一种整合异构多模态特征的稀疏注意力深度学习模型，用于帕金森病严重程度分析

Dristi Datta[a]，Tanmoy Debnath[1a, c]，Minh Chau[b]，Manoranjan Paul[a]，Gourab Adhikary[d] 和 Md Geaur Rahman[a]

[a]查尔斯特大学计算、数学与工程学院，新南威尔士州，澳大利亚[b]查尔斯特大学牙科学院与医学科学学院，新南威尔士州，澳大利亚[c]人工智能与网络未来中心（*AICF*），新南威尔士州，澳大利亚[d]霍金斯诊所全科医疗机构，南澳大利亚州，澳大利亚

arXiv:2601.00519v1 [cs.LG] 2 Jan 2026

文章信息

关键词：帕金森病 多模态深度学习 可解释人工智能 基于注意力的融合 类别不平衡学习

摘 要

刻画帕金森病（PD）表现的异质性，需要将生物学和临床标志物整合到统一的预测框架中。尽管多模态数据能够提供互补信息，许多现有计算模型仍难以兼顾可解释性、类别不平衡问题，或难以有效融合高维影像与表格型临床特征。为解决这些局限，我们提出类别加权稀疏注意力融合网络（SAFN），这是一种可解释的深度学习框架，可用于稳健的多模态分析。所提出的 SAFN 框架通过模态专属编码器和对称交叉注意力机制，整合 MRI皮层厚度、MRI体积测量指标、全面的临床评估以及人口统计学变量，并捕捉影像和临床来源表征之间的非线性交互。稀疏性约束注意力门控融合层可动态优先关注信息量较高的模态，而类别平衡焦点损失（$\beta = 0.999$，$\gamma = 1.5$）则在不进行合成过采样的情况下缓解数据集不平衡。基于帕金森病进展标志物倡议（PPMI）的参与者（570 名 PD 患者、133 名健康对照），采用按受试者划分的五折交叉验证进行评估，SAFN 的准确率达到 0.98 ± 0.02，精确率-召回率曲线下面积达到 1.00 ± 0.00，优于既有的机器学习和深度学习基线模型。模型的可解释性分析揭示了符合临床逻辑的决策过程：模态门控将约 60% 的预测权重分配给临床评估，这与运动障碍学会的诊断原则一致，而 MRI 衍生特征则提供了互补的分层信息。因此，SAFN 提供了一种可复现、透明的多模态建模范式，将高性能人工智能与临床推理相结合，为神经退行性疾病的计算分析奠定了可信赖的基础。

1. 引言

帕金森病（PD）是一种复杂的进行性神经退行性疾病，会显著降低生活质量。其临床特征包括运动迟缓、僵直、震颤和姿势不稳等运动症状，以及认知障碍、睡眠障碍和自主神经功能障碍等多种非运动表现。从病理学角度看，PD的特征是黑质致密部（SNpc）中的多巴胺能神经元丢失；然而，通常只有在估计有50–70%的这些神经元发生变性后，才会出现明显的运动症状（Kempster，2025；Twala，2025）。

帕金森病也是全球增长最快的神经系统疾病之一。当代流行病学预测估计，到2050年，全球帕金森病患病人数将超过2,500万，较2021年增加112%，这主要是由人口老龄化和社会人口结构转变所推动（Li, Jin, Wang, Yu and Zhang，2025b；Su,Wei, Cao, Zhang, Li, Xu, Zhou, Shang, Luo, Yue and Miao，2025）。全球疾病负担研究的分析进一步显示，发病率、死亡率和伤残调整寿命年持续上升，尤其是在55岁及以上成年人中（Peng, Liu, WangandLi，2025）。因此，帕金森病给医疗卫生系统、患者和照护者带来了日益沉重的负担。

[1]通讯作者：tdebnath@csu.edu.au
ORCID(s):

尽管治疗管理不断进步，全面刻画疾病严重程度和表型仍颇具挑战。临床评估（如统一帕金森病评定量表（MDS-UPDRS）和Hoehn–Yahr 分期）是功能性评估的金标准，但高度依赖症状表现。患者状态的波动以及与非典型帕金森综合征的症状重叠，进一步增加了诊断的难度（Khan, Lavu and Neal，2024；Mattia,Chougar, Foubert-Samier, Meissner, Fabbri, Pavy-Le Traon,Rascol, Grabli, Degos, Pyatigorskaya, Faucher, Vidailhet,Corvol, Lehéricy and Péran，2025a；Welton, Hartono,Lee, Teh, Hou, Chen, Chen, Lim, Prakash, Tan, Tan and Chan，2024）。此外，尽管这些量表能够量化功能损害，却无法直接反映其潜在的神经解剖学退化。疾病轨迹的异质性和临床表现的显著差异，也阻碍了可靠、统一的疾病特征描述的制定。

这些挑战凸显了对客观、可扩展计算工具的迫切需求，这类工具能够综合异质数据，并将功能性临床信号与神经影像衍生的生物标志物相结合。这推动了人们对应用机器学习（ML）和深度学习（DL）方法的浓厚兴趣，以提升帕金森病表征的可重复性和临床相关性。近期研究表明，自动化生物标志物有望检测神经退行性变化，并将其与临床表型联系起来。因此，该领域正日益转向旨在捕捉帕金森病复杂多系统特性的多模态人工智能（AI）框架。

为使本研究置于这一不断发展的领域背景之中，我们回顾了近期涵盖结构、功能和微结构磁共振成像、影像组学、电生理及脑电图标志物、行为与数字表型、多模态融合策略、鉴别诊断建模，以及帕金森病计算生物标志物研究系统评估的研究。结构磁共振成像研究一致表明，帕金森病涉及黑质、基底神经节、苍白球、丘脑、脑干、小脑及额颞皮层的广泛宏观结构改变（Almgren, Hanganu, Camacho, Kibreab, Camicioli, Ismail, Forkert and Monchi, 2023年；Alrawis, Mohammad, Al Ahmadi and Al Muhtadi, 2025；Basaia, Sarasso, Sciancalepore, Balestrino, Musicco, Pisano, Stankovic, Tomic, Micco, Tessitore, Salvi, Meiburger, Kostic, Molinari, Agosta and Filippi, 2024；Camacho, Wilms, Mouches, Almgren and Souza, 2023；Camacho, Wilms, Almgren, Hanganu, Kibreab, Camicioli, Ismail, Forkert and Monchi, 2024；Hussain, Shah, Dawood, Xu, Alshamayleh, Khan and Ghazal, 2025；Islam and Khanam, 2024；Li, Ao, Wu, Wen, Ul Haq and Yin, 2024b；Mahajan, 2025；Welton等人，2024；Zhou, Zhu, Wang and Gao, 2025）。方法学涵盖手工设计的形态计量特征集，以及多中心卷积神经网络；后者的显著性图一致将深部灰质和额颞区域列为重点（Balıkçı Çiçek, Küçükakçalı, Deniz and Algül, 2025；Camacho 等人，2024）。包括 Swin Transformer 和基于注意力的模型在内的更先进架构，正越来越多地将病理线索定位于中脑区域，尽管外部泛化仍是一项挑战（Basaia等，2024；Hussain等，2025）。其他互补研究方向（包括规范建模）报告了与疾病存在相关的皮质下、额叶和小脑区域萎缩模式（Zheng, Zhou, Mao等，2024；Zhou等人，2025）。然而，一个反复出现的发现是，仅使用磁共振成像训练的模型——无论采用体素级方法还是基于特征的方法——在独立队列上的准确率往往会下降，这进一步凸显了构建能够稳健整合影像与稳定临床信息的框架的必要性。

除结构形态外，功能磁共振成像（fMRI）和微结构磁共振成像还能提供互补信息。静息态连接分析揭示了感觉运动网络和默认模式网络的变化（Cao、Wang、Xue、Zhang、Huang 和 Liu，2020年；Mattia、Chougar、Foubert Samier、Meissner、Fabbri、Pavy Le Traon、Rascol 和 Péran，2025b；Shi、Zhang、Wang、Wang、Yao、Li、Guo、Zheng 和 Ren，2022），而弥散张量成像（DTI）和神经黑色素敏感磁共振成像则可为黑质纹状体变性提供敏感的生物标志物（Camacho 等，2024；Chen、Liu、Luo 等，2024；Li、Liu、Wang、Liu、Lin 和 Xiong，2022；Mattia 等，2025b；Welton 等，2024；Zhao、Wang 和 Zhang，2022）。此外，采用脑电图（EEG）的电生理学研究将帕金森病进一步刻画为一种神经振荡病，并识别出α和β节律异常（Afonso、Edla 和 Ramesh，2025；Bunterngchit、Baniata、Albayati、Baniata、Alharbi、Alshammari 和 Kang，2025；Jibon、Tasbir、Talukder 等，2024；Li 等，2024a；Zhao、Dai、Li、Zhang、Liu、Wang、Wang 和 Lu，2024）。尽管这些模态能够提供丰富的生物学信息，但它们往往被孤立地进行研究，或需要复杂的预处理流程，从而限制了其临床应用性。

因此，多模态融合已成为整合生物标志物分析的一项主流策略。融合结构磁共振成像、遗传变异和临床评估的框架，其表现优于单模态基线（Dentamaro,Impedovo, Musti, Pirlo and Taurisano，2024；Li 等人，2025b；Mattia et al.，2025b；Sar, Puri, Nazand Al Khudhairy，2025；Yang, Hu, Chen, Gu, Lin and Xie，2025）。例如，PIDGN 模型利用门控注意力将磁共振成像与遗传单核苷酸多态性融合，体现了结合生物学和临床数据的价值（Li 等人，2025b）。其他方法

整合语音、步态、传感器衍生数据或表格数据的方法也展现出强劲的预测性能（Colautti,Mastrolia, Corbo, Nardone, Cifelli, Giustini, Di Giovanni, Sancesarioand Fusco，2025；Esan, Oyewole,Babalolaand Adekoya，2025；Jin, Li, Han,Qiu, Zhang,Xu and Han，2025；Yang 等人，2025）。尽管取得了这些进展，仍有若干持续存在的局限阻碍临床转化：（i）模态主导，即高维磁共振成像特征掩盖了信息量丰富的临床或人口统计学输入；（ii）缺乏可解释性，尤其是黑箱融合模型会掩盖各模态的相对贡献；（iii）类别不平衡，这会使模型在帕金森病患病率分布不均的真实世界数据集中产生学习偏差。

为应对这些挑战，我们开发了类别加权稀疏注意力融合网络（SAFN）。这是一种多模态架构，明确针对现有帕金森病模型的关键局限：异质特征结构、模态主导以及有限的可解释性。SAFN利用模态专属编码器整合MRI皮层厚度、MRI体积特征、临床评估和人口统计学变量，以保留模态内结构；利用对称交叉注意力对影像与临床流之间具有临床意义的交互进行建模；并利用稀疏性约束注意力门控融合层，防止高维磁共振成像特征掩盖具有信息价值的临床信号。通过将类别平衡焦点损失直接纳入优化目标，SAFN无需合成重采样即可缓解数据集不平衡问题。该框架为多模态帕金森病分析提供了一种均衡、透明且符合临床需求的方法，弥合了高性能人工智能与可解释决策之间的差距。

基于当前多模态研究中发现的局限性，本研究作出了以下几项重要贡献：

1. **统一的多模态分析框架**。 我们提出类别加权稀疏注意力融合网络（SAFN），通过选择性融合异质的MRI衍生模态和临床模态，应对模态异质性，并防止高维影像特征主导稀疏但具有信息价值的临床信号。

2. **内置类别不平衡处理**。 SAFN采用类别平衡焦点损失和EMA稳定化AdamW优化，明确应对真实世界队列中常见的严重帕金森病–健康对照组不平衡，无需合成过采样，同时保留少数类完整性。

3. **与临床相契合的可解释性**。 学习得到的模态门通过量化各模态层级的贡献，提供内在透明度；结果显示，SAFN将约60%的预测权重分配给临床评估，这与公认的帕金森病诊断和分期原则一致。

4. **可复现的基准测试方案**。 我们建立了严格的评估流程，以确保比较公平且可复现；采用一致的预处理、分层五折交叉验证，并系统地与经典机器学习基线和深度学习基线进行评估。

5. **整合四个互补的生物标志物领域**。 该框架整合了MRI皮层厚度、MRI体积测量特征、全面的临床评估和人口统计学因素，通过单一端到端预测模型，捕捉帕金森病的多系统和异质性特征。

6. **卓越的性能与架构验证**。通过广泛的消融研究和分层评估，我们证明，SAFN 的架构组件——交叉注意力、稀疏门控和类别均衡优化——在准确率、平衡准确率和精确率-召回率曲线下面积等度量方面，相较于强大的集成模型和深度学习基线，能够共同带来一致的改进。

本文其余部分安排如下。第 2 节介绍数据集、预处理流程和模态组织。第 3节阐述方法框架，包括基线模型和所提出的 SAFN 架构。第 4 节报告定量结果，包括 SAFN 与现有机器学习/深度学习基线的比较。第 5 节结合现有帕金森病文献解读这些结果，重点阐述临床相关性、消融研究结果和模态层级贡献，并概述未来工作的方向。最后，第 6节总结本研究的主要见解。

2. 数据准备

2.1. 数据获取

数据来源于帕金森病进展标志物倡议（PPMI）。这是一项多中心纵向生物标志物研究，提供经过协调统一的多模态数据，用于帕金森病（PD）研究。研究人员利用 PPMI 参与者标识符提取影像衍生特征和表格特征，并将其整合到统一的受试者级数据库中。

该数据集涵盖四个主要特征领域：

- **MRI皮层厚度：** 使用 FreeSurfer v7.1.1 从 T1加权磁共振成像中提取的区域皮层厚度测量值；
- **MRI体积特征：** 区域及全局体积指标，包括灰质、白质、小脑结构和颅内容积；
- **临床评估：** 运动和非运动量表，例如 MDS-UPDRS、蒙特利尔认知评估量表、NMSS、病程和用药状态；
- **人口统计学属性：** 年龄、性别、利手和受教育程度。

所有模态均映射至统一的参与者标识符体系，以确保影像学与临床数据源之间的可重复性和一致性。

2.2. 数据清洗、数据筛选和数据协调

初始数据集由703名参与者组成（570名帕金森病患者、133名健康对照者）。通过完整性检查识别重复条目、不匹配的标识符和不一致的变量编码。缺失情况超过20%的变量被排除。对于存在零星缺失值的连续特征，采用中位数替代法进行插补；而分类属性（例如性别、利手）则采用标签编码。

具有不完整磁共振成像采集数据或未能完成FreeSurfer重建的参与者被排除在分析之外。使用 *z*-分数和直方图检查对连续变量中的离群值进行筛查，以确保异质特征分布之间的尺度一致性。鉴于队列中固有的类别不平衡（约81%为帕金森病患者），未在数据层面应用重采样或再平衡；相反，在模型优化过程中对不平衡问题进行了明确处理。

清理后，特征被归入上述四个特定模态领域，以支持模态感知建模。所有数值变量均采用 *z*-分数标准化（零均值、单位方差）。MRI 在各扫描仪站点间的数据协调遵循 PPMI 的标准化采集方案，并通过对照 PPMI MRITechnicalOperationsManual 进行验证，从而确保多中心数据的一致性。

2.3. 用于分析的最终文件

用于建模的最终数据集包括：

- 70 MRI皮层厚度特征 es,
- 13 MRI体积指标,
- 409临床变量,
- 7人口统计学属性。

所有特征均整合为统一的表格（.csv）格式，同时保留相应的 NIfTI 文件，以便通过参与者标识符（PATNO）进行追溯和交叉参照。这种标准化结构确保了下游预处理的一致性，并使本研究评估的所有机器学习和深度学习模型能够进行公平比较。

