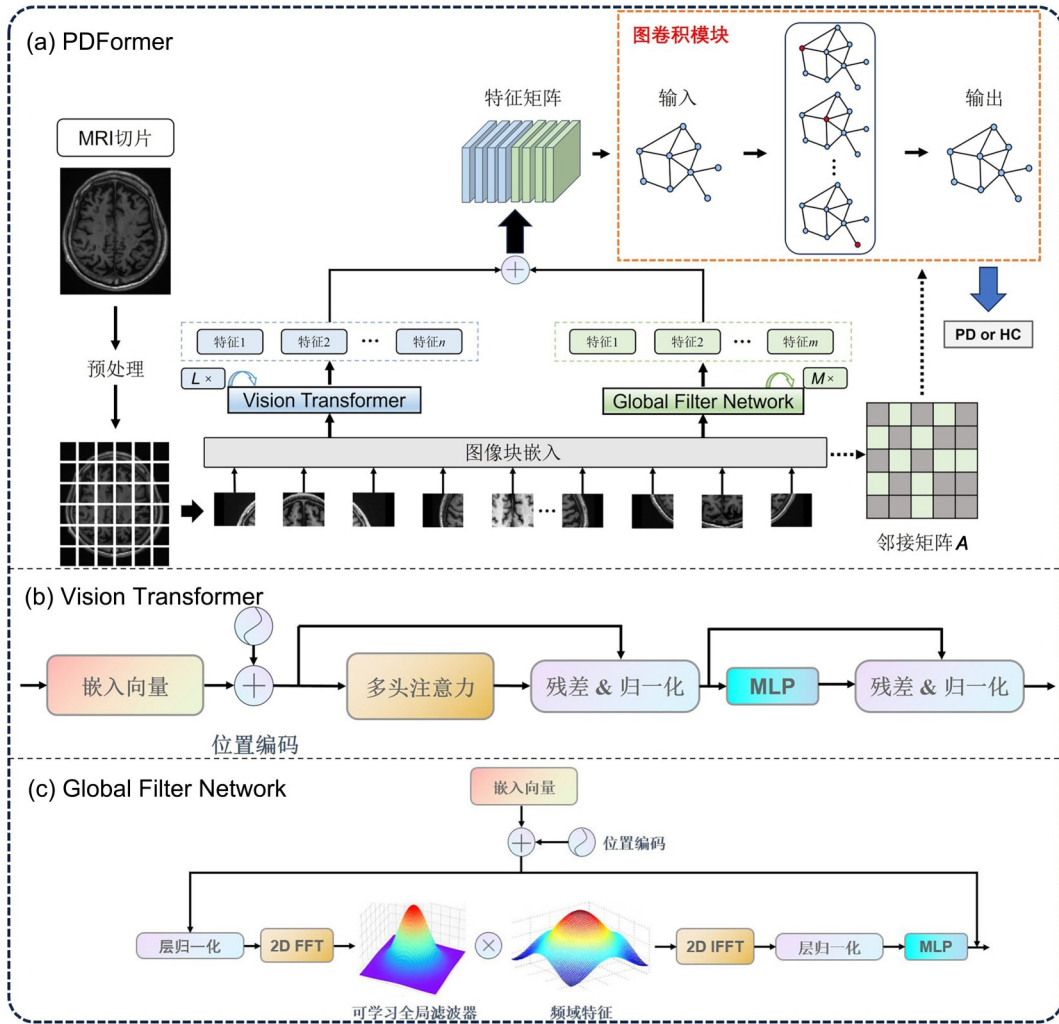
2.1 PDFormer 框架概述 考虑到 PD 相关的脑部变化在早期阶段较微弱且具有异质性,因此,仅依赖单一空间域或频域特征的建模方法往往无法

充分刻画病变区域的信息. 将PD图像分类视为多层次特征融合与拓扑关系建模问题, 提出PDFormer模型. 图1为提出的模型的框架示意图.

具体地, 所提模型包含四个关键步骤. 首先, 选择关键轴向切片以构建输入数据. 其次, 设计一个双域特征提取框架, 其中一个分支用于处理空间域表征, 另一个分支则用于提取频域特征. 随后, 构建图拓扑结构, 其中节点表示MRI图像块, 边则基于欧式距离编码空间邻接关系, 利用GCN实现高阶特征融合与增强特征表示. 最后, 模型对每个轴向切片生成初步的分类结果, 并通过多数投票策略整合多个切片的预测结果, 以获得受试者级别的PD诊断决策.