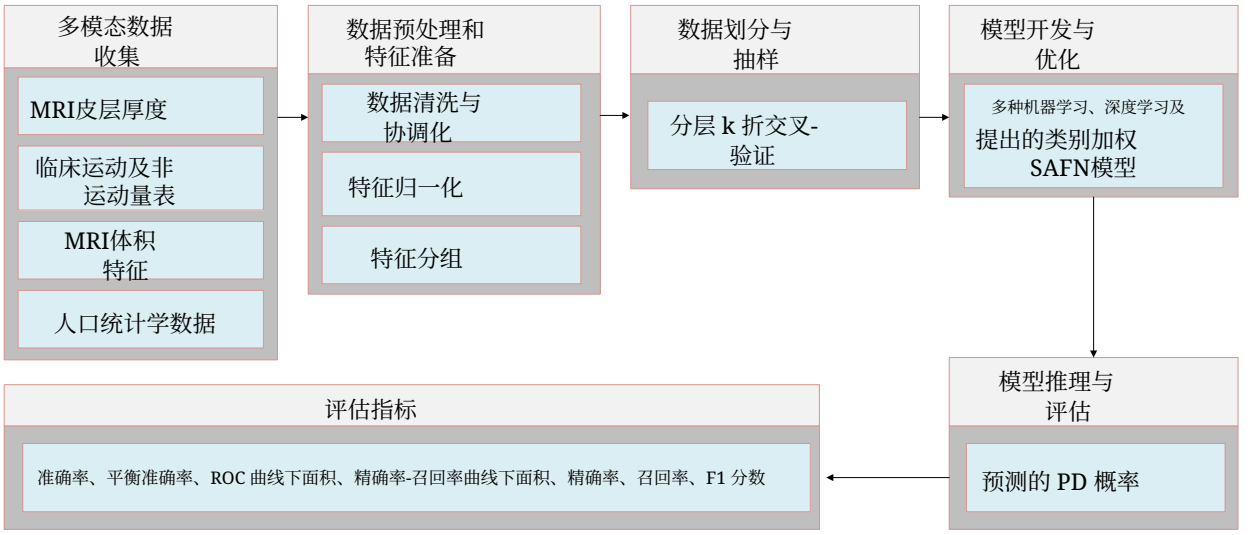


图 1: 帕金森病分类的整体方法学流程。多模态数据在模型训练前经过预处理、归一化和分层抽样。机器学习、深度学习及提出的类别加权 SAFN 模型均在相同的实验设置下进行训练，并采用准确率、平衡准确率、ROC 曲线下面积、精确率-召回率曲线下面积、精确率、召回率和 F1 分数等度量进行评估。

3. 方法学

本研究设计了这一方法框架，旨在利用多模态生物医学数据开发一种稳健且可解释的帕金森病分类系统。本节分为五个部分：（i）概述端到端处理和评估流程的整体工作流；（ii）基线机器学习和深度学习分类器的模型开发与优化；（iii）对所提出的类别加权 SAFN 模型的详细说明；（iv）采用互补性能指标进行模型推理与评估；以及（v）方法学总结。所有模型均使用相同的预处理特征集和分层五折交叉验证划分进行训练，以确保方法的一致性，并能够与所提出的 SAFN 模型进行公平比较。

3.1. 总体工作流程

端到端工作流程如图1所示。多模态输入包括 MRI 皮质厚度特征、MRI 体积指标、临床运动和非运动量表以及人口统计学变量；所有数据均通过统一的受试者标识符进行协调，并合并为一个表格数据集（参见第2节）。

模型训练所需的所有预处理步骤（插补、编码和标准化）均在数据准备阶段完成。在每个交叉验证折中，插补器和缩放器仅使用训练子集重新拟合，以防止数据泄漏，随后应用于验证集。特征被整理为四个考虑模态的组——MRI皮层厚度、MRI体积测量、临床和人口统计学——从而为所有基线模型提供一致的结构化输入，并支持 SAFN 中基于模态的编码。

采用分层 k 折交叉验证（$k = 5$）确保评估均衡，每一折依次作为验证集，其余折则构成训练集。对于有多条记录的参与者，StratifiedGroupKFold 确保受试者层面的独立性。包括 SAFN 在内的深度学习模型根据验证损失进行早停训练，最终结果报告为各折的平均值 ± 标准差。

3.2. 模型开发与优化

我们实现了一系列基线模型以评估性能，其完整的超参数设置汇总于表1中。机器学习基线模型包括逻辑回归、支持向量机、随机森林、XGBoost、LightGBM 和 K近邻，而深度学习基线模型则包括全连接人工神经网络和一维卷积神经网络。所有模型均使用相同的预处理特征集和分层交叉验证划分进行训练，以确保比较公平且可控。通过内部验证调整超参数，并使用 Adam 或 AdamW 优化器对深度学习模型进行优化，同时采用早停。所提出的类别加权 SAFN 整合了自注意力编码、

表 1基线机器学习、深度学习及所提出的 SAFN 模型的超参数设置。符号定义： C =逆正则化强度； γ = 径向基函数核宽度； λ, α = L2/L1正则化项； p = 闵可夫斯基距离指数； D_{model} = 嵌入维度。

模型	关键超参数
逻辑回归	L2正则化； $C = 1.0$；求解器：liblinear；类别权重：平衡；最大迭代次数：2000； 随机种子：42； 预处理：中位数填补 + 标准化；类别独热编码
SVM	核：径向基函数； $C = 1.0$； γ =缩放；类别权重：平衡；概率-性=True；随机种子：42；与上文相同的预处理
随机森林	树的数量：600；最大深度：无；最小分裂样本数：5；最小叶节点样本数：2；类别权重：平衡
XGBoost	估计器数量：400；学习率：0.05；最大深度：6；子采样比例：0.8；列采样比例：0.8；正则化： $\lambda = 1, \alpha = 0$
LightGBM	估计器数量：400；学习率：0.05；最大深度：6；叶节点数：31；子采样比例：0.8；按树列采样比例：0.8；正则化： $\lambda = 1, \alpha = 0$；目标函数：二分类；类别权重：平衡；随机种子：42
KNN	邻居数：7；权重：距离；度量：闵可夫斯基（ $p = 2$, 欧-氏距离）；标准化数值输入；独热编码分类变量
ANN	隐藏层： [128, 64]；激活函数： ReLU； Dropout：0.4； 损失函数：BCEWithLogits损失；通过 pos_weight 进行类别平衡；优化器：Adam ($\eta = 10^{-3}$)；批量大小：64；训练轮数：50；早停耐心值：8
一维卷积神经网络	卷积层： [32, 64]；卷积核尺寸： [5, 3]；激活函数：ReLU；全局平均池化；全连接层：64；Dropout：0.25；优化器：AdamW ($\eta = 2 \times 10^{-4}$, 权重衰减= 10^{-4})；批量大小：64；训练轮数：60；早停耐心值：10；通过 pos_weight 实现类别平衡
提出的 SAFN	$D_{\text{model}} = 64$；注意力头：4；层：2；Dropout：0.3；优化器：AdamW ($\eta = 2 \times 10^{-4}$, 权重衰减= 10^{-4})；学习率调度器： 预热-余弦退火；指数移动平均=0.999；训练轮数：60；早停耐心值：12

对称交叉注意力和稀疏正则化融合，并通过类别加权损失函数而非过采样来处理类别不平衡。

3.3. 提出的类别加权稀疏注意力融合网络*3.3.1.* 概述

提出的类别加权 SAFN 旨在利用异构表格模态对帕金森病进行分类，包括 MRI 皮层厚度特征、MRI 体积测量特征、临床运动和非运动量表以及人口统计学变量。传统的早期或后期融合策略往往存在以下问题：(i) 不考虑预测价值的差异而将所有模态一视同仁；(ii) 在帕金森病与健康对照受试者之间存在类别不平衡时表现不佳。SAFN 通过两项关键创新缓解了这些局限：(i) 引入带稀疏正则化的基于注意力的门控融合模块，有选择地强调信息丰富的模态，同时抑制噪声；(ii) 采用基于类别平衡焦点损失的不平衡感知目标函数，重新加权各类别的有效贡献，并使训练聚焦于难分类样本。此外，在临床数据流与 MRI 皮层厚度数据流之间引入对称交叉注意力机制，以捕获简单拼接常常忽略的模态间依赖关系。

3.3.2. 架构设计

图 2 展示了网络结构，其中包括模态特定的分词器/编码器、用于模态间信息交换的交叉注意力模块、稀疏注意力–门控融合模块，以及归一化分类头。

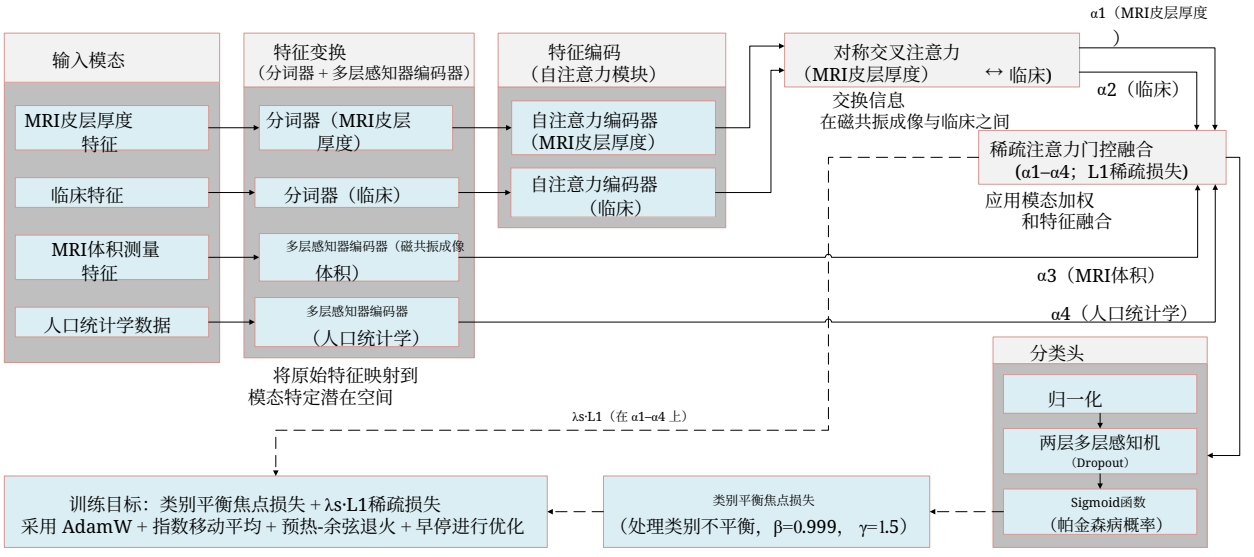


图 2： 所提出的用于帕金森病分类的类别加权稀疏注意力融合网络架构。四种输入模态——MRI皮层厚度、临床、MRI体积测量和人口统计学特征——分别通过模态专属分词器或多层感知机编码器进行处理。MRI皮层厚度和临床嵌入通过对称交叉注意力进行交互，随后采用带有可学习模态权重的稀疏注意力门控多模态融合（α_1–α_4）。所得融合表示 $\mathbf{H}$ 被传递至分类头，以预测帕金森病概率。实线箭头表示数据流，虚线箭头表示与损失相关的优化路径。

(a) 模态专属标记化与编码器。 令 $\mathbf{x}^{mri_ct} \in \mathbb{R}^{F_{mri_ct}}$、 $\mathbf{x}^{mri_vol} \in \mathbb{R}^{F_{mri_vol}}$、 $\mathbf{x}^{clin} \in \mathbb{R}^{F_{clin}}$ 和 $\mathbf{x}^{demo} \in \mathbb{R}^{F_{demo}}$ 表示四种模态——MRI皮层厚度、MRI体积、临床和人口统计学——经过中位数填补、标准化和类别编码后对应的特征向量。这里，F_{mri_ct}, F_{mri_vol}, F_{clin} 和 F_{demo} 表示各模态中的特征数量。

对于 *MRI*皮层厚度 和 临床 数据流，每个标量特征首先通过轻量级分词器映射为可学习的嵌入向量，从而实现按特征进行上下文建模。由此生成标记序列 $\mathbf{T}^{mri_ct} \in \mathbb{R}^{F_{mri_ct} \times D}$ 和 $\mathbf{T}^{clin} \in \mathbb{R}^{F_{clin} \times D}$，其中 D 是所有模态共享的嵌入维度。随后，每个标记序列均由一个两层Transformer编码器处理（四个注意力头、GELU激活函数、Dropout p=0.4），得到上下文表示：

$$\mathbf{E}^{mri_ct} = \mathrm{Enc}(\mathbf{T}^{mri_ct}) \in \mathbb{R}^{F_{mri_ct} \times D}, \tag{1}$$

$$\mathbf{E}^{clin} = \mathrm{Enc}(\mathbf{T}^{clin}) \in \mathbb{R}^{F_{clin} \times D}, \tag{2}$$

其中，Enc(·)表示特定于模态的Transformer编码器。随后，这些编码后的序列会进行对称交叉注意力和注意力池化（见部分(b)），以获得 $\mathbb{R}^D$中的定长模态表示。

对于*MRI*体积测量和人口统计学输入，由于它们是表格数据且维度较低，因此采用两层多层感知机（MLP）编码器，将原始特征向量直接投影到相同的 D维潜在空间：

$$\mathbf{Z}^{mri_vol} = \mathrm{MLP}(\mathbf{x}^{mri_vol}) \in \mathbb{R}^{D}, \qquad \mathbf{Z}^{demo} = \mathrm{MLP}(\mathbf{x}^{demo}) \in \mathbb{R}^{D}. \tag{3}$$

（*b*）*MRI*皮层厚度与临床特征之间的交叉注意力。为建模MRI皮层厚度与临床特征之间的互补关系，在特定于模态的Transformer编码器之后、注意力池化之前，应用对称交叉注意力。因此，交叉注意力作用于部分(a)中定义的编码后令牌序列 $\mathbf{E}^{mri_ct} \in \mathbb{R}^{F_{mri_ct} \times D}$ 和 $\mathbf{E}^{clin} \in \mathbb{R}^{F_{clin} \times D}$。

形式上，交叉注意力后的词元表示计算如下：

$$\tilde{\mathbf{E}}^{mri_ct} = \mathrm{CrossAttn}\left(\mathbf{E}^{mri_ct} \leftarrow \mathbf{E}^{clin}\right), \tag{4}$$

$$\tilde{\mathbf{E}}^{clin} = \text{交叉注意力}\big(\mathbf{E}^{clin} \leftarrow \mathbf{E}^{mri_ct}\big), \tag{5}$$

其中，记号 $\mathbf{A} \leftarrow \mathbf{B}$表示 $\mathbf{A}$ 中的词元充当查询，而 $\mathbf{B}$中的词元提供键和值值。此处，CrossAttn$(\cdot)$ 表示一个多头交叉注意力模块，其后依次连接残差连接、层归一化和逐位置前馈子层。

这种双向设计使两种模态能够交换互补信息：MRI皮层厚度特征通过临床背景得到丰富，而临床特征则融入MRI衍生的结构线索。在交叉注意力之后，注意力池化操作将每个词元序列聚合为固定长度的 D维向量，从而得到用于后续融合的精炼模态表示 $\tilde{\mathbf{Z}}^{mri_ct}$，$\tilde{\mathbf{Z}}^{clin} \in \mathbb{R}^D$。