2.2 切片选择 在进行经验分析和多次实验(详见4.1)后, 在Z轴方向选取轴向平面索引位于76~90序列的连续15张切片来训练网络, 最终获得2190张PD患者的轴向切片和2325张健康对照(Healthy Control, HC)个体的轴向切片, 每张切片均为 182×218 的灰度图像. 对切片后的数据进行预处理, 使用数据增强技术, 包括随机旋转($-20^\circ \sim 20^\circ$)和随机翻转, 以此来增加数据多样性. 将图像尺寸调整为 224×224 , 并对像素值归一化到 $[0, 1]$.

2.3 空间表征学习 本研究中ViT是左侧主干网络, 其结构类似标准ViT模型, 由多层Transformer编码器组成, 每层由多头自注意力和多层



(a) The overall picture of the proposed model, (b) the details of Vision Transformer encoder, (c) the structure of Global Filter Network

图1 PDFormer框架的示意图

Fig. 1 Illustration of the proposed PDFormer framework

感知机交替堆叠而成. 在每个 Transformer 层之前都会进行层归一化, 并在每个模块后加入残差连接以优化梯度流动.

对输入的 2D MRI 切片 $x \in R^{H \times W \times C}$ (空间分辨率为 $H \times W$ 且包含 C 个通道) 进行分块, 将其划分为多个 $P \times P$ 的 patch, $x_p \in R^{N \times (P^2 \cdot C)}$, $P \times P$ 表示每个 patch 的大小, $N = HW/P^2$ 表示总的 patch 个数, 也是 Transformer 的输入序列长度. 每个 patch 被展平后通过可训练的线性投影矩阵 $W_p \in R^{(P^2 \cdot C) \times D}$ 映射到固定维度 D , 形成 Transformer 的输入序列, 具体计算如下:

$$z_0 = x_p W_p + E_{\text{pos}} \quad (1)$$

其中, $E_{\text{pos}} \in R^{N \times D}$ 为可学习的位置编码, 用于弥补 Transformer 结构中缺乏空间位置信息的不足. ViT 的 Transformer 编码器计算如下:

$$z'_l = \text{MSA}(\text{LN}(z_{l-1})) + z_{l-1} \quad (2)$$

$$l = 1, \dots, L$$

$$z_l = \text{MLP}(\text{LN}(z'_l)) + z'_l \quad (3)$$

$$l = 1, \dots, L$$

其中, L 为 Transformer 的层数. 本研究使用 12 层 Transformer 编码器.

2.4 频域感知特征编码 在 MRI 过程中, 数据首先在频域中采集, 然后经过傅里叶逆变换转换到空间域供临床观察, 因此, 利用频域信息来辅助分类任务具有重要的医学意义. 然而, 直接在高维频域数据上进行深度学习计算代价较高, 特别是在处理高分辨率影像时, 为此, 采用 GFNet 进行高效的频域特征提取. GFNet 的核心思想是用快速傅里叶变换替代自注意力机制, 以降低计算复杂度.

在 PDFormer 右侧分支中对输入的 2D MRI 切片进行非重叠 patch 划分. 与 ViT 的处理方式一致, 将每个 patch 展平为 $L = H \times W$ 个维度为 D 的 tokens 序列, 每个空间域 token $x \in R^{H \times W \times D}$ 由 2D FFT 转换为频域中的复杂张量 $X \in R^{N \times D}$:

$$X = 2D \text{FFT}[x] \quad (4)$$

其次, 引入可训练的全局滤波器 K 来对 FFT 变换后的特征进行调制, 以增强对高频和低频信息的建模能力:

$$\tilde{X} = X \odot K \quad (5)$$

其中, \odot 表示逐元素乘法; K 是一个全局滤波器, 其维度与 X 相同, 并通过网络训练进行优化.

最后, 经过频域调制后的特征仍然位于频域, 因此需要通过二维傅里叶逆变换 (IFFT) 将其转换回空间域, 以便进行后续特征融合和分类:

$$x' = \text{IFFT}(\tilde{X}) \quad (6)$$

其中, x' 表示逆变换后的特征, 它既包含原始 MRI 影像的空间信息, 也结合了频域增强后的特征表示. GFNet 采用全局平均池化代替 ViT 结构中的分类 token, 以减少参数数量, 同时保留重要特征信息. 本研究使用 GFNet-B 变体, 其包含 19 层网络结构, 嵌入维度为 512.