（c）稀疏注意力门控融合。四种模态嵌入经过拼接，并通过可学习的Sigmoid门进行调制，以获得融合后的多模态表征。令 $\tilde{\mathbf{Z}}^{mri_ct}$、$\tilde{\mathbf{Z}}^{clin}$、$\mathbf{Z}^{mri_vol}$、$\mathbf{Z}^{demo} \in \mathbb{R}^D$ 表示从（a）和（b）部分获得的特定于模态的潜在向量。这些向量首先拼接为

$$\mathbf{Z} = [\tilde{\mathbf{Z}}^{mri_ct};\, \tilde{\mathbf{Z}}^{clin};\, \mathbf{Z}^{mri_vol};\, \mathbf{Z}^{demo}] \in \mathbb{R}^{4D}, \tag{6}$$

其中，分号“;”表示沿特征维度进行拼接，D 是每个模态的嵌入维度。

随后计算模态级门控向量：

$$\boldsymbol{\alpha} = \sigma(\mathbf{W}_g \mathbf{Z} + \mathbf{b}_g) \in (0,1)^4, \tag{7}$$

其中，$\mathbf{W}_g \in \mathbb{R}^{4\times 4D}$ 和 $\mathbf{b}_g \in \mathbb{R}^4$ 是可学习参数，$\sigma(\cdot)$ 表示逐元素 Sigmoid 函数激活。向量 $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \alpha_3, \alpha_4]^\top$ 分别编码 MRI皮层厚度、临床、MRI体积和人口统计学模态的相对重要性。

融合表征 是在拼接之前，将这些门应用于各模态嵌入后得到的：

$$\mathbf{H} = [\alpha_1\, \tilde{\mathbf{Z}}^{mri_ct};\, \alpha_2\, \tilde{\mathbf{Z}}^{clin};\, \alpha_3\, \mathbf{Z}^{mri_vol};\, \alpha_4\, \mathbf{Z}^{demo}] \in \mathbb{R}^{4D}. \tag{8}$$

因此，首先使用对应的门α_j 对每个模态向量进行缩放，然后将其拼接以形成融合多模态表征 $\mathbf{H}$。

为鼓励选择性地利用各模态，对门控系数施加 ℓ_1 稀疏性惩罚：

$$\mathcal{L}_{\text{sparse}} = \lambda_s \sum_{j=1}^{4} |\alpha_j|, \tag{9}$$

其中，$\lambda_s > 0$ 控制稀疏性强度，促使模型侧重于信息量最大的模态，同时抑制贡献较弱的模态。

(d) 分类头。首先使用层归一化对融合后的多模态表征 $\mathbf{H} \in \mathbb{R}^{4D}$ 进行归一化，然后将其输入一个采用GELU激活函数和Dropout正则化的两层多层感知机（MLP），以生成标量对数几率 $s \in \mathbb{R}$。最终的帕金森病（PD）概率通过Sigmoid激活函数获得：

$$s = \text{MLP}\big(\text{LayerNorm}(\mathbf{H})\big), \tag{10}$$

$$\hat{y} = \sigma(s), \tag{11}$$

其中，LayerNorm$(\cdot)$表示层归一化，MLP$(\cdot)$ 表示分类头，$\sigma(\cdot)$ 为Sigmoid函数，$\hat{y} \in (0,1)$ 表示PD的预测概率。

3.3.3. 不平衡感知目标函数

为缓解 PD 与健康对照样本之间固有的类别不平衡，SAFN 采用类别平衡焦点损失（CB-Focal），将有效样本数加权与焦点调制相结合，以强调难分类及代表性不足的样本。给定分类头输出的对数几率 s及其对应的概率$p = \sigma(s)$，单样本损失定义如下：

$$\mathcal{L}_{\text{CB-Focal}} = -\alpha_1 (1-p)^\gamma\, y \log(p+\epsilon) - \alpha_0\, p^\gamma\, (1-y) \log(1-p+\epsilon), \tag{12}$$

其中， $y \in \{0,1\}$ 表示真实标签（PD $=1$，健康对照 $=0$），γ 为聚焦参数，用于控制对难分类样本的强调程度；$\epsilon = 10^{-7}$ 为用于数值稳定性的小常数（通过将概率截断至 $[\epsilon, 1-\epsilon]$ 来实现）。

类别平衡系数 α_c（其中 α_1 用于 PD，α_0 用于 HC）根据 有效样本数 （Cui、Jia、Lin、Song 和 Belongie，2019）计算得出：

$$\alpha_c \propto \frac{1-\beta}{1-\beta^{n_c}}, \qquad \beta \in (0,1), \quad n_c = \text{number of samples in class } c. \tag{13}$$

在我们的实现中，有效样本数权重按小批次 计算，依据每个批次特定的类别计数 n_c。这种动态公式在保留类别平衡重加权原则的同时，确保了小批次训练过程中的稳定优化。所得类别权重与有效样本数（$w_{\text{pos}} \propto 1/\text{EN}_{\text{pos}}$、$w_{\text{neg}} \propto 1/\text{EN}_{\text{neg}}$）成反比，且不强制满足约束条件 $\alpha_0 + \alpha_1 = 1$。

总体训练目标将类别平衡焦点损失部分与模态门控层所引入的稀疏正则化相结合：

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CB-Focal}} + \lambda_s \mathcal{L}_{\text{sparse}}, \tag{14}$$

其中，$\lambda_s > 0$ 控制施加于模态门的 ℓ_1 稀疏约束的强度。

3.3.4. 训练与优化

所有组件均使用 AdamW（学习率 2×10^{-4}，权重衰减 10^{-4}，β_1=0.9，β_2=0.999，$\epsilon = 10^{-8}$）进行端到端训练。前 10% 个优化步骤采用线性预热，随后使用余弦衰减调度，将学习率逐渐降至零。在整个训练过程中，对模型参数维护指数移动平均（EMA，衰减率为 0.999）；所有验证和早停决策均使用 EMA 权重，以提高稳定性。为进一步稳定更新，将梯度裁剪至 ℓ_2-范数 1.0。训练最多进行 60 轮，并采用早停机制（耐心值为 12）；监控指标为复合验证度量，即在固定阈值 0.5 下计算的曲线下面积、平衡准确率和 F1 分数的均值。小批次大小设为 64。除非另有说明，各层均采用 PyTorch 默认的初始化方案。所有超参数均事先 从标准范围内选定，并在 5 折中保持不变，以确保 SAFN 与基线模型之间进行公平且可复现的比较。

3.3.5. 交叉验证与预处理

采用 5 折分层交叉验证策略，以确保各折中帕金森病/健康对照类别的比例保持一致。在每一折中，仅在训练集上拟合插补器、缩放器和标签编码器，然后将其应用于验证划分，以防止数据泄漏。当多个观测值属于同一受试者时，使用 StratifiedGroupKFold 并依据受试者标识符进行划分，确保训练集与验证集之间受试者级别的独立性。评估指标分别在固定决策阈值 0.5 和每折特定的最佳 F1 阈值下报告；后者通过扫描 $t \in [0.05, 0.95]$ 来确定使 F1 分数最大化的阈值。

3.3.6. 通过注意力门实现可解释性

SAFN 提供两个互补层面的可解释性。首先，磁共振成像（皮层厚度和体积测量）编码器及临床编码器中的词元级 注意力图突出显示了对预测贡献最大的特征。其次，模态级 门控系数 $\boldsymbol{\alpha}$ 量化了各模态对融合表示及最终输出的相对贡献。对不同受试者的 $\boldsymbol{\alpha}$ 进行汇总，可得到模态重要性的全局度量；通常结果显示，临床特征的贡献更大，其次是 MRI皮层厚度和磁共振体积输入，而人口统计学变量则提供辅助性背景信息。这些层次化注意力机制使模型的决策过程能够被透明地审视，在无需依赖外部事后解释方法的情况下，提升了科学可解释性和临床可信度。

3.4. 模型推理与评估

模型训练完成后，在每一折的留出验证子集上进行推理，以在一致的实验条件下评估泛化性能。对于所有模型，均使用验证轮次中表现最佳时的权重，在相应的验证划分上生成预测结果。对于每位受试者 i，获得一个标量对数几率 s_i，并将其转换为概率 $\hat{y}_i = \sigma(s_i)$，表示预测为PD的可能性。固定的0.5决策阈值作为主要运行点；此外，还通过选择每一折中使F1分数最大化的

$t \in [0.05, 0.95]$ 值来确定自适应阈值。这种双阈值策略为分类器校准、精确率–召回率权衡以及跨验证折的稳健性提供了互补视角。

评估指标。 模型性能采用标准二分类指标进行量化，这些指标综合评估整体准确率、区分能力以及 PD 与健康对照预测之间的平衡。令 TP、 TN、 FP 和 FN 分别表示真正例、真负例、假正例和假负例。各项指标定义如下：

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{15}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \qquad \text{Recall (Sensitivity)} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{16}$$

$$\text{F1 分数} = 2 \times \frac{\text{精确率} \times \text{召回率}}{\text{精确率} + \text{召回率}}, \tag{17}$$

$$\text{Balanced Accuracy} = \frac{1}{2}\left(\frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FP}}\right). \tag{18}$$

此外，使用受试者工作特征曲线下面积（ROC 曲线下面积）和精确率-召回率曲线下面积（精确率-召回率曲线下面积）评估与阈值无关的区分能力：

$$\text{ROC-AUC} = \int_0^1 \text{TPR}(x)\, \text{dFPR}(x), \tag{19}$$

$$\text{PR-AUC} = \int_0^1 \text{Precision}(r)\, \text{dRecall}(r), \tag{20}$$

其中， TPR 和 FPR 分别表示所有阈值下的真正例率和假正例率。所有指标均针对每个交叉验证折独立计算，并汇总为均值 ±标准差，以确保对不同模型进行统计上公平的比较。

可视化与报告。 对于每个模型，均生成了逐折混淆矩阵、ROC 曲线和PR曲线，以可视化类别区分和校准行为。通过在统一网格上对每个折的曲线进行插值并跨折平均，计算得到平均ROC和PR曲线，从而生成平滑的交叉验证性能汇总表示。为保持清晰度并避免冗余，结果部分仅报告平均混淆矩阵、ROC曲线和PR曲线。最终性能值以所有评估指标的均值 ± 标准差呈现，从而能够全面比较基线机器学习模型、深度学习模型和所提出的类别加权 SAFN。所有评估均使用 scikit-learn（v1.5）和 PyTorch（v2.1）实现，确保完全可复现。

3.5. 方法学总结

总而言之，所提出的框架在统一的实验设置中，整合了多模态特征预处理、基线机器学习和深度学习建模，以及可解释的类别加权 SAFN 的开发。所有模型均采用相同的预处理流程和分层五折交叉验证进行训练，以确保方法学公平性和可复现性。所有模型均以一致的方式计算评估指标，性能报告为各折的均值 ± 标准差。该连贯的设计为客观比较所提出的 SAFN 与传统机器学习和深度学习方法的预测性能及可解释性奠定了坚实基础。

4. 结果

本节展示了所有已实现的帕金森病分类模型的比较结果。为确保公平性和可复现性，所有传统机器学习方法和深度学习方法均在统一的五折交叉验证框架下进行评估。按照方法学部分的描述，采用互补的性能指标评估各模型，包括准确率、平衡准确率、ROC 曲线下面积、精确率-召回率曲线下面积、精确率、召回率和 F1 分数。为确保比较公平，所有模型均使用相同的特征集和相同的数据划分进行训练。针对类别不平衡问题，采用了适合各模型的策略：线性分类器和基于核的分类器（逻辑回归、支持向量机、K近邻）使用

表 2传统机器学习与深度学习模型的性能比较（五折交叉验证）。数值为均值 ± 标准差，四舍五入至小数点后两位。

模型	准确率	平衡 Acc	ROC 曲线下面积	精确率-召回率曲线下面积	精确率	召回率	F1
逻辑回归	0.94 ± 0.01	0.93 ± 0.03	0.98 ± 0.00	0.99 ± 0.00	0.97 ± 0.01	0.95 ± 0.02	0.96 ± 0.01
SVM	0.94 ± 0.01	0.89 ± 0.03	0.97 ± 0.01	0.99 ± 0.00	0.95 ± 0.01	0.97 ± 0.01	0.96 ± 0.01
随机森林	0.95 ± 0.06	0.95 ± 0.03	0.95 ± 0.03	0.98 ± 0.01	0.99 ± 0.01	0.95 ± 0.07	0.97 ± 0.04
XGBoost	0.96 ± 0.04	0.96 ± 0.02	0.97 ± 0.01	0.98 ± 0.01	0.99 ± 0.01	0.96 ± 0.05	0.97 ± 0.03
LightGBM	0.96 ± 0.01	0.95 ± 0.01	0.98 ± 0.01	0.98 ± 0.01	0.98 ± 0.01	0.98 ± 0.01	0.98 ± 0.01
KNN	0.88 ± 0.01	0.83 ± 0.04	0.90 ± 0.04	0.97 ± 0.01	0.94 ± 0.02	0.91 ± 0.01	0.92 ± 0.01
ANN	0.93 ± 0.03	0.91 ± 0.03	0.98 ± 0.01	0.99 ± 0.01	0.97 ± 0.01	0.94 ± 0.05	0.95 ± 0.02
一维卷积神经网络	0.58 ± 0.20	0.61 ± 0.07	0.70 ± 0.06	0.90 ± 0.03	0.70 ± 0.25	0.57 ± 0.30	0.62 ± 0.31
类别加权 SAFN（所提出的方法）	**0.98 ± 0.02**	**0.97 ± 0.05**	**0.98 ± 0.02**	**1.00 ± 0.00**	**0.99 ± 0.03**	**0.99 ± 0.01**	**0.99 ± 0.01**

注释。 帕金森病（阳性）类别的精确率-召回率曲线下面积、精确率、召回率和 F1分数均有报告。随机森林和 SAFN 的训练未使用合成少数类过采样技术；SAFN 通过类别加权损失处理类别不平衡。所有数值均以五折交叉验证的均值 ±标准差表示；由于各折之间存在差异，均值 ± 标准差区间可能超出 $[0, 1]$, 的理论范围，但所有折级指标均处于有效范围内。

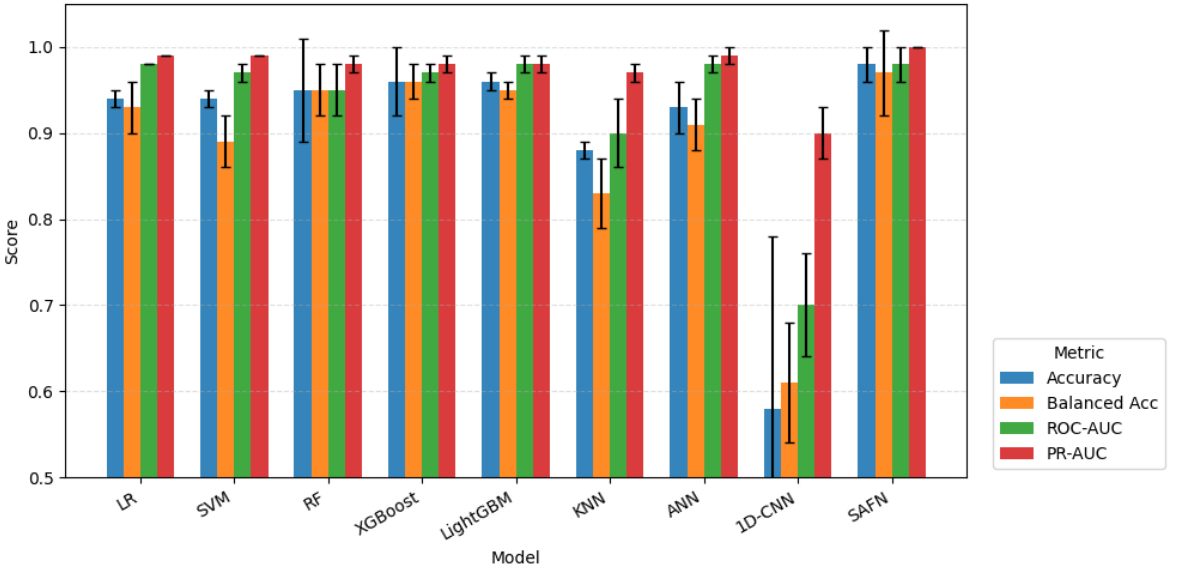


图 3： 分组柱状图展示了传统机器学习模型和深度学习模型在四项关键评估指标上的性能：准确率、平衡准确率、ROC 曲线下面积和精确率-召回率曲线下面积。误差线表示五个交叉验证折的标准差。与基线模型相比，所提出的类别加权 SAFN 始终表现出较高的性能，且方差极小。

类别_权重=平衡策略进行训练；基于树的集成模型（随机森林、XGBoost、LightGBM）采用内置的样本重新加权；深度学习模型（人工神经网络、一维卷积神经网络、SAFN）则使用类别加权二元交叉熵损失，以确保帕金森病患者和健康对照样本之间的学习平衡。

表2 汇总了定量结果，随后图 3 中的分组柱状图对关键度量进行了直观比较。图 4中的雷达图进一步展示了全部七项度量的整体性能概况。最后，图 5–7 展示了各模型组的详细混淆矩阵、PR曲线和 ROC 曲线。

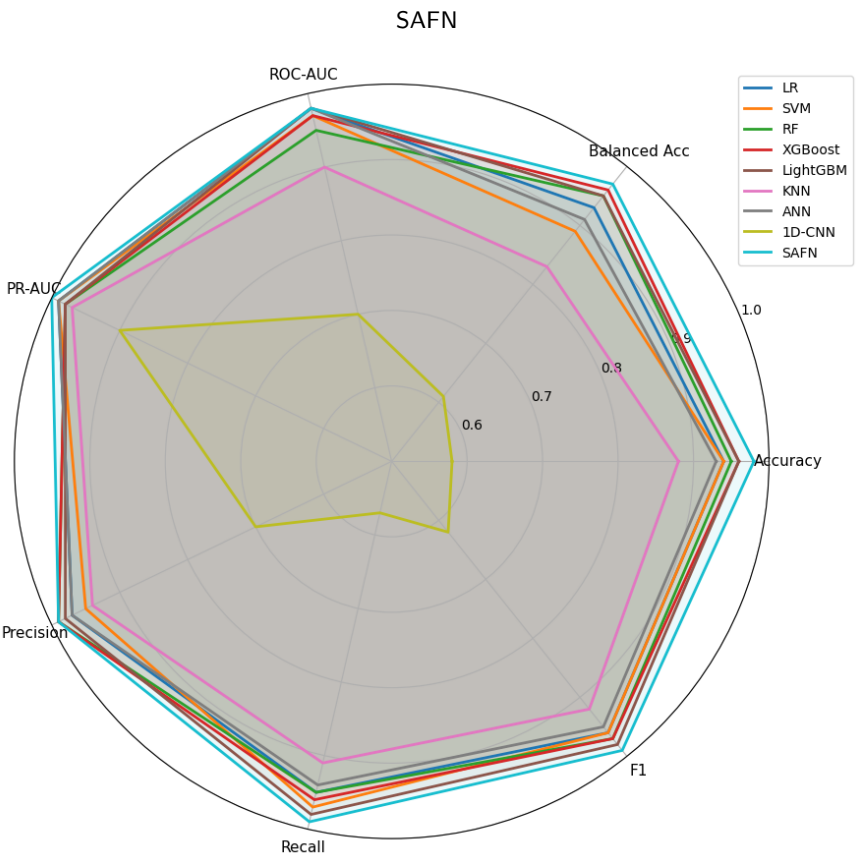


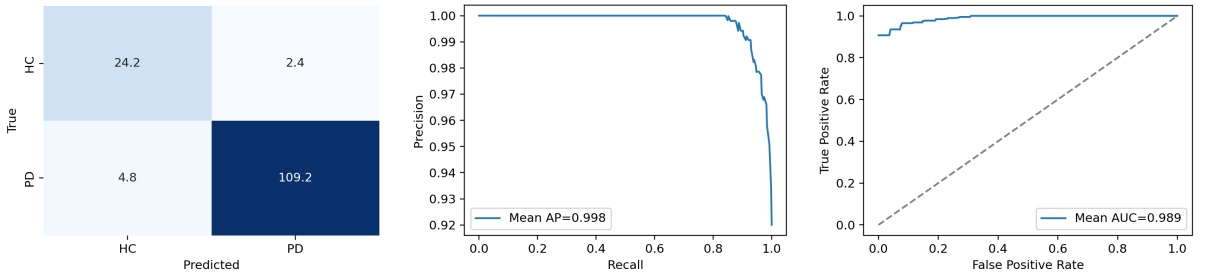
图 4： 雷达图比较了所有基线模型和所提出的类别加权 SAFN 在七项评估指标（准确率、平衡准确率、ROC 曲线下面积、精确率-召回率曲线下面积、精确率、召回率和 F1 分数）上的性能。数值对应五折交叉验证的平均得分。SAFN 模型形成最外层轮廓，表明其性能始终优于其他模型。

在机器学习模型中，大多数分类器均取得了较高的准确率（>0.9），凸显了由临床、运动和非运动属性构成的多模态特征表示具有较强的判别能力。在基于线性和核的方法中，逻辑回归与支持向量机的性能几乎相同（准确率：0.94 ± 0.01），二者的 ROC-AUC值均超过 0.97。其稳定的性能体现了有效正则化（通过 L_2 惩罚项和最大化间隔实现）以及对中等非线性边界的适用性。随机森林的平衡准确率（0.95 ± 0.03）和 F1 分数（0.97 ± 0.04）略高，这得益于其集成结构能够捕捉特征交互并减轻过拟合。不过，其在各折间较高的方差（标准差最高达 0.06）可能是由折规模不大时的抽样变异所致。

在提升树模型中，XGBoost 和 LightGBM 展现出最强且最稳定的性能。两种模型都能有效捕捉多模态输入空间中的非线性结构，但 LightGBM 实现了最高的准确率（0.96–0.97）和平衡准确率（0.96 ± 0.01），这得益于其按叶生长策略和高效的正则化。相比之下，K近邻表现出明显较低的准确率（0.88 ± 0.01）和 ROC– 曲线下面积（0.90 ± 0.04），反映出其对高维特征缩放的敏感性，以及基于距离的指标在异质特征空间中的判别价值降低。

图5和6 对这些趋势进行了可视化解读。图5 展示了逻辑回归、支持向量机和随机森林的混淆矩阵及精确率-召回率曲线/ROC曲线，它们均表现出较高的真阳性率以及迅速上升的精确率-召回率曲线，表明其帕金森病判别近乎完美。图6 展示了XGBoost、LightGBM和K近邻：提升模型的混淆矩阵紧凑，ROC曲线趋近左上角；而K近邻的精确率-召回率曲线较宽，这与其较弱的定量性能相符。

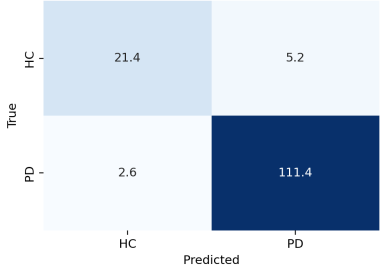
在深度学习组（图7）中，基线人工神经网络取得了良好的结果（准确率：0。93 ± 0.03；ROC– 曲线下面积：0。98 ± 0.01），这证实了只要



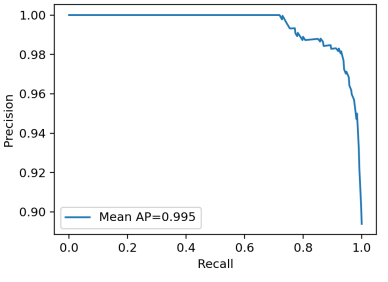
(a) 逻辑回归：混淆矩阵（平均值）

(b) 逻辑回归：精确率-召回率曲线（均值）

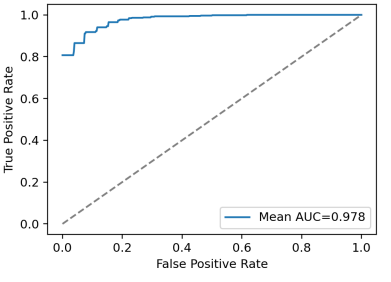
(c) 逻辑回归：ROC曲线（均值）



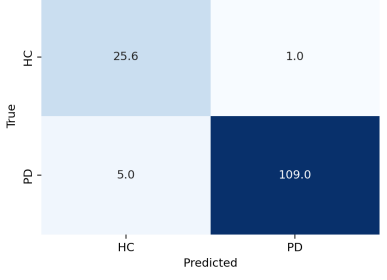
(d) 支持向量机：混淆矩阵（平均值）



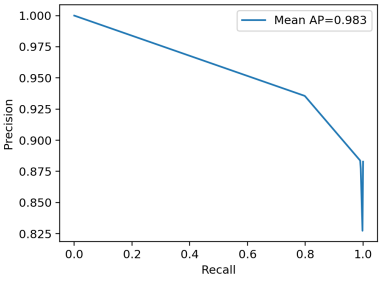
(e) 支持向量机：精确率-召回率曲线（均值）



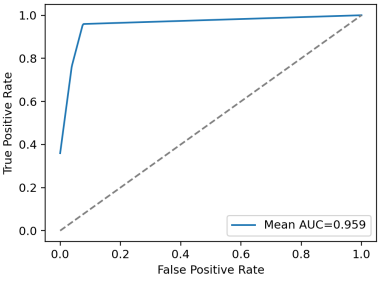
(f) 支持向量机：ROC曲线（均值）



(g) 随机森林：混淆矩阵（平均值）



(h) 随机森林：精确率-召回率曲线（平均值）



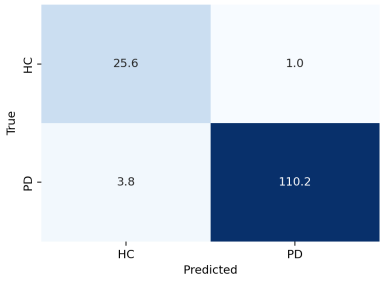
(i) 随机森林：ROC曲线（平均值）

图 5：传统机器学习分类器（逻辑回归、支持向量机和随机森林）的性能可视化。对于每个模型，该图展示了通过五折交叉验证计算得出的平均混淆矩阵、平均精确率-召回率曲线和平均 ROC 曲线。

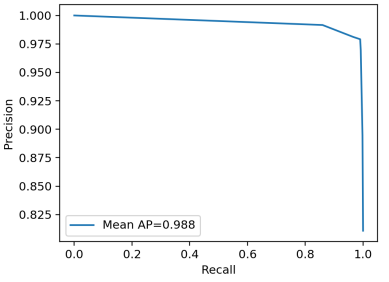
进行适当的正则化，全连接网络便能有效地建模非线性依赖关系。然而，一维卷积神经网络的性能不稳定（准确率：0.58 ± 0.20），这可能是由于过拟合，以及在缺乏显著局部序列依赖关系的特征空间中，卷积结构带来的益处有限。

所提出的类别加权 SAFN 在所有模型中取得了最佳综合性能，准确率为 0.98 ± 0.02，精确率-召回率曲线下面积达到完美的 1.00 ± 0.00。SAFN 将自注意力与特征归一化相结合，从而能够对信息性特征进行动态重新加权，并抑制无关或冗余维度。其类别加权优化进一步确保少数类和多数类均衡学习，从而提高召回率和精确率（两者均为 0.99）。在所有模型中，SAFN 的混淆矩阵对角线特征最为明显，表明其错误分类极少，且在各折之间具有较强的泛化能力。

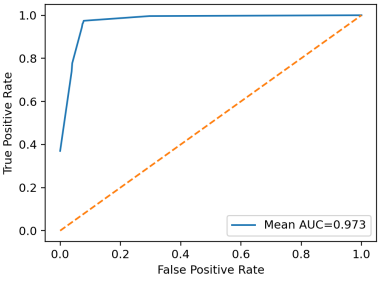
总体而言，表2 和图5–7 的结果表明，LightGBM 和 XGBoost 等集成模型为结构化临床数据提供了强有力的基线。然而，所提出的 SAFN 利用注意力驱动的特征融合和类别加权优化，超越了这些基线，在帕金森病分类中实现了更强的泛化能力和更均衡的预测能力。