2.5 基于图的拓扑建模 对于每个图像块 p_i ($i \in 1, 2, \dots, N$), 空间特征 $f_i^v \in R^{D_s}$ 由 ViT 编码器提取, 频域特征 $f_i^s \in R^{D_s}$ 通过 GFNet 编码器在执行二维傅里叶变换后获得. 最终的图像块级表征 $f_i \in R^{D_s + D_s}$ 通过拼接空间特征和频域特征得到:

$$f_i = \text{Concat}(f_i^v, f_i^s) \quad (7)$$

为了建模图像块之间的拓扑结构, 构建一个无向图 $G = (V, A)$, 其中, 每个节点 $v_i \in V$ 对应一个图像块, 邻接矩阵 $A \in R^{N \times N}$ 根据图像块中心之间的欧几里得距离计算得到, 并通过高斯核进行调制. 具体地, 首先计算每个图像块的几何中心坐标. 给定一个 $H \times W$ 的图像和 $P \times P$ 的图像块, 图像被划分为 $h = \lfloor H/P \rfloor$ 行和 $w = \lfloor W/P \rfloor$ 列, 总计 $N = h \times w$ 个图像块. 位于第 i 行第 j 列的图像块的中心坐标 (c_h, c_w) 计算如下:

$$c_h = i \cdot P + \frac{P}{2} \quad i \in \{0, 1, \dots, h-1\} \quad (8)$$

$$c_w = j \cdot P + \frac{P}{2} \quad j \in \{0, 1, \dots, w-1\}$$

对于任意两个图像块 $p_m = (c_{h_m}, c_{w_m})$ 和 $p_n = (c_{h_n}, c_{w_n})$, 其中心点之间的欧几里得距离 d_{mn} 为:

$$d_{mn} = \sqrt{(c_{h_m} - c_{h_n})^2 + (c_{w_m} - c_{w_n})^2} \quad (9)$$

图像块 m 与 n 之间的连接强度通过高斯核定义为:

$$w_{mn} = \exp\left(-\frac{d_{mn}^2}{2\sigma^2}\right) \quad (10)$$

这里的高斯核是基于欧几里得距离的高斯权重函数, 用来衡量两个图像块中心之间的连接强

度,距离越近权重越大,距离越远权重越小, σ 是控制衰减速率的超参数.遍历所有块之间的中心点组合,将权重存入邻接矩阵 A .

邻接矩阵 A 的定义为:

$$A[m,n]=A[n,m]=\begin{cases} w_{mn}, & m \neq n \\ 1, & m = n \end{cases} \quad (11)$$

将拼接后的图像块表征 $X=[f_1, f_2, \dots, f_N] \in R^{N \times (D_s + D_e)}$ 与邻接矩阵 A 一同输入GCN网络,以执行基于拓扑的特征聚合.GCN的运算定义为:

$$H^{(l+1)} = \sigma(\tilde{A}H^{(l)}W^{(l)}) \quad (12)$$

其中, $H^{(l)}$ 是第 l 层的输入特征矩阵($H^{(0)}=X$ 为初始输入特征矩阵), $W^{(l)}$ 是第 l 层的可训练权重矩阵, \tilde{A} 为归一化后的邻接矩阵, σ 为激活函数.输出 $H^{(l+1)}$ 为经过信息传递后更新的图像块级表征.

2.6 分类与多数投票策略 在通过空间-频域融合与基于图的拓扑建模获得判别性表征后,每张2D MRI切片将被单独分类.具体地,对图卷积模块输出的特征进行全局平均池化,以生成每张切片的切片级特征向量.随后,该向量通过一个全连接层,并接入 $Softmax$ 激活函数,以预测该切片在两个类别上的概率分布.

对于每位受试者,模型独立处理其多张轴位切片,生成切片级的分类预测.为了获得受试者级的诊断结果,采用多数投票策略,即将所有切片预测结果中出现频率最高的类别作为最终的预测:

$$y_{\text{final}} = \text{mode}(\{y_1, y_2, \dots, y_S\}) \quad (13)$$

其中, $y_i \in \{0, 1\}$ 表示第 i 张切片的预测标签, S 为该受试者的切片总数. mode 函数返回在所有切片预测中出现次数最多的类别标签.