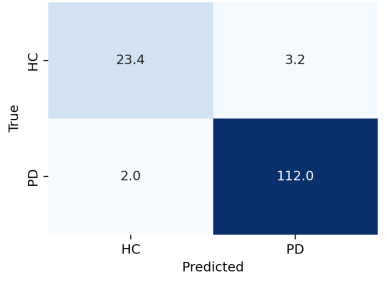
(a) XGBoost：混淆矩阵（平均值）



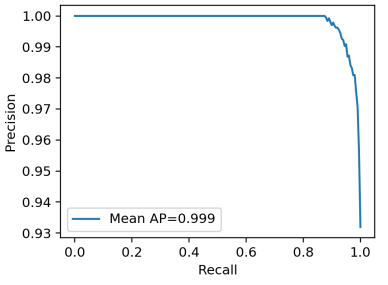
(b) XGBoost：精确率-召回率曲线（均值）



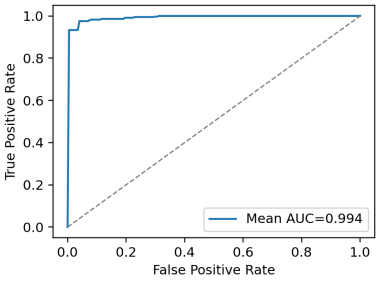
(c) XGBoost：ROC曲线（均值）



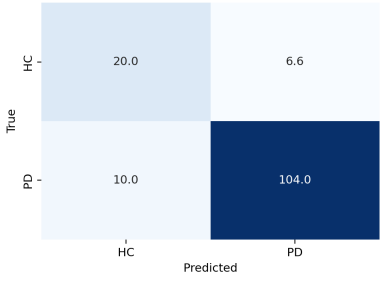
(d) LightGBM：混淆矩阵（平均值）



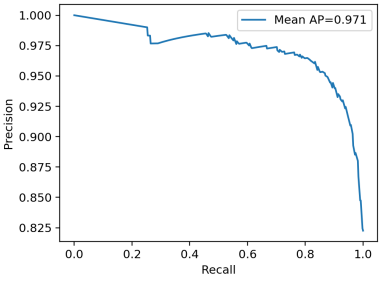
(e) LightGBM：精确率-召回率曲线（均值）



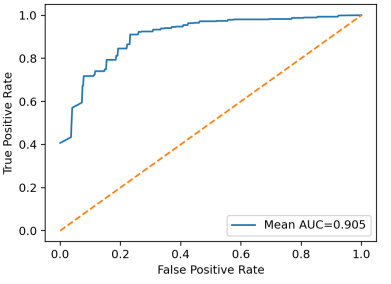
(f) LightGBM：ROC曲线（均值）



(g) K近邻：混淆矩阵（平均值）



(h) K近邻：精确率-召回率曲线（平均值）



(i) K近邻：ROC曲线（平均值）

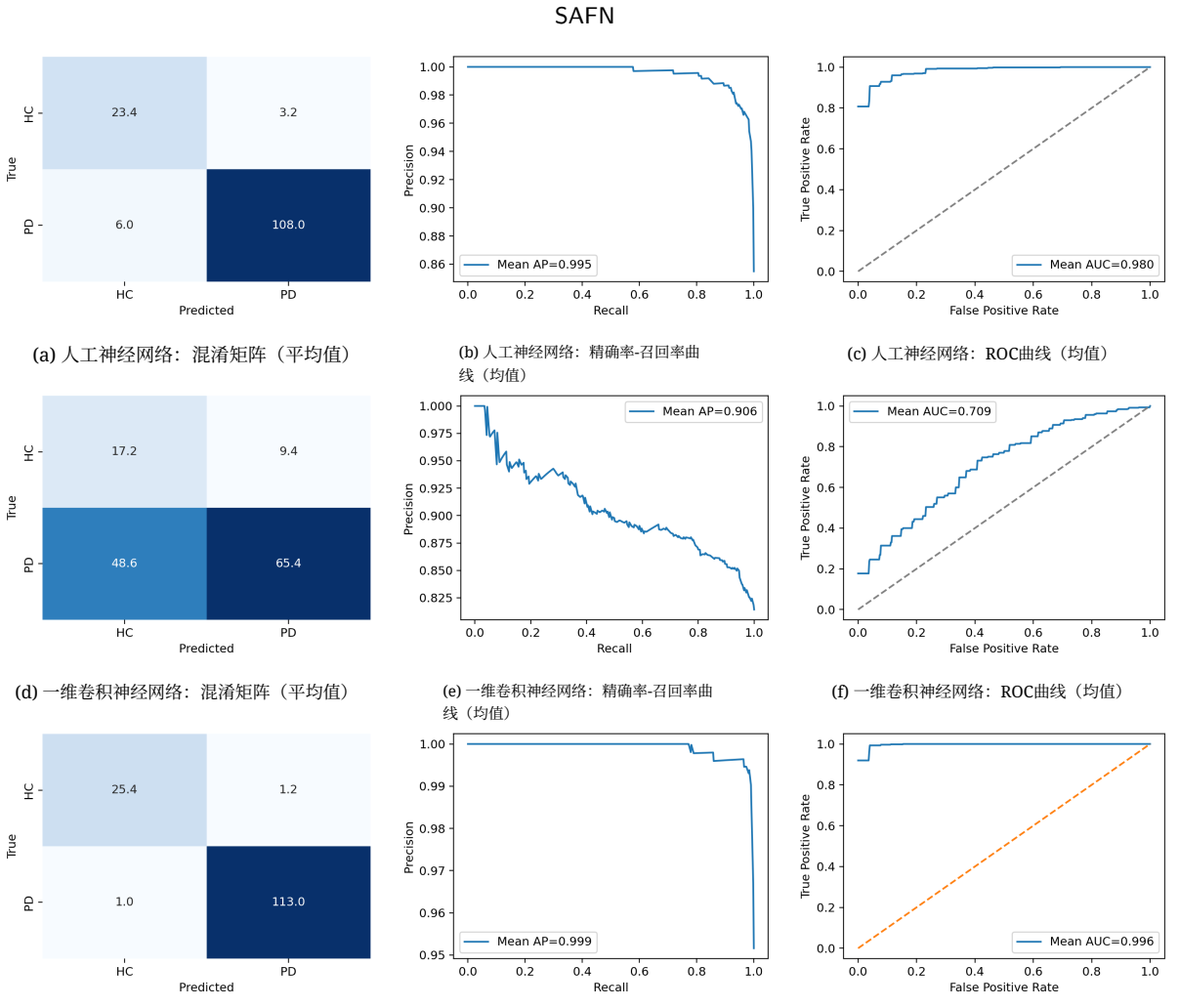
图 6： 高级机器学习分类器（XGBoost、LightGBM 和 K近邻）的性能可视化。对于每个模型，该图展示了通过五折交叉验证计算得出的平均混淆矩阵、平均精确率-召回率曲线和平均 ROC 曲线。

5. 讨论

5.1. 统计组间差异

在建模之前，先进行基线组水平统计分析，以刻画帕金森病组与健康对照组参与者之间的差异。首先使用夏皮罗–威尔克检验评估连续变量是否符合正态分布。由于大多数变量的分布偏离正态，采用非参数曼–惠特尼U检验比较帕金森病组与健康对照组。使用 Cliff's δ 量化效应量，其中 $|\delta| = 0.2$、0.5 和 0.8 分别代表小、中和大效应。分类变量采用 χ^2 独立性检验进行分析；当任一期望单元格计数低于 5 时，则使用费希尔精确检验。分类变量之间的关联使用 Cramér's V 进行汇总，并解释为小效应（$V < 0.1$）、中效应（$0.1 \leq V < 0.3$）或大效应（$V \geq 0.3$）。所有 p-值均采用 Benjamini–Hochberg错误发现率（FDR）程序进行多重比较校正，显著性阈值为 $q = 0.10$。连续变量以中位数 [四分位距], 报告，分类变量则以计数和百分比报告。

表 3 总结了经 FDR 校正后组间差异具有统计学意义且判别力最强的连续变量。正如预期，帕金森病参与者的运动严重程度评分明显更高，包括统一帕金森病评定量表 III 总分（NP3TOT）、Hoehn–Yahr分期（NHY）和运动迟缓（NP3BRADY），这些指标的效应量均较大



(a) 人工神经网络：混淆矩阵（平均值）

(b) 人工神经网络：精确率-召回率曲线（均值）

(c) 人工神经网络：ROC曲线（均值）

(d) 一维卷积神经网络：混淆矩阵（平均值）

(e) 一维卷积神经网络：精确率-召回率曲线（均值）

(f) 一维卷积神经网络：ROC曲线（均值）

(g) SAFN：混淆矩阵（平均）

(h) SAFN：精确率-召回率曲线（平均）

(i) SAFN：ROC曲线（平均）

图7： 深度学习模型（人工神经网络、一维卷积神经网络和所提出的类别加权 SAFN）的性能可视化。对于每个模型，该图展示了在 5 折交叉验证中计算得到的平均混淆矩阵、平均精确率-召回率曲线和平均 ROC 曲线。

至极大（例如 $|\delta| \geq 0.84$），且校正后的 p值极小。其他运动分项得分，如 NP3FACXP、NP3RTCON、NP3FTAPR、NP3POSTR 和 NP3RIGRU，也显示出较大的效应。非运动功能性指标 MSEADLG 表现出中到大的效应（Cliff 的 $\delta = 0.60$），反映出帕金森病组在日常生活活动中的独立能力有所下降。

这些组间差异与帕金森病研究中已有及新近出现的证据相符。NP3TOT反映总体运动严重程度，反复被确认为帕金森病进展最有力的临床预测因子之一（Germani,Bhagwat, Dugré, Gau, Montillo, Nguyen, Sokolowski, Sharp, Polineand Glatard，2025）。近期一项重复验证研究“利用临床和功能性磁共振成像特征预测帕金森病病程”发现，MDS-UPDRS第三部分指标始终是最主要的预测特征；移除运动指标后，模型性能显著下降（Germani et al.，2025）。这支持了NP3TOT作为高信息量临床标志物的重要性，也与表3中观察到的大效应量直接吻合。运动迟缓特征（NP3BRADY、NP3FTAPR）也具有高度相关性。一项采用PPMI五种时间序列模态的多模态可解释机器学习重要研究显示，在多个可解释性框架（SHAP、LIME、SHAPASH）中，运动迟缓都是最主要且最稳定的预测因子（Junaid, Ali, Eid,El-Sappagh and Abuhmed，2023年）。作者得出结论，运动迟缓相关评分不仅具有预测作用，也是区分早期帕金森病的重要医学基础

表3基线时区分度最高的连续变量（PD 与 HC）。数值以中位数 [四分位距]表示； p_{FDR} 采用 Benjamini–Hochberg 方法校正（q=0.10）。效应量为 Cliff's δ （非参数；负值表示这些变量在 PD 组中的值更大）。

变量	PD 中位数 [四分位距]	健康对照中位数 [四分位距]	Test	p_{FDR}	Cliff's δ
NP3TOT	22.00 [15.00, 30.00]	1.00 [0.00, 3.00]	曼-惠特尼U检验	3.25e-54	-0.98
NHY	2.00 [1.00, 2.00]	0.00 [0.00, 0.00]	曼-惠特尼U检验	2.05e-77	-0.94
NP3BRADY	1.00 [1.00, 2.00]	0.00 [0.00, 0.00]	曼-惠特尼U检验	1.70e-47	-0.84
NP3FACXP	1.00 [1.00, 2.00]	0.00 [0.00, 0.00]	曼-惠特尼U检验	3.22e-46	-0.79
NP3RTCON	1.00 [0.00, 2.00]	0.00 [0.00, 0.00]	曼-惠特尼U检验	3.00e-34	-0.66
NP3FTAPR	1.00 [1.00, 2.00]	0.00 [0.00, 0.00]	曼-惠特尼U检验	1.59e-33	-0.66
NP3POSTR	0.00 [0.00, 1.00]	0.00 [0.00, 0.00]	曼-惠特尼U检验	2.79e-31	-0.64
NP3RIGRU	1.00 [0.00, 2.00]	0.00 [0.00, 0.00]	曼-惠特尼U检验	9.37e-31	-0.61
MSEADLG	90.00 [90.00, 100.00]	100.00 [100.00, 100.00]	曼-惠特尼U检验	2.93e-28	0.60
NP3RIGLU	1.00 [0.00, 2.00]	0.00 [0.00, 0.00]	曼-惠特尼U检验	6.65e-28	-0.60

注释。NP3TOT：MDS-UPDRS第三部分总分；NHY：Hoehn-Yahr分期；NP3BRADY：全身运动迟缓；NP3FACXP：面部表情（面具脸）；NP3RTCON：静止性震颤持续性；NP3FTAPR：手指敲击（右手）；NP3POSTR：姿势；NP3RIGRU：僵直（右上肢）；NP3RIGLU：僵直（左上肢）；MSEADLG：改良Schwab和England日常生活活动评分（分值越低表示残疾程度越重）。

表4分类基线差异（帕金森病 vs 健康对照）。 p_{FDR} 经 Benjamini–Hochberg 校正；效应量为 Cramér's V。

变量：PDSTATE	
帕金森病人数（%）	OFF状态：86（15.1%）；ON状态：173（30.4%）；nan：311（54.6%）
健康对照计数（%）	OFF状态：0 (0.0%)；ON状态：0 (0.0%)；nan：133 (100.0%)
Test	χ^2 检验
p_{FDR}	4.22×10^{-20}
Cramér's V	0.37
变量：COGCAT _TEXT	
帕金森病计数 (%)	认知主诉：20（3.5%）；痴呆：5（0.9%）；轻度认知障碍：56（9.8%）；正常：93（16.3%）；nan：381（66.8%）
健康对照计数（%）	认知主诉：2（1.5%）；痴呆：0（0.0%）；轻度认知障碍：6（4.5%）；正常：53（39.8%）；nan：72（54.1%）
Test	Fisher/χ^2
p_{FDR}	1.45×10^{-2}
Cramér's V	0.13

注释。 PDSTATE：运动检查时的用药状态（OFF = 停药期；ON = 服用多巴胺能药物后）。COGCAT_TEXT：认知诊断类别（正常、轻度认知障碍、痴呆或自述认知主诉）。“nan”表示PPMI基线数据集中缺少注释。

与非帕金森病状态的关键指标，这有力支持了本研究观察到的运动迟缓相关效应。NHY的显著区分同样与文献结论相符：Hoehn–Yahr分期反映了进行性双侧受累、轴向功能障碍和姿势不稳，这些都是帕金森病的核心诊断和预后指标（Rodriguez-Blazquez, Rojo-Abuin, Alvarez-Sanchez, Arakaki, Bergareche-Yarza, Chade,Garretto, Gershanik, Kurtis,Martinez-Castrillo etal.，2013）。同样，MSEADLG的效应量反映了功能衰退，这与相关研究一致；这些研究将日常生活活动指标与独立能力下降及疾病负担增加的早期迹象联系起来（Lai,Su,Cheng,

Huang 和 Lu，2018；Shulman、Pretzer-Aboff、Anderson、Stevenson、Vaughan、Gruber-Baldini、Reich 和 Weiner，2006）。

面部表情（NP3FACXP）呈现出较大的组间效应（$\delta = 0.79$），这与面具脸——面部表情减弱——作为帕金森病早期且敏感的运动体征相一致，其成因是多巴胺能功能及基底节功能障碍（Maycas-Cepeda、López-Ruiz、Feliz-Feliz、Gómez-Vicente、García-Cobos、Arroyo 和 García-Ruiz，2021）。可解释人工智能研究已证实，面具脸与运动严重程度及诊断效用之间存在显著相关性（Filali Razzouki、Jeancolas、Sambin、Mangone、Chalançon、Gomes、Lehéricy、Vidailhet、Arnulf、Corvol 等人，2025）。用于评估静止性震颤持续性的 NP3RTCON 显示出较大的效应量（$\delta = -0.66$），反映了大多数新诊断帕金森病患者均有的一种典型不对称症状。即使震颤程度较轻，它仍能可靠地区分帕金森病患者与健康对照者，这与早期未用药帕金森病亚型中报告的模式（Hou、Ou、Yang、Song、Gong 和 Shang，2018）、近期关于对侧震颤扩散的描述（Pasquini、Pavese、Ceravolo、Helmich 和 Deuschl，2025），以及肌张力增加的神经生理学证据（Asci、Falletti、Zampogna、Patera、Hallett、Rothwell 和 Suppa，2023；Korkmaz、Yaşa 和 Sonkaya，2025）相符。NP3POSTR评估的姿势异常也显示出较大的效应（$\delta = -0.64$），反映了与疾病严重程度、认知衰退及跌倒风险相关的轴向功能障碍。近期采用Kinect衍生姿势指标的研究证实了这些指标区分帕金森病患者与健康对照者的判别能力，与我们的研究结果一致（Hong、Zhang、Zhang、Wu、Lin、Su、Jin、Gao、Peng、Li 等人，2022；Pokhabov、Tunik、Pokhabov、Katunina 和 Zalyalova，2024）。

僵直分项评分 NP3RIGRU 和 NP3RIGLU 显示出较大的效应（$\delta = -0.61$ 和 $\delta = -0.60$）。肢体僵硬——在疾病早期通常呈不对称分布——是帕金森病的典型运动特征，在 PPMI 的新发患者中，超过 80% 会出现这一症状。在 Rasch 模型研究中，僵直始终是 MDS-UPDRS第三部分中具有区分性的条目，并且与表型分类和疾病进展密切相关（Regnault、Prato、Quéré、Benoit、Massat 和 Morel，2025年；Fereshtehnejad、Zeighami、Dagher 和 Postuma，2017）。生物力学和神经生理学证据也表明，上肢僵硬与总体运动障碍相关（Asci 等，2023年；Falletti、Asci、Zampogna、Patera、Pinola、Centonze、Hallett、Rothwell 和 Suppa，2025年）。总体而言，这些发现揭示了帕金森病中多维运动损害的一致模式，涵盖僵直、运动迟缓、轴性症状和功能独立性。总体运动功能（NP3TOT）、肢体运动及精细运动统一帕金森病评定量表分项评分均显示出较大且一致的效应量，证实它们是早期帕金森病可靠的临床标志物；这也与当代机器学习研究的结论一致，即这些指标对于帕金森病识别和预测建模具有基础性重要意义（Regnault 等，2025年）。

表4展示了在帕金森病组与健康对照组之间存在显著差异的分类变量。PDSTATE 与诊断之间存在很强的关联（Cramér's $V = 0.37$，$p_{\text{FDR}} = 4.22 \times 10^{-20}$），而认知类别分布（COGCAT_TEXT）也呈现较弱但显著的关联（Cramér's $V = 0.13$，$p_{\text{FDR}} = 1.45 \times 10^{-2}$）。这些发现证实，该数据集呈现出帕金森病预期的临床模式，适用于后续预测建模。这些分类差异的临床相关性也得到了充分支持。用药状态（开/关）对测得的运动表现有公认的影响，并且在多模态建模流程中是一个关键的背景特征（Junaid 等人，2023年）；研究表明，用药史有助于提高预测准确率。认知障碍——涵盖帕金森病轻度认知障碍和帕金森病痴呆——同样有充分的文献记载；近期多模态研究表明，将认知状态与运动及功能测量指标相结合，可显著改善预测建模（Junaid 等人，2023年）。这里也体现了相同的模式。

5.2. 提议的 SAFN 模型衍生特征重要性

为了细致解读所提 SAFN 模型的预测行为，我们采用 Gradient ×Input 方法进行了特征级归因分析。与模型无关扰动方法不同，Gradient × Input 通过梯度与输入的逐元素乘积，直接量化模型输出对各输入特征微小变化的敏感度。这为各变量对帕金森病预测概率的贡献程度提供了可微估计。我们在每个交叉验证折的验证划分上计算归因值，并汇总所有样本的结果，再进行归一化，使各特征的贡献总和为 100%；我们将其作为百分比重要性进行报告，以提高可解释性。

图8 展示了影响力最大的前 20 项特征。结果呈现出清晰的规律：临床评估变量在模型决策中占据绝对主导地位，这与模态门控分析的结果一致；在该分析中，临床特征约占融合表示的 60%。四项自主神经症状条目（SCAU26AT、SCAU26CT、SCAU26BT、SCAU26DT）是最强的预测因子，合计贡献了超过 18% 的总重要性。这些条目对应于帕金森病结局量表——自主神经量表（SCOPA-AUT）的关键要素，

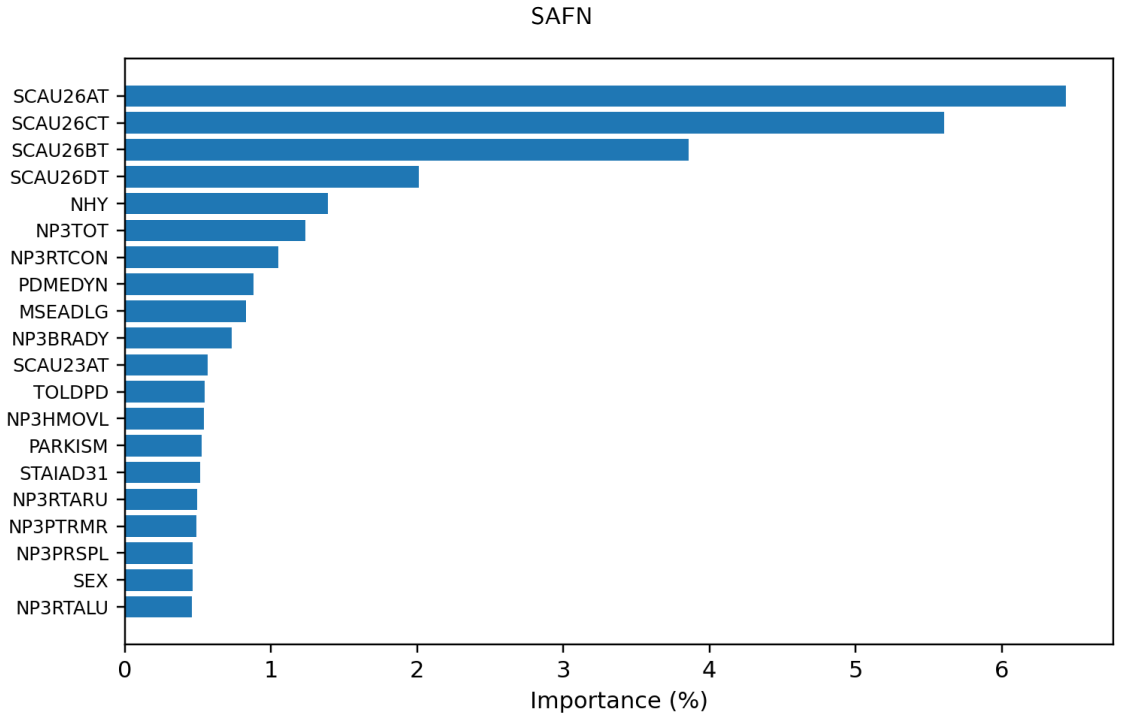


图8： 采用梯度 × 输入归因方法，所提出的类别加权 SAFN 模型识别出的影响力最大的前 20 个特征。条形表示对预测的 PD 概率的特征贡献归一化值，取所有折和样本的平均值。

这与自主神经功能障碍作为帕金森病早期且普遍存在的非运动表现这一公认作用相符。

MDS-UPDRS第三部分中的运动检查变量也表现出显著的重要性。值得注意的贡献包括 NP3TOT（总体运动严重程度）、运动迟缓相关评分（NP3BRADY、NP3HMOVL）、僵直条目（NP3RTCON、NP3RTARU、NP3RTALU）、震颤评分（NP3PTRMR）以及与姿势相关的变量（NP3PRSPL）。这些发现进一步证实了 SAFN 在生物学上的合理性，因为运动障碍仍是帕金森病的核心诊断标志。

全局疾病分期和诊断状态变量，包括 Hoehn–Yahr分期（NHY）、帕金森病用药状态（PDMEDYN）、临床医生报告的帕金森综合征（PARKISM）以及医生确认的诊断（TOLDPD），也显示出较高的重要性。这种模式表明，SAFN 不仅捕捉了细粒度的运动和自主神经变化，还整合了更广泛的疾病背景，以区分帕金森病患者与健康对照者。

有趣的是，心理和行为指标（例如状态-特质焦虑量表中的 STAIAD31）以及人口统计学变量（性别）也具有可测量的贡献。这些关联与已知的流行病学模式一致：焦虑症状常与帕金森病共存，性别差异也已有充分记录，其中男性的帕金森病患病率更高。

总体而言，特征重要性分析表明，SAFN 学习到了具有临床依据的关系：（i）自主神经功能障碍、（ii）运动障碍，以及（iii）整体疾病状态成为影响最大的预测因子，而人口统计学变量和行为变量则提供了次要但互补的信息。这些结果验证了所提出模型的可解释性和临床相关性，并以透明的方式展现了其潜在的决策机制。

5.3. 联合解释：统计效应与学习得到的重要性

为统一呈现数据集中存在的帕金森病-健康对照差异，以及所提出的 SAFN 模型如何利用这些差异，我们将单变量统计组间差异（表3–4）与使用梯度 × 输入（图8）计算得出的模型特征归因进行了比较。尽管这两种分析都考察了帕金森病-健康对照之间的差异，但它们反映了截然不同的视角：统计检验量化的是每个特征的边际组间差异

表5所提出的 SAFN 的平均模态级门控权重。门控值是相互独立的基于Sigmoid的缩放因子。百分比表示四种模态的归一化贡献。

模态	平均门控权重	贡献（%）
临床	0.118	59.9
MRI皮层厚度	0.032	16.2
人口统计学	0.029	14.7
MRI体积	0.018	9.1

而 SAFN 的归因则反映了在完整模型中学习到的多变量、非线性贡献。

两种视角之间呈现出高度一致性。若干单变量组间差异较大的变量——如 NP3TOT、NHY、NP3BRADY、NP3RTCON（静止性震颤持续性）和 MSEADLG——也出现在 SAFN 特征排名的前列。这些变量代表了帕金森病的核心运动和功能特征，包括运动迟缓、僵直、总体运动严重程度和日常生活能力受损。它们在两种分析中均表现出一致性，进一步印证了 SAFN 的临床有效性，并表明该模型优先关注已确立的帕金森病症状。

然而，SAFN 也识别出一些未出现在单变量判别能力排名前列的模式。值得注意的是，SCOPA–AUT量表中的四个自主神经症状条目（SCAU26AT、SCAU26CT、SCAU26BT、SCAU26DT）跻身预测因子之列，尽管它们的边际组间差异仅为中等程度。这体现了多变量深度模型的一项关键优势：单独看来作用较弱、但联合起来具有信息价值的特征，在与相关症状域相互作用时会产生显著影响。自主神经功能障碍被广泛认为是帕金森病早期且异质性较高的非运动表现；它在 SAFN 中的重要性表明，该模型能够检测这些细微的早期信号，并将其与运动领域标志物结合起来。

相反，若干单变量效应显著的特征——例如 NP3FACXP、NP3FTAPR（右手手指敲击）、NP3POSTR 及其他运动分项得分——并未出现在 SAFN排名前20的特征中。这并不意味着这些特征毫无用处；相反，SAFN 可能会降低这些特征的权重，因为它们与已选出的相关运动变量存在冗余，而后者编码了相似的损伤模式。通过标记化、自注意力和门控融合，SAFN 对相关信号进行压缩，并为每个症状簇优先选取最具代表性的一组特征。

最后，PDMEDYN、TOLDPD、PARKISM、STAIAD31 和性别等变量也在归因分析中显示出重要作用。这些变量的单变量差异较弱，但通过与运动症状和自主神经症状的交互作用而变得更为重要。纳入这些变量符合帕金森病中已知的流行病学和行为关联，包括患病率的性别差异、焦虑相关特质、用药状态以及经临床医生确认的诊断指标。

总而言之，这一比较表明，统计分析与基于模型的分析具有很强的互补性。单变量检验突出了PD与健康对照之间最显著的个体 差异，而SAFN则捕捉到更丰富的由交互驱动的 表征，整合了冗余的运动模式、细微的自主神经功能障碍以及情境化的临床信息。这两种视角相互印证，进一步增强了我们对所提出的SAFN模型的可解释性和临床合理性的信心。

5.4. 模态贡献分析

为描述所提出的SAFN如何整合异质信息，我们分析了稀疏注意力门控融合层生成的模态级门控系数。每种模态都被赋予独立的Sigmoid激活，而非Softmax归一化权重，这意味着四个门控值不一定加总为1。为便于解释，我们将每折的原始门控系数除以四个门控值之和进行归一化。表5 报告了平均原始门控权重及其对应的归一化百分比贡献。

表5 中的结果表明，临床模态获得了最高的相对权重（约为60%）。这说明症状严重程度评分、运动和非运动评估、功能测量指标以及用药相关变量，为区分PPMI队列中的帕金森病患者与健康对照提供了最强的判别信息。这一模式符合临床认知：MDS-UPDRS评分、治疗状态和日常生活能力受损，直接反映了帕金森综合征的核心表现。

表6SAFN模型在五折交叉验证下的综合消融研究。该表评估了选择性移除或单独使用各输入模态（MRI皮层厚度、临床、MRI体积测量和人口统计学）以及消融关键架构组件（交叉注意力、模态门控和类别加权）的影响。另将基于拼接特征训练的普通多层感知机作为基线。所有结果均以均值 ± 标准差报告。

模型	准确率	平衡 Acc	ROC 曲线下面积	精确率-召回率曲线下面积	精确率	召回率	F1
普通多层感知机（拼接所有特征）	0.94 ± 0.02	0.92 ± 0.03	0.97 ± 0.03	0.99 ± 0.01	0.97 ± 0.01	0.95 ± 0.02	0.96 ± 0.01
仅临床 SAFN	0.97 ± 0.01	0.93 ± 0.04	0.99 ± 0.00	1.00 ± 0.00	0.97 ± 0.02	0.99 ± 0.01	0.98 ± 0.01
MRI 皮层厚度专用 SAFN	0.68 ± 0.03	0.58 ± 0.04	0.61 ± 0.05	0.87 ± 0.02	0.85 ± 0.02	0.75 ± 0.05	0.79 ± 0.02
SAFN w/o 临床模态	0.73 ± 0.02	0.65 ± 0.03	0.76 ± 0.03	0.94 ± 0.01	0.88 ± 0.02	0.77 ± 0.04	0.82 ± 0.02
不含磁共振成像的 SAFN 皮质-厚度	0.97 ± 0.02	0.95 ± 0.05	1.00 ± 0.01	1.00 ± 0.00	0.98 ± 0.02	0.98 ± 0.01	0.98 ± 0.01
SAFN w/o 跨注意力	0.97 ± 0.01	0.95 ± 0.03	1.00 ± 0.00	1.00 ± 0.00	0.98 ± 0.01	0.98 ± 0.01	0.98 ± 0.01
SAFN w/o 模态门	0.97 ± 0.02	0.95 ± 0.04	1.00 ± 0.01	1.00 ± 0.00	0.98 ± 0.02	0.99 ± 0.01	0.98 ± 0.01
SAFN (no 类别-加权)	0.96 ± 0.02	0.92 ± 0.04	0.99 ± 0.01	1.00 ± 0.00	0.97 ± 0.02	0.98 ± 0.01	0.97 ± 0.01
完整 SAFN（本文提出）	**0.98 ± 0.02**	**0.97 ± 0.05**	**0.98 ± 0.02**	**1.00 ± 0.00**	**0.99 ± 0.03**	**0.99 ± 0.01**	**0.99 ± 0.01**

MRI 皮层厚度和人口统计学模态的权重适中（分别为16%和15%）。虽然在基线时，帕金森病患者与健康对照之间的皮层厚度差异通常较为细微，但与丰富的临床症状表现相结合时，这些差异可提供互补的解剖学线索。同样，年龄、性别和利手等人口统计学因素也能提供背景信息；已知这些因素会影响帕金森病风险和表型表达。

MRI 容积模态的贡献最低（约为 9%）。这表明，全局和局部容积测量指标——颅内容积、灰质/白质体积以及小脑指标——在早期帕金森病中单独使用时的判别信号相对较弱。既往神经影像学文献充分支持这一观察：帕金森病的容积性萎缩模式存在异质性，与疾病进展的关联比与早期病例-对照区分的关联更强。

总体而言，门控分析表明 SAFN 会根据学习到的预测效用自适应地优先选择各模态。临床评估在决策过程中占主导地位，而 MRI（皮层厚度）、人口统计学变量和 MRI（体积测量）则提供互补但次要的信息。这种行为通过揭示不同模态如何影响用于分类的融合表示，提升了预测性能和可解释性。

5.5. 消融研究与架构意义

为评估各模态及所提出 SAFN 模型中各架构组件的作用，本研究采用五折交叉验证进行了全面的消融研究。消融实验包括：(i) 移除或单独使用特定输入模态（MRI皮层厚度、临床、MRI体积和人口统计学），(ii) 移除单个架构组件，例如交叉注意力、模态门控和类别加权，以及 (iii) 将模型与基于拼接特征训练的普通多层感知机基线进行比较。定量结果汇总于表 6。

基线比较。 PlainMLP基线的准确率为 0.94 ± 0.02，ROC 曲线下面积为 0.97 ± 0.03，说明简单拼接虽然性能较强，但明显不如 SAFN。这凸显了 SAFN 各项架构组件的重要性，包括特征标记化、基于注意力的编码和模态感知融合；这些组件共同捕捉到比标准全连接网络更丰富的特征交互。

模态子集的影响。 呈现出明显的模态依赖模式。仅临床 *SAFN* 的性能接近完整模型（准确率 0.97 ± 0.01；F1 分数 0.98 ± 0.01），这反映了临床评估的高判别价值，例如 MDS-UPDRS、PDSTATE、用药状态以及运动/非运动症状。相比之下，仅皮层厚度 *SAFN* 的表现明显较差（ROC 曲线下面积 0.61 ± 0.05），这与既往研究结果一致：在基线时，帕金森病与健康对照之间的结构磁共振成像差异较为细微，通常不足以单独用于分类。