3 实验结果

3.1 数据集与预处理 采用的数据集涵盖了患者组和健康对照组,共301例受试者,其中,PD组146例,HC组155例.其中,PD受试者:(1)符合PD的诊断标准;(2)接受MRI扫描;(3)未接受过脑部手术;(4)没有颅内占位性病变、头部外伤史或任何其他影响大脑重要结构的疾病.排除标准:(1)MRI检查禁忌证;(2)排除运动伪影和金

属伪影.HC组无任何已知的神经系统疾病或认知障碍.

所有影像数据均使用配备24个通道头线圈的3-T磁共振成像扫描仪(GE Signa)采集,T1加权成像的扫描参数如表1所示.扫描图像经过筛选后分别归入相应研究组.

表1 采集的T1加权扫描图像规格

Table 1 Specifications of T1 weighted scanned images collected

影像参数	值
尺寸	256×256
切片厚度	1.0 mm
空间	1.0 mm×1.0 mm×1.0 mm
平面	冠状面

由于原始扫描图像之间存在空间差异,需要对图像进行预处理,以便将所有扫描图像统一到同一空间.使用蒙特利尔神经研究所(Montreal Neurological Institute)创建的MNI152_T1_1mm模板,借助FSL工具(<https://fsl.fmrib.ox.ac.uk>)完成线性配准.线性配准后,不同扫描对象的头颅大小和空间朝向得以统一.图2展示了配准前后的MRI扫描结果.

本研究中,为了标准化模型输入并降低计算复杂度,从每例配准后的三维MRI中,沿轴状面方向提取15张连续切片(切片索引76至90),每张图像的分辨率为182×218像素.这些二维切

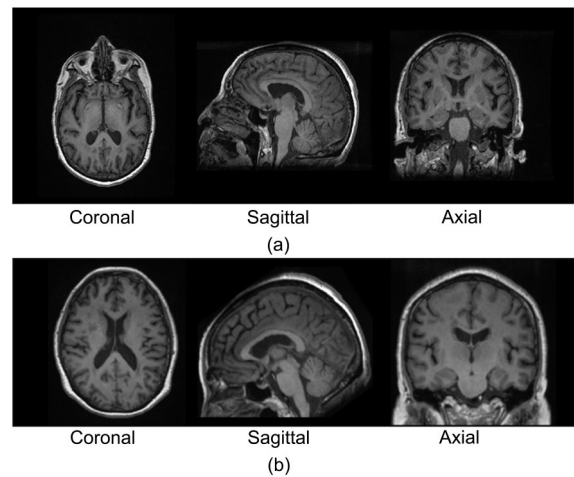


图2 同一患者在配准前后的MRI扫描结果:(a)配准前;(b)配准后

Fig. 2 MRI scan of the same patient before (a) and after (b) registration

片被用于后续模型的训练与分类,所有数据均以 NIfTI 格式(.nii)存储。

3.2 实验设置 采用十折交叉验证来对数据集进行训练与测试,以确保实验结果的稳定性和可靠性,并获得更具泛化能力的评估指标。此外,为了进一步确保实验结果的可靠性,通过计算十折的平均性能来报告最终的分类结果。采用迁移学习策略,具体地,ViT 与 GFNet 的参数初始化均来自 ImageNet 数据集上预训练的权重。

所有实验均在单个 NVIDIA RTX 4090 GPU 上进行,使用的深度学习框架为 PyTorch 2.0.1。采用 SGD 优化器对所有网络进行训练,并设置动量值为 0.9,以加速收敛并减少梯度更新过程中的振荡。初始学习率设为 0.0001,每批次数据的大小为 16,总训练 epoch 数为 150。以上参数配置均适用于两种 Transformer 网络。表 2 简要总结了提出的框架的网络参数细节。

表 2 本文提出的框架的网络参数信息摘要

Table 2 Summary of the proposed network parameter

参数	左分支/空间域	右分支/频率域
图像尺寸	224×224	224×224
patch 大小	16×16	16×16
层数/深度	$L = 12$	$M = 19$
嵌入维度	$n = 768$	$m = 512$
激活函数	<i>GeLU</i>	<i>GeLU</i>
MLP 尺寸	3072	2048
dropout 率	0.1	0.25