移除所有临床变量（去除临床模态的 *SAFN*）后，平衡准确率显著下降（0.65 ± 0.03），这表明MRI皮层厚度、MRI体积和人口统计学特征提供了互补但判别能力相对较弱的信号。相比之下，移除MRI皮层厚度（去除磁共振成像皮层厚度的 *SAFN*）后，模型性能几乎没有下降，证实在该队列中，临床和人口统计学信息在区分 PD 与 HC 方面占主导地位。

架构组件的影响。 省略交叉注意力或模态门控后，平衡准确率和 F1 分数均出现小幅但持续的下降。尽管编码器能够捕捉强烈的模态内结构，但跨模态交互（通过交叉注意力实现）和自适应加权（通过门控实现）仍带来了可衡量的性能提升。移除类别加权后，平衡准确率从 0.97 降至 0.92，表明在处理 PD/HC 比例失衡时，采用类别平衡优化至关重要。

完整模型性能。 完整 SAFN 在所有度量上均取得了最强且最稳定的性能：准确率 0.98 ± 0.02，平衡准确率 0.97 ± 0.05，F1 分数 0.99 ± 0.01。尽管若干消融模型达到了相近的 ROC 曲线下面积和精确率-召回率曲线下面积值，但这些度量已接近上限并趋于饱和。决定性的改进体现在平衡准确率和 F1 分数上，这证实了完整的架构设计——标记化、注意力编码、交叉注意力、稀疏门控和类别加权——共同带来了最稳健的多模态分类器。

解读。 总体而言，消融研究揭示了三个关键结论：(1) 临床模态提供了主要的判别信号；(2) MRI皮层厚度和 MRI体积特征单独使用时贡献有限，但能够补充临床信息；(3) SAFN 的架构组件（交叉注意力、模态门控、类别加权优化）增强了模型的稳健性和性能，即使相较于强大的仅临床基线模型，改进幅度也较为渐进。

这些发现验证了 SAFN 的设计选择，并证明其适用于对帕金森病分类中的异构多模态生物医学数据进行建模。

5.6. 与现有文献的比较

机器学习和深度学习方法已广泛应用于基于临床、影像、遗传和多模态数据集的帕金森病分类。表 7总结了九项与本研究直接相关的近期代表性研究。这些研究共同凸显了该领域的三大趋势：(i) 临床量表和运动量表具有较强的预测价值；(ii) 对多模态融合的关注日益增加；(iii) 越来越重视可解释性，尤其是对于纳入高维磁共振成像数据的模型。

一些完全基于临床或表格数据的研究已取得出色的性能。例如，Esan 等人（Esan 等人，2025）报告称，利用集成学习处理人口统计学、运动、认知和生活方式特征时，曲线下面积最高可达 0.97，其中统一帕金森病评定量表和蒙特利尔认知评估量表得分成为主要预测因子。这与我们的发现一致：在所提出的 SAFN 模型中，疾病严重程度量表（例如 SCAU 条目、统一帕金森病评定量表第三部分组件和 Hoehn–Yahr分期）也被确定为影响最显著的特征。

近年来，融合结构性磁共振成像和基因组学的复杂多模态研究相继涌现。Li 等人（Li 等人，2025a）提出了 PIDGN，这是一种融合单核苷酸多态性和磁共振成像的混合 Transformer–ResNet 模型，曲线下面积达到 = 0.90；而 Yang 等人（Yang 等人，2025）采用自适应集成堆叠，将磁共振成像形态特征与遗传变异融合，平衡准确率达到 0.95。这些结果展现了结合互补模态的潜力，不过这些方法主要依赖高维磁共振成像或单核苷酸多态性特征，且通常需要复杂的预处理流程。

仅基于磁共振成像的深度学习模型也已得到广泛研究。Camacho 等人（Camacho 等人，2023年）和 Camacho 等人（Camacho 等人，2024）将轻量级 SFCN 网络的变体应用于多中心 T1加权成像和弥散张量成像数据，曲线下面积介于 0.87 至 0.89 之间。Alrawis 等人（Alrawis 等人，2025）提出的 FCN-PD 模型，以及基于 Transformer 的 TransPD-Net（Hussain 等人，2025）等更专业的磁共振成像架构，报告了

表 7采用临床、磁共振成像、遗传及多模态数据进行帕金森病分类的近期研究概述。该表总结了关键模态、模型类型和性能指标，以便将所提出的 SAFN 与最先进的方法进行对比。

研究	数据模态	模型 / 任务	性能说明	
(Esan 等人，2025)	临床（人口统计学信息，运动，认知，生活方式）	表格数据	SVM，KNN，LR，RF，极端梯度提升，堆叠集成；帕金森病 vs 健康对照	准确率 = 0.93，曲线下面积 = 0.97	统一帕金森病评定量表 MoCA and 功能性 量表 最强的 预测因子；使用合成少数类过采样技术
（Li 等人，2025a）	结构 MRI + 单核苷酸多态性遗传学	PIDGN （Transformer + 三维残差网络 + 门控注意力）；帕金森病与健康对照对比	准确率 = 0.86，曲线下面积 = 0.90	SHAP 突出显示 HLA-DRA 及相关 单核苷酸多态性; 中脑 via Grad-CAM
(Dentamaro等人, 2024)	三维 T1 磁共振成像 + 临床（年龄，sex，统一帕金森病评定量表-III）	密集连接网络 + 激励网络；帕金森病 vs 健康对照	准确率 = 0.97，曲线下面积 = 0.99	集成 梯度 显示 前额叶和 颞叶 区域 as 显著
(Yang 等人，2025)	磁共振成像 + 遗传单核苷酸多态性特征	自适应集成堆ing（AE_叠）；PD 与健康对照相比	平衡准确率 = 0.95，F1 = 0.93	SNCA- and VPS52- 相关 变异 most 预测性
（Camacho 等人，2023年）	大型 多中心 T1加权磁共振成像	3D SFCN （轻量级卷积神经网络）；帕金森病 vs 健康对照	准确率 = 0.79，曲线下面积 = 0.87	显著性 图: 额-颞皮层和深部灰质核团
(Camacho 等人，2024)	T1 + 弥散张量成像（各向异性分数、平均扩散率、径向扩散率、轴向扩散率）	三维卷积神经网络（SFCN 变体）；帕金森病 vs 健康对照	准确率 = 0.81，曲线下面积 = 0.89	DTI 度量 more 显著的 than 形态测量学 in 显著性分析
(Alrawis 等人，2025)	磁共振成像（混合局部 + 全局特征）	FCN-PD （U-Net + EfficientNet + 注意力）；帕金森病 vs 健康对照	准确率 = 0.97–0.96 跨数据集	注意力 maps 高光 黑质 and 皮层厚度
(Hussain 等人，2025)	磁共振成像切片（T1 加权成像）	TransPD-Net （CNN + Swin Transformer）；帕金森病 vs HC	准确率 = 0.96，曲线下面积 = 0.96	模式 性能 降至 曲线下面积 ≈ 0.67，针对未见过的数据集
（Basaia等人，2024）	磁共振成像 + 临床数据	3D CNN with 迁移学习；PD 严重程度 vs HC	Acc = 0.64–0.74（轻度 / 中度–重度 vs HC)	类别激活图：脑干、颞-叶、基底神经节
本研究（SAFN）	多模态 MRI（皮层厚度），MRI（体积测量），临床量表，人口统计学信息（PPMI）	表格数据: 类别加权 SAFN with 交叉注意力和门控 多模态融合；帕金森病与健康对照（五折交叉验证）	准确率 = 0.98，曲线下面积 = 0.98	模态 门 and Grad×输入 突出显示 临床量表（统一帕金森病评定量表-III， SCAU Hoehn–Yahr分级） 项目，as 主要的 预测因子，with MRI and 人口统计学信息提供互补线索

准确率更高（0.96–0.97），但在外部数据集上的泛化能力有所下降，这证实了基于磁共振成像的帕金森病分类存在已知的鲁棒性挑战。类似地，Dentamaro等人（Dentamaro等人，2024）和Basaia等人（Basaia等人，2024）等采用多模态磁共振成像与临床数据融合的方法，提升了帕金森病与健康对照的可分性，但依赖计算开销较大的三维卷积神经网络，并且需要大量图像处理。

相较于这些研究，所提出的 SAFN 带来了三项关键进展。首先，SAFN 不再依赖体素级 MRI 或端到端影像组学，而是将精简的表格数据 MRI 衍生形态测量学与临床、人口统计学及 MRI（体积）汇总数据相结合，从而以最少的预处理实现高效的多模态融合。其次，该模型引入了门控融合机制和交叉注意力层，专门用于学习模态间的交互；传统多层感知机和许多早期融合框架均不具备这些能力。第三，SAFN 通过模态级门控和特征级 Grad×Input 归因提供明确的可解释性，从而能够识别具有临床意义的预测因子。由此得到的重要性分布显示，总体运动严重程度与运动迟缓（例如 NP3TOT、NP3BRADY）、轴向/姿势特征以及自主神经 SCAU26 条目是主要贡献因素，这与基线分析中观察到的最显著统计组间差异高度吻合。经典统计分析与模型归因结果的一致性，增强了我们对所学习表征的生理和临床相关性的信心。

在所评述的研究中，SAFN 实现了具有竞争力或更优的分类性能（准确率 $=0.98$，曲线下面积 $=0.98$），同时具备明确的可解释性，且计算开销显著低于大型三维卷积神经网络或基于 Transformer 的 MRI 模型。此外，SAFN 在不同折之间表现出较强的稳健性，特征归因也始终如一，因此进一步区别于以往的多模态帕金森病分类器；这些分类器反复受到模型稳定性或外部泛化能力不足的限制。

总体而言，这项比较分析表明，所提出的 SAFN 通过在单一统一架构中兼顾高准确率、计算效率、多模态敏感性和透明的可解释性，对现有研究形成了补充并推动了其发展。

5.7. 局限性与未来工作

尽管所提出的类别加权 SAFN 在区分帕金森病患者与健康对照参与者方面展现出良好的性能和可解释性，但仍需承认其存在若干局限。

首先，分析基于单一、经过协调处理且规模中等的 PPMI 衍生队列开展（$N = 703$；570 名帕金森病患者、133 名健康对照），且类别分布严重不平衡。尽管采用了类别加权损失（类别平衡焦点损失）来缓解这种偏斜，但依赖单一数据集可能引入队列特异性偏倚，并限制外部泛化能力。因此，未来工作应优先在规模更大、相互独立的多中心队列以及不同扫描仪平台上开展验证，并探索通过领域自适应或联邦学习策略实现跨数据集和跨站点泛化。

其次，本研究聚焦于静态横断面数据，而帕金森病进展本质上具有时间性。对重复评估进行纵向建模，有助于预测疾病轨迹、转化风险和进展速率。通过引入时序架构和不确定性量化扩展 SAFN，将有助于开展进展预测、鉴别诊断和个体化风险分层。

第三，部分模态中的特征存在缺失或仅部分可用，这可能限制了模型充分利用跨模态关系的能力。尽管 SAFN 能够处理不完整的表格输入，未来研究仍应探讨有原则的数据插补、概率建模或不确定性感知融合机制，以更好地应对缺失情况和异质数据可用性。此外，还需处理扫描仪协调和站点特异性变异问题，尤其是在多中心部署时。

第四，磁共振成像衍生输入基于预处理后的区域形态测量和体积汇总，而非原始三维图像。虽然这种设计实现了高效的多模态整合，但可能降低对细微结构改变和微结构改变的敏感度。未来的扩展可纳入轻量级影像骨干网络或直接处理原始磁共振成像或正电子发射断层成像体积数据的视觉变换器，并整合更丰富的模态，例如微结构磁共振成像、电生理学、遗传学或数字表型分析，以捕捉多尺度生物标志物。

最后，尽管 SAFN 通过模态级门控和特征级归因（例如，梯度 $\times$ 输入）提供了全局可解释性，但仍需进一步研究，才能将这些见解转化为患者层面的、可在临床中采取行动的解释。结合以用户为中心的界面设计，SHAP、显著性图或 Grad-CAM 等事后分析技术或许能够生成更清晰、易于理解的解释，以支持临床决策。前瞻性、以用户为中心的评估和真实世界基准测试对于评估工作流程整合、临床可用性和监管就绪度至关重要。

总体而言，未来研究应重点关注：（i）外部验证与跨站点验证；（ii）纵向及时间维度的拓展；（iii）更丰富、更完整的多模态输入，包括原始影像；以及（iv）以患者为中心的可解释性，最终目标是推动 SAFN 发展为稳健、适用于真实临床场景的临床决策支持工具。

6. 结论

本研究提出了类别加权 SAFN，这是一种可解释的多模态深度学习框架，旨在解决帕金森病计算机辅助诊断中的关键局限，包括类别不平衡、模态异质性，以及高维磁共振成像特征掩盖具有临床意义信息的风险。SAFN 将结构磁共振成像皮层厚度、MRI 体积测量、丰富的运动及非运动临床评估和人口统计学数据整合至统一的基于注意力的架构中，实现了符合生物学规律且具有情境感知能力的多模态整合。

该框架的核心是一个经过稀疏性正则化的注意力门控融合层，并结合对称交叉注意力机制，从而能够对异质输入进行选择性且精简的加权。这些组件促进了跨模态的均衡表征学习，并增强了融合模型的可解释性。层次归因分析进一步表明，SAFN能够捕捉具有临床意义的模式，其全局模态加权和细粒度特征归因与既有的PD症状学及诊断原则高度一致。

总体而言，所提出的SAFN框架展示了精心构建的基于注意力的多模态融合在PD分类中的价值，为传统的高维成像流程提供了一种稳健、可解释且计算高效的替代方案。其设计为未来拓展至纵向建模、多中心验证以及临床神经病学中的现实决策支持应用奠定了坚实基础。

致谢

本研究获得人工智能与网络未来中心（AICF）的经费支持；该中心的活动经费主要由澳大利亚查尔斯特大学提供。

声明 写作过程中使用生成式人工智能和人工智能辅助技术

在准备本文期间，作者使用 Grammarly 对稿件进行校对，并改进英文写作的清晰度。在使用此工具/服务后，作者根据需要审阅并编辑了相关内容，并对所发表文章的内容承担全部责任。