3.3 评价指标 为了全面评估所提方法的分类性能,采用准确率、灵敏度、特异性和 F1-score 来进行评估。这些指标通过四个值进行计算:真正例 (TP)、真负例 (TN)、假正例 (FP) 和假负例 (FN)。

准确率 (Accuracy, ACC) 表示分类模型正确识别的样本占总样本数的比例:

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

灵敏度 (Sensitivity, SEN), 也称召回率 (Recall), 表示模型正确识别患病患者的能力:

$$SEN = \frac{TP}{TP + FN} \quad (15)$$

特异性 (Specificity, SPE), 也称为真负率

(True Negative Rate), 是衡量模型正确识别健康个体的能力, 可表示为:

$$SPE = \frac{TN}{TN + FP} \quad (16)$$

F1-score 是精确度和召回率的调和平均数, 综合考虑了精确度与召回率的平衡:

$$F1\text{-score} = \frac{2TP}{2TP + FP + FN} \quad (17)$$

3.4 对比实验 为了验证所提模型的有效性, 将其与五个常见的深度学习模型进行对比, 包括 DenseNet (2D)^[27], MobileNetV2 (2D)^[28], ResNet (2D)^[29], ResNet (3D) 和 SFCN (3D)。对于 2D CNN 模型, 步骤包括输入切片、数据增强、采用多数投票策略等。对于 3D 网络架构, 输入为完整的 3D 大脑 MRI 数据, 以进行分类结果预测。

图 3 展示了不同方法在 PD 数据集上分类的性能对比, 图中的误差棒表示每个指标的标准差, 反映了模型性能的波动范围。PDFormer 的误差棒较短, 表明其性能更稳定。相比之下, 某些方法的误差棒较长, 表明其性能波动较大, 可能受到数据分布或模型复杂度的影响。

为了提供更详细的对比分析, 表 3 列出了不同模型在各个性能指标上的平均值和标准差, 表中黑体字表示最优的性能。

3.5 消融实验 为了探究所提模型各核心组件对最终性能的具体贡献, 进行了五组消融实验。在每个实验中去除模型的一个关键部分并评估其对性能的影响。表 4 展示了在 PD 数据集第四折上的消融实验结果, 作为十折交叉验证实验中的代表性结果, 表中黑体字表示最优的性能。

由表可见, 仅使用 ViT 提取空间域特征时, 模型在负样本识别 ($SPE = 0.700$) 上存在一定的改进空间。仅使用 GFNet 提取频域特征时, 模型能有效识别健康对照, 但对 PD 患者的识别效果较差。去除多数投票机制后的模型仍然使用 ViT, GFNet 和 GCN 层进行特征提取和融合, 此时模型的 $ACC = 80\%$, $SEN = 93.3\%$, 说明去除多数投票机制后, 模型对单一切片的分类性能有所提升, 但整体分类效果 (SPE 和 $F1\text{-score}$) 有所下降。去除 GCN 层后, 模型虽然仍使用 ViT 和 GFNet 进行特征提取, 但不再对提取的特征进行

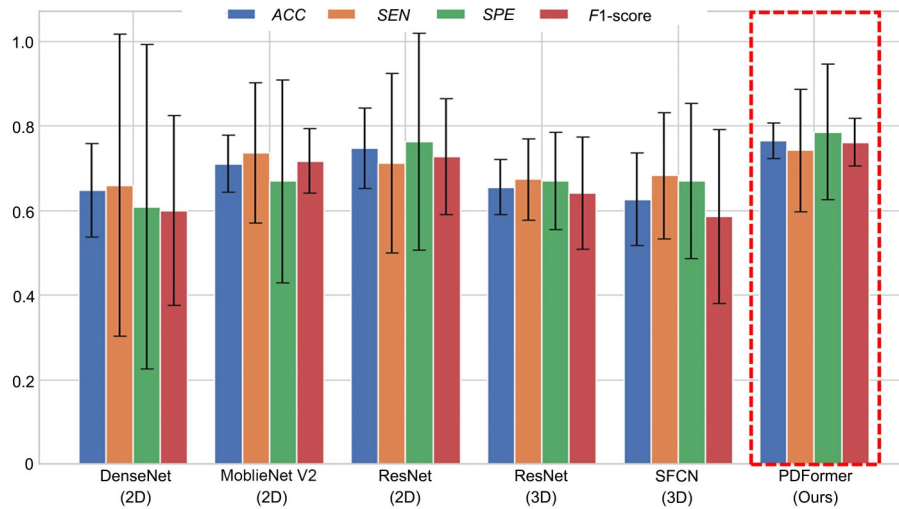


图3 不同方法在PD数据集上分类的性能对比

Fig. 3 Classification performance of different methods on PD dataset

表3 不同方法在PD数据集上分类的性能对比

Table 3 Classification performance of different methods on PD dataset

方法	ACC	SEN	SPE	F1-score
DenseNet (2D)	0.648±0.111	0.660±0.357	0.609±0.384	0.600±0.225
MobileNetV2 (2D)	0.711±0.067	0.737±0.166	0.670±0.240	0.718±0.076
ResNet (2D)	0.747±0.095	0.713±0.213	0.763±0.257	0.728±0.137
ResNet (3D)	0.656±0.066	0.674±0.097	0.670±0.115	0.641±0.133
SFCN (3D)	0.627±0.110	0.683±0.149	0.670±0.184	0.587±0.206
PDFormer (Ours)	0.766±0.042	0.743±0.145	0.786±0.160	0.762±0.057

表4 PD数据集第四折消融实验结果

Table 4 Ablation studies on Fold-4 of the PD dataset

方法	ACC	SEN	SPE	F1-score
w/o majority voting	0.800	0.933	0.576	0.792
w/o GCN Fusion	0.767	0.867	0.667	0.788
ViT only	0.780	0.850	0.700	0.800
GFNet only	0.733	0.533	0.933	0.667
PDFormer (Ours)	0.833	0.733	0.933	0.815

高阶关系建模,此时模型的 $ACC = 76.67\%$,和PDFormer相比,性能有所下降.

综合来看,完整的PDFormer表现最优.

4 讨论

4.1 切片选择的影响 完成配准及尺寸归一化等标准预处理操作后,将所有MRI图像体积统一调整为固定大小,从处理后的体积图像中提取 224×224 的轴向2D切片.考虑到MRI体积的首

尾部分通常包含的结构信息较少,注意力集中在体积图像的中部区域,以获得更具判别力的特征.

首先以体积中间位置为中心,提取连续15张切片进行训练,分析其分类性能.随后,逐步扩大切片跨度,分别构建30张和45张连续切片的输入组合,并对三种切片数量对应的分类性能进行比较,结果如图4所示.

实验结果表明,在使用连续15张切片时(图4a),模型的 ACC 和 $F1$ -score已经表现良好,并能较稳定地区分样本差异.但进一步将跨度扩大至30或45张时,模型的性能没有得到显著的提升.如图4b和图4c所示,虽然部分切片组合的 ACC 或 $F1$ -score略有浮动,但整体趋势趋于平稳,甚至在某些范围内还略有下降.

此外,切片数量的增加也显著增加了模型训练的计算开销.在当前的GPU环境下,15张切片的训练时长约为4h,30张切片为7h,45张切片则

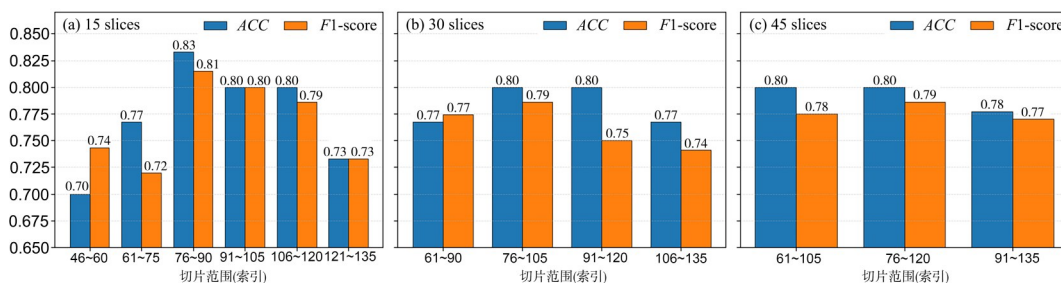


图 4 不同切片跨度的性能比较

Fig. 4 Model performance of different slice spans

超过 10 h. 因此,为了在保证模型性能的同时兼顾训练效率,最终选择切片范围为第 76~90 张的 15 张切片组合作为模型输入. 在该设置下,模型取得了最佳的分类表现,实现了性能与计算成本之间的良好权衡.