参考文献

Afonso, M.M., Edla, D.R., Ramesh, D., 2025. 基于脑电图和斯托克韦尔变换检测帕金森病的基于Transformer的深度学习. Procedia Computer Science 258, 4094–4104. URL: https://www.sciencedirect.com/science/article/ pii/S1877050925017648, doi:10.1016/j.procs.2025.01.764. Almgren, H., Hanganu, A., Camacho, M., Kibreab, M., Camicioli, R., Ismail, Z., Forkert, N.D., Monchi, O., 2023. 帕金森病运动症状与皮层曲率和厚度之间的相互作用有关. NeuroImage: Clinical 37, 103300. doi:10.1016/j.nicl.2022. 103300. Alrawis, M., Mohammad, F., Al Ahmadi, S., Al Muhtadi, J., 2025. FCN PD：一种利用磁共振成像数据诊断帕金森病的先进深度学习框架. Diagnostics 15, 992. URL: https://doi.org/10.3390/diagnostics15080992, doi:10.3390/diagnostics15080992. Asci, F., Falletti, M., Zampogna, A., Patera, M., Hallett, M., Rothwell, J., Suppa, A., 2023. 帕金森病的僵直：来自生物力学和神经生理学测量的证据. Brain 146, 3705–3718. Balıkçı Çiçek, İ., Küçükakçalı, Z., Deniz, B., Algül, F.E., 2025. 利用对比解释方法CEM对帕金森病诊断进行可解释分析. Diagnostics 15, 2069. doi:10.3390/diagnostics15162069. Basaia, S., Sarasso, E., Sciancalepore, F., Balestrino, R., Musicco, S., Pisano, S., Stankovic, I., Tomic, A., Micco, R., Tessitore, A., Salvi, M., Meiburger, K.M., Kostic, V.S., Molinari, F., Agosta, F., Filippi, M., 2024. 利用临床和T1加权磁共振成像数据进行帕金森病诊断与预后的多中心三维卷积神经网络. Diagnostics 14, 51178. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC12351178/. Bunterngchit, C., Baniata, L.H., Albayati, H., Baniata, M.H., Alharbi, K., Alshammari, F.H., Kang, S., 2025. 一种用于准确检测帕金森病的卷积与Transformer混合方法（基于脑电图（EEG））. Bioengineering 12, 583. URL: https://doi.org/ 10.3390/bioengineering12060583, doi:10.3390/bioengineering12060583. Camacho, M., Wilms, M., Almgren, H., Hanganu, A., Kibreab, M., Camicioli, R., Ismail, Z., Forkert, N.D., Monchi, O., 2024. 利用脑宏观与微观结构变化提升基于磁共振成像数据的帕金森病分类效果. npj Parkinson's Disease 10, 43. doi:10.1038/s41531-024-00647-9. Camacho, M., Wilms, M., Mouches, P., Almgren, H., Souza, R., 2023. 在大型多中心T1加权磁共振成像数据集上训练深度学习模型，实现帕金森病的可解释分类. NeuroImage: Clinical 38, 103405. URL: https://doi.org/10.1016/j.nicl. 2023.103405, doi:10.1016/j.nicl.2023.103405. Cao, X., Wang, X., Xue, C., Zhang, S., Huang, Q., Liu, W., 2020. 一种通过整合全脑功能活动和灰质结构预测帕金森病的影像组学方法. Frontiers in Neuroscience 14, 751. doi:10.3389/fnins.2020.00751。

Chen, H., Liu, X., Luo, X., et al., 2024. 一种结合深度学习与影像组学、聚焦中脑和黑质的自动化混合方法，用于检测早期帕金森病. Frontiers in Aging Neuroscience 14, 1397896. URL: https://doi.org/10.3389/fnagi. 2024.1397896, doi:10.3389/fnagi.2024.1397896. Colautti, L., Mastrolia, V., Corbo, M., Nardone, A., Cifelli, V., Giustini, V., Di Giovanni, G., Sancesario, G., Fusco, F.R., 2025. 肠道微生物群与帕金森病关系的系统综述：治疗潜力. Brain Sciences 15, 782. doi:10.3390/brainsci15080782. Cui, Y., Jia, M., Lin, T.Y., Song, Y., Belongie, S., 2019. 基于有效样本数的类别平衡损失，载于：IEEE/CVF计算机视觉与模式识别会议论文集，第9268–9277页. Dentamaro, V., Impedovo, D., Musti, L., Pirlo, G., Taurisano, P., 2024. 通过多模态深度学习和可解释人工智能提升帕金森病早期检测：来自PPMI数据库的见解. Scientific Reports 14, 20941. URL: https://doi.org/10.1038/ s41598-024-70165-4, doi:10.1038/s41598-024-70165-4. Esan, O., Oyewole, O., Babalola, R., Adekoya, A., 2025. 肠道微生物群失调与帕金森病的关联：系统综述与荟萃分析. Current Research in Translational Medicine 73, 103541. doi:10.1016/j.crtrans.2025.103541. Falletti, M., Asci, F., Zampogna, A., Patera, M., Pinola, G., Centonze, D., Hallett, M., Rothwell, J., Suppa, A., 2025. 帕金森病的僵直：左旋多巴的客观效果. Movement Disorders 40, 727–738. Fereshtehnejad, S.M., Zeighami, Y., Dagher, A., Postuma, R.B., 2017. 帕金森病分型的临床标准：生物标志物与纵向进展. Brain 140, 1959–1976. Filali Razzouki, A., Jeancolas, L., Sambin, S., Mangone, G., Chalançon, A., Gomes, M., Lehéricy, S., Vidailhet, M., Arnulf, I., Corvol, J.C., et al., 2025. 阐释面部动作单元与帕金森病面具脸及临床评分的相关性. npj Parkinson's Disease 11, 53. Germani, E., Bhagwat, N., Dugré, M., Gau, R., Montillo, A.A., Nguyen, K.P., Sokolowski, A., Sharp, M., Poline, J.B., Glatard, T., 2025. 利用临床和功能性磁共振成像特征预测帕金森病病程：一项重复实验与重复验证研究. PloS one 20, e0317566. Hong, R., Zhang, T., Zhang, Z., Wu, Z., Lin, A., Su, X., Jin, Y., Gao, Y., Peng, K., Li, L., et al., 2022. 基于Kinect得出的综合指数，用于评估帕金森病患者姿势异常的严重程度. npj Parkinson's Disease 8, 96. Hou, Y., Ou, R., Yang, J., Song, W., Gong, Q., Shang, H., 2018. 早期未用药帕金森病患者不同亚型的纹状体与小脑功能连接模式. Neuroradiology 60, 1323–1333. Hussain, S.S., Shah, P.M., Dawood, H., Xu, D., Alshamayleh, A., Khan, M.A., Ghazal, T.M., 2025. 一种Swin Transformer与卷积神经网络融合框架，用于准确分类磁共振成像中的帕金森病. Scientific Reports 15, 15117. URL: https://doi.org/10.1038/ s41598-025-93671-5, doi:10.1038/s41598-025-93671-5. Islam, N.U., Khanam, R., 2024. 基于机器学习的帕金森病检测与进展评估：利用三维磁共振成像中的灰质和白质分割. Journal of Engineering and Applied Science 71, 522. doi:10.1186/s44147-024-00522-7. Jibon, F.A., Tasbir, A., Talukder, M.A., et al., 2024. 利用基于自编码器和径向基函数神经网络的混合深度学习框架及功率谱密度，从脑电信号中检测帕金森病. Digital Health 10. URL: https://doi.org/10.1177/20552076241297355, doi:10.1177/20552076241297355. Jin, Y., Li, X., Han, X., Qiu, Y., Zhang, H., Xu, J., Han, M., 2025. 基于SHAP的可解释机器学习用于帕金森病严重程度预测：临床和环境特征的综合分析. Frontiers in Neurology 16, 1678463. doi:10.3389/fneur.2025. 1678463. Junaid, M., Ali, S., Eid, F., El-Sappagh, S., Abuhmed, T., 2023. 基于多模态时间序列数据的可解释机器学习模型，用于帕金森病的早期检测. Computer Methods and Programs in Biomedicine 234, 107495. Kempster, P.A., 2025. 了解帕金森病的进展：综述. BMJ Neurology Open 7, e001215. URL: https: // neurologyopen.bmj.com/content/bmjno/7/2/e001215.full.pdf, doi:10.1136/bmjno-2025-001215. Khan, A.Z., Lavu, D., Neal, R.D., 2024. 帕金森病：关于其在初级保健中诊断准确性的定量和定性证据的范围综述. British Journal of General Practice 74, e227–e232. doi:10.3399/BJGP.2023.0409. Korkmaz, B., Yaşa, M.E., Sonkaya, R., 2025. 帕金森病患者的上肢功能、脊柱姿势和轴性强直. Acta Neurologica Belgica 125, 119–126. Lai, Y.R., Su, Y.J., Cheng, K.Y., Huang, C.C., Lu, C., 2018. 与帕金森病患者生活质量相关的临床因素. Neuropsychiatry 8, 119–125. Li, H., et al., 2024a. 基于深度学习的帕金森病影像检测与分类. BMC Medical Imaging 24, 1364. URL: https: //doi.org/10.1186/s12880-024-01364-8, doi:10.1186/s12880-024-01364-8. Li, J., Liu, X., Wang, X., Liu, H., Lin, Z., Xiong, N., 2022. 用于帕金森病诊断的弥散张量成像影像组学. Brain Sciences 12, 851. doi:10.3390/brainsci12070851. Li, K., Ao, B., Wu, X., Wen, Q., Ul Haq, E., Yin, J., 2024b. 基于深度CNN-LSTM模型的脑电图帕金森病检测与分类. Biotechnology & Genetic Engineering Reviews 40, 2577–2596. doi:10.1080/02648725.2023.2200333. Li, W., Rao, Q., Dong, S., Zhu, M., Yang, Z., Huang, X., Liu, G., 2025a. PIDGN：一种用于帕金森病早期预测的可解释多模态深度学习框架. Journal of Neuroscience Methods 415, 110363. doi:10.1016/j.jneumeth.2025.110363. Li, X., Jin, X., Wang, J., Yu, Y., Zhang, R., 2025b. 绘制早发型帕金森病的全球疾病负担：1990–2021年的地域和社会人口学不平等. Frontiers in Public Health 13, 1618533. URL: https://doi.org/10.3389/fpubh.2025.1618533, doi:10.3389/ fpubh.2025.1618533. Mahajan, R.P., 2025. 利用磁共振成像识别帕金森病的深度学习技术. https://www.semanticscholar.org/paper/Deep-Learning-Techniques-for-Identification-of-on-Mahajan-Jain/557ad9757e2ce2ca86edb1087210348247fa6273. Advance online publication. Mattia, G.M., Chougar, L., Foubert-Samier, A., Meissner, W.G., Fabbri, M., Pavy-Le Traon, A., Rascol, O., Grabli, D., Degos, B., Pyatigorskaya, N., Faucher, A., Vidailhet, M., Corvol, J.C., Lehéricy, S., Péran, P., 2025a. 利用多模态磁共振成像通过深度学习区分帕金森综合征：一项概念验证研究. Movement Disorders 40, e30300. URL: https://movementdisorders.

onlinelibrary.wiley.com/doi/10.1002/mds.30300, doi:10.1002/mds.30300. Mattia, G.M., Chougar, L., Foubert Samier, A., Meissner, W.G., Fabbri, M., Pavy Le Traon, A., Rascol, O., Péran, P., 2025b. 利用多模态磁共振成像鉴别帕金森综合征的深度学习：一项概念验证研究。《Movement Disorders》40, e30300. URL: https://doi.org/10.1002/mds.30300, doi:10.1002/mds.30300. Maycas-Cepeda, T., López-Ruiz, P., Feliz-Feliz, C., Gómez-Vicente, L., García-Cobos, R., Arroyo, R., García-Ruiz, P.J., 2021. 帕金森病中的面具脸：它向我们传递了什么信息？Frontiers in Neurology 11, 603582. Pasquini, J., Pavese, N., Ceravolo, R., Helmich, R.C., Deuschl, G., 2025. 帕金森病中不对称性震颤的对侧扩散。《Movement Disorders Clinical Practice》. Peng, S., Liu, P., Wang, X., Li, K., 2025. 55岁以上人群帕金森病的全球、区域和国家疾病负担：对1991–2021年全球疾病负担研究的系统分析。《BMC神经病学》25. URL: https://doi.org/10.1186/s12883-025-04191-8, doi:10.1186/s12883-025-04191-8. Pokhabov, D., Tunik, M., Pokhabov, D., Katunina, E., Zalyalova, Z., 2024. 帕金森病患者的姿势稳定性与步态。《神经科学与行为生理学》54, 410–413. Regnault, A., Prato, M.K., Quéré, S., Benoit, A., Massat, N.J., Morel, T., 2025. 优化早期帕金森病的MDS-UPDRS第三部分：支持肢体相关运动迟缓/僵直子评分的初步证据。npj Parkinson's Disease 11, 297. Rodriguez-Blazquez, C., Rojo-Abuin, J.M., Alvarez-Sanchez, M., Arakaki, T., Bergareche-Yarza, A., Chade, A., Garretto, N., Gershanik, O., Kurtis, M.M., Martinez-Castrillo, J.C., et al., 2013. MDS-UPDRS第二部分（日常生活的运动体验）有助于评估帕金森病患者的残疾程度。《帕金森综合征及相关障碍》19, 889–893. Sar, A., Puri, P.S., Naz, H., Al Khudhairy, A., 2025. 用于帕金森病早期检测的多模态深度学习框架：结合神经学和生理数据实现高保真诊断。Scientific Reports 15, 34835. URL: https://doi.org/10.1038/ s41598-025-21407-6, doi:10.1038/s41598-025-21407-6. Shi, D., Zhang, X., Wang, Y., Wang, Q., Yao, M., Li, Z., Guo, L., Zheng, D., Ren, J., 2022. 基于ALFF影像组学特征的帕金森病机器学习检测：潜在的神经影像学生物标志物。Frontiers in Aging Neuroscience 14, 806828. doi:10.3389/fnagi.2022. 806828. Shulman, L.M., Pretzer-Aboff, I., Anderson, K.E., Stevenson, R., Vaughan, C.G., Gruber-Baldini, A.L., Reich, S.G., Weiner, W.J., 2006. 帕金森病患者日常生活活动的主观报告与客观测量。《Movement Disorders》21, 794–799. Su, D., Wei, S., Cao, S., Zhang, M., Li, F., Xu, Y., Zhou, W., Shang, X., Luo, L., Yue, W., Miao, M., 2025. 预测至2050年195个国家和地区的帕金森病患病率及其驱动因素：基于2021年全球疾病负担研究的建模研究。《英国医学杂志》388, e080952. URL: https://doi.org/10.1136/bmj-2024-080952, doi:10.1136/bmj-2024-080952. Twala, B., 2025. 帕金森病人工智能驱动的精准诊疗：综述与实验分析。Frontiers in Aging Neuroscience 17, 1638340. URL: https://www.frontiersin.org/articles/10.3389/fnagi.2025.1638340/full, doi:10.3389/fnagi.2025.1638340. Welton, T., Hartono, S., Lee, W., Teh, P.Y., Hou, W., Chen, R.C., Chen, C., Lim, E.W., Prakash, K.M., Tan, L.C., Tan, E.K., Chan, L.L., 2024. 基于中脑磁共振成像深度学习的帕金森病分类。Frontiers in Aging Neuroscience 16, 1425095. URL: https://www.frontiersin.org/articles/10.3389/fnagi.2024.1425095/full, doi:10.3389/fnagi.2024.1425095. Yang, Y., Hu, L., Chen, Y., Gu, W., Lin, G., Xie, Y., 2025. 利用PPMI队列的磁共振成像和遗传数据识别帕金森病：改进的机器学习融合方法。Frontiers in Aging Neuroscience 17, 1510192. URL: https://doi.org/10.3389/fnagi. 2025.1510192, doi:10.3389/fnagi.2025.1510192. Zhao, S., Dai, G., Li, J., Zhang, Y., Liu, T., Wang, Y., Wang, X., Lu, B., 2024. 基于图学习的可解释模型：利用与语音相关的脑电图诊断帕金森病。npj Digital Medicine 7, 3. doi:10.1038/s41746-023-00983-9. Zhao, W., Wang, Z., Zhang, H., 2022. 基于卷积神经网络和迁移学习的高分辨率磁化率图中脑结构自动分割。Frontiers in Neuroscience 16, 801618. doi:10.3389/fnins.2022.801618. Zheng, L., Zhou, C., Mao, C., , et al., 2024. 对比机器学习揭示与帕金森病疾病严重程度和进展相关的疾病特异性特征。《通信生物学》7, 954. doi:10.1038/s42003-024-06648-x. Zhou, X., Zhu, H., Wang, X., Gao, Q., 2025. 基于脑成像特征的帕金森病相关脑衰老预测模型。Frontiers in Neurology 16, 1584226. doi:10.3389/fneur.2025.1584226.