4.2 基于 Grad-CAM 的敏感区域可视化 为了更深入地理解所提模型在区分 PD 与 HC 时的决策过程,在测试集上进行了基于 Grad-CAM 的可解释性分析. 测试集包含 30 例 MRI 扫描(PD 患者 15 例,HC 15 例). 对于每个 3D 数据,选取索引为 76~90 的轴向切片,共获得 450 张用于可视化的 2D 图像.

基于训练好的 PDFormer 模型,将 Grad-CAM 应用于 ViT 和 GFNet 分支所提取的深层特

征图,为每一张切片生成注意力图. 这些注意力图能突出地显示模型在判别过程中所依赖的关键区域,图 5 展示了来自 PD (图 5a) 与 HC (图 5b) 两类受试者的代表性样例,每幅图被划分为两组(每类各包含两个样本). 在每个可视化样例中,上排显示的是 ViT 分支的 Grad-CAM 结果,下排对应 GFNet 分支的结果,可以直接比较两个子网络的空间关注差异. 图中红色区域表示模型关注度高的敏感区域. 由图可见,ViT 分支的 Grad-CAM 热力图较分散,模型在多个区域均给予一定的权重. GFNet 分支的 Grad-CAM 热力图更集中. 两个分支在同一样本上的关注区域存在一定互补性,证明双分支结构在全局-局部特征建模上的合理性. 这种差异性与互补性也为后续的特征

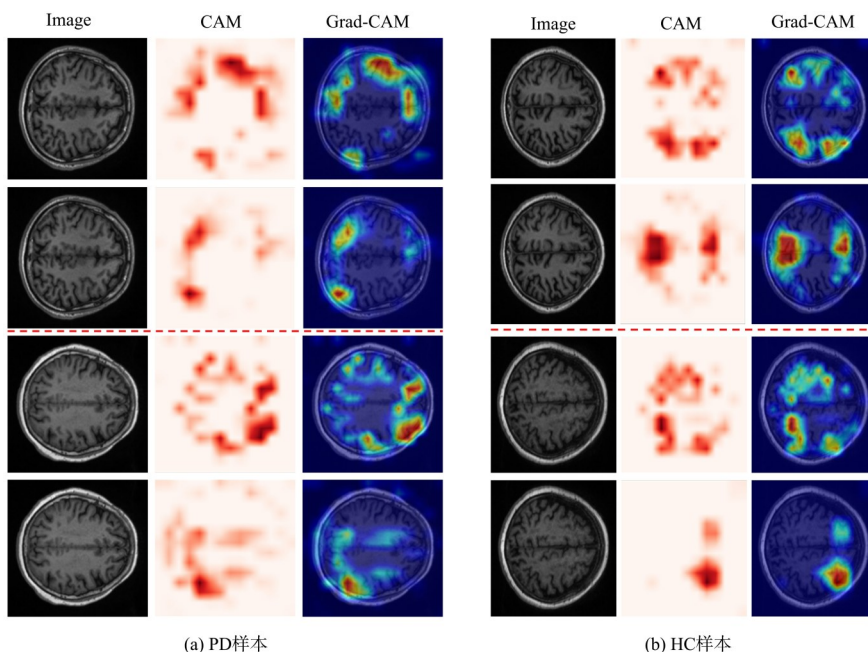


图 5 PD 样本(a)和 HC 样本(b)的 Grad-CAM 可视化

Fig. 5 Grad-CAM visualization of PD (a) and HC (b) samples

融合提供了理论支持,有助于提升整体分类性能.

5 结论

本文提出的PDFormer双分支网络模型结合了ViT与GFNet,联合建模MRI图像中的空间与频率特征,同时引入GCN建模图像块间的空间拓扑结构,在PD分类任务中取得了优于现有方法的性能表现,提高了其鲁棒性和预测能力.

未来将尝试引入多模态影像信息,如R2*, FLAIR, DTI等多模态影像,以获得更全面的特征表达,进一步提高分类性能.

参考文献

- [1] Bidesi N S R, Vang Andersen I, Windhorst A D, et al. The role of neuroimaging in Parkinson's disease. *Journal of Neurochemistry*, 2021, 159(4):660—689.
- [2] Qu Y, Li J, Qin Q, et al. A systematic review and meta - analysis of inflammatory biomarkers in Parkinson's disease. *npj Parkinson's Disease*, 2023, 9(1):18.
- [3] Khanna K, Gambhir S, Gambhir M. A novel technique for classifying Parkinson's disease using structural MRI scans. *Multimedia Tools and Applications*, 2023, 82(29):46011—46036.
- [4] Haq A U, Li J P, Memon M H, et al. Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings. *IEEE Access*, 2019 (7):37718—37734.
- [5] Xu S, Pan Z. A novel ensemble of random forest for assisting diagnosis of Parkinson's disease on small handwritten dynamics dataset. *International Journal of Medical Informatics*, 2020, 144:104283.
- [6] Litjens G, Kooi T, Bejnordi B E, et al. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 2017, 42:60—88.
- [7] Sivaranjini S, Sujatha C M. Deep learning based diagnosis of Parkinson's disease using convolutional neural network. *Multimedia Tools and Applications*, 2020, 79(21):15467—15479.
- [8] Zhang Y, Teng Q, Liu Y, et al. Diagnosis of Alzheimer's disease based on regional attention with sMRI gray matter slices. *Journal of Neuroscience Methods*, 2022, 365:109376.
- [9] Luo W, Li Y, Urtasun R, et al. Understanding the effective receptive field in deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 2016:29.
- [10] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: Transformers for image recognition at scale. 2020, arXiv:2010.11929.
- [11] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Montreal, Canada:IEEE, 2021:10012—10022.
- [12] Touvron H, Cord M, Douze M, et al. Training data-efficient image transformers & distillation through attention//*International Conference on Machine Learning*. Online:PMLR, 2021:10347—10357.
- [13] 李西,姜孟. 机器学习在帕金森病诊断中的应用研究. *电子科技大学学报*, 2024, 53(2):315—320.
- [14] Abós A, Baggio H C, Segura B, et al. Discriminating cognitive status in Parkinson's disease through functional connectomics and machine learning. *Scientific Reports*, 2017, 7(1):45347.
- [15] Anita S. Improved classification accuracy for diagnosing the early stage of Parkinson's disease using alpha stable distribution. *IETE Journal of Research*, 2023, 69(1):92—103.
- [16] Babu G S, Suresh S, Mahanand B S. A novel PBL-McRBFN-RFE approach for identification of critical brain regions responsible for Parkinson's disease. *Expert Systems with Applications*, 2014, 41(2):478—488.
- [17] 孙志军,薛磊,许阳明,等. 深度学习研究综述. *计算机应用研究*, 2012, 29(8):2806—2810.
- [18] Mary G, Suganthi N. Detection of Parkinson's disease with multiple feature extraction models and darknet CNN classification. *Computer Systems Science & Engineering*, 2022, 43(1):333—345.
- [19] Chakraborty S, Aich S, Kim H C. Detection of Parkinson's disease from 3T T1 weighted MRI scans using 3D convolutional neural network. *Diagnostics*, 2020, 10(6):402.
- [20] Zhu S. Early diagnosis of Parkinson's disease by analyzing magnetic resonance imaging brain scans

- and patient characteristic//2022 10th International Conference on Bioinformatics and Computational Biology. Hangzhou, China: IEEE, 2022: 116–123.
- [21] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding. 2018, arXiv:1810.04805.
- [22] Brown T B, Mann B, Ryder N, et al. Language models are few-shot learners. 2020, arXiv: 2005.14165.
- [23] Rao Y, Zhao W, Zhu Z, et al. Global filter networks for image classification. Advances in Neural Information Processing Systems, 2021, 34: 980–993.
- [24] Atwood J, Towsley D. Diffusion-convolutional neural networks. 2016, arXiv: 1511.02136v6.
- [25] Tran D V, Navarin N, Sperduti A. On filter size in graph convolutional networks. 2018, arXiv: 1811.10435
- [26] Gilmer J, Schoenholz S S, Riley P F, et al. Neural message passing for quantum chemistry//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: 2017: 1263–1272.
- [27] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HA, USA: IEEE, 2017: 2261–2269.
- [28] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 4510–4520.
- [29] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, CA, USA: IEEE, 2016: 770–778.

(责任编辑 杨可盛)