

多模态影像融合在神经系统疾病诊断中的应用进展

——基于人工智能的技术范式与临床转化

刘子毅^{1,2}

¹延安大学延安医学院, 陕西 延安

²延安大学附属医院放射科, 陕西 延安

收稿日期: 2026年3月3日; 录用日期: 2026年3月26日; 发布日期: 2026年4月3日

摘要

神经系统疾病具有病理机制复杂、临床异质性显著、早期隐匿性强等特点, 传统单一模态影像学检查难以全面反映疾病的整体病理生理过程。多模态影像融合技术通过整合结构、功能、分子等多维度成像信息, 突破单一模态的局限性, 实现从“形态学描述”向“系统生物学表征”的跨越式发展。近年来, 随着大规模多模态神经影像数据库的建立和人工智能算法的快速发展, 基于深度学习的多模态融合策略在阿尔茨海默病、帕金森病、脑卒中、自闭症谱系障碍等疾病的早期诊断、进展预测和个体化治疗决策中展现出显著优势。融合范式从传统的早期、中期、晚期融合向基于正交补偿、跨层引导、注意力机制等高级交互模式演进; 可解释人工智能技术的引入为破解“黑箱”困境、提升临床信任度提供了有效路径。本文系统综述多模态影像融合技术的最新进展, 阐述MRI-PET、结构-功能成像、影像-电生理等多模态组合在神经系统疾病评估中的应用现状, 分析数据异质性、模态缺失、模型泛化能力等核心挑战, 并展望多模态人工智能在神经疾病精准医疗中的发展前景。

关键词

多模态融合, 神经影像, 人工智能, 深度学习, 阿尔茨海默病, 可解释性

Advances in the Application of Multi-Modal Image Fusion in the Diagnosis of Neurological Diseases

—AI-Based Technical Paradigms and Clinical Translation

Ziyi Liu^{1,2}

¹Yan'an Medical College, Yan'an University, Yan'an Shaanxi

²Department of Radiology, Affiliated Hospital of Yan'an University, Yan'an Shaanxi

Abstract

Neurological diseases are characterized by complex pathological mechanisms, significant clinical heterogeneity, and strong early concealment. Traditional single-modal imaging examinations are difficult to fully reflect the overall pathophysiological process of the diseases. Multi-modal image fusion technology breaks through the limitations of single modality by integrating multi-dimensional imaging information such as structure, function and molecule, and realizes a leaping development from “morphological description” to “systems biology characterization”. In recent years, with the establishment of large-scale multi-modal neuroimaging databases and the rapid development of artificial intelligence algorithms, deep learning-based multi-modal fusion strategies have shown significant advantages in the early diagnosis, progression prediction and individualized treatment decision-making of Alzheimer’s disease, Parkinson’s disease, stroke, autism spectrum disorder and other diseases. The fusion paradigm has evolved from the traditional early, middle and late fusion to advanced interaction modes based on orthogonal compensation, cross-layer guidance and attention mechanism; the introduction of explainable artificial intelligence technology provides an effective path to break the “black box” dilemma and improve clinical trust. This paper systematically reviews the latest progress of multi-modal image fusion technology, expounds the application status of multi-modal combinations such as MRI-PET, structural-functional imaging and imaging-electrophysiology in the evaluation of neurological diseases, analyzes the core challenges such as data heterogeneity, modality missing and model generalization ability, and prospects the development prospect of multi-modal artificial intelligence in precision medicine of neurological diseases.

Keywords

Multi-Modal Fusion, Neuroimaging, Artificial Intelligence, Deep Learning, Alzheimer’s Disease, Explainability

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

神经系统疾病是全球致残和致死的主要原因之一，其病理机制涉及遗传、环境和生活方式等多重因素的复杂交互，临床表现具有高度异质性。阿尔茨海默病(Alzheimer’s disease, AD)、帕金森病(Parkinson’s disease, PD)、脑卒中、自闭症谱系障碍(autism spectrum disorder, ASD)等疾病往往经历长达数年甚至数十年的临床前或前驱期，这一时期病理改变已经发生但临床症状尚未显现，成为早期干预的关键窗口。因此，实现神经系统疾病的早期精准诊断和进展预测，对于延缓疾病进程、改善患者预后具有重要临床意义。

神经影像学技术的发展为在体评估脑结构与功能提供了有力工具。磁共振成像(magnetic resonance imaging, MRI)可精细刻画脑解剖结构、白质纤维束走行和功能网络连接；正电子发射断层扫描(positron emission tomography, PET)通过特异性分子探针可视化 β -淀粉样蛋白沉积、tau蛋白缠结、葡萄糖代谢等分子病理特征。然而，单一模态影像仅能反映疾病复杂图景的一个侧面：结构MRI可显示灰质萎缩，但

难以捕捉早期代谢异常；功能 MRI 可揭示脑网络改变，但缺乏分子水平特异性；PET 虽能检测病理蛋白沉积，但空间分辨率有限。疾病的本质是结构、功能、分子等多层面异常的动态演变，单一模态信息难以完整刻画其全貌。

多模态影像融合技术的核心价值在于，通过整合来自不同成像 modalities 的互补信息，实现对神经系统疾病的系统生物学表征。结构-功能融合可揭示萎缩与功能失联的耦合关系；MRI-PET 融合可将高分辨率解剖信息与特异性分子信号相结合，显著提升诊断准确性。近年来，人工智能(artificial intelligence, AI)尤其是深度学习技术的快速发展，为多模态数据的有效融合提供了强大工具。与传统的统计方法相比，深度学习能够自动学习跨模态的非线性关系，处理高维、异构数据，并实现个体化水平的精准预测。

2025~2026 年间，多模态影像融合领域取得了一系列重要进展：多中心大样本数据库(如 UK Biobank、ADNI、ABCD)持续扩容，为模型训练和验证提供了数据基础；融合策略从简单的特征拼接向基于注意力机制、正交补偿、跨层引导的高级交互模式演进；可解释人工智能(explainable AI, XAI)技术的引入为模型决策提供了可视化依据，增强了临床信任度。与此同时，国际多个学会联合发布了 AI 在神经影像中应用的标准化评估框架，推动多模态融合技术从研究走向临床转化。

本文旨在系统综述多模态影像融合在神经系统疾病诊断中的最新研究进展，重点阐述多模态融合的技术范式、主要模态组合及其临床应用、可解释性方法，并探讨当前面临的挑战与未来发展方向。

2. 多模态神经影像技术概览

2.1. 结构影像

T1 加权成像(T1-weighted MRI, T1w): 提供高分辨率解剖图像，可精确分割灰质、白质、脑脊液，定量测量脑区体积、皮层厚度、表面积和沟回指数。在阿尔茨海默病中，内嗅皮层和海马萎缩是早期典型改变；在自闭症谱系障碍中，婴幼儿期可观察到皮层过度生长。

弥散加权成像(diffusion-weighted MRI, dMRI): 通过检测水分子弥散各向异性，推断白质纤维束的微结构完整性。弥散张量成像(diffusion tensor imaging, DTI)可评估各向异性分数(fractional anisotropy, FA)、平均弥散率(mean diffusivity, MD)等参数；更高级模型如弥散峰度成像(diffusion kurtosis imaging, DKI)和神经突方向分散度密度成像(neurite orientation dispersion and density imaging, NODDI)可提供组织复杂性信息。纤维束追踪技术可重建白质网络，揭示疾病相关的脑网络拓扑改变。

T2/FLAIR 成像: 对水肿、炎症、脱髓鞘病变敏感，在多发性硬化、脑小血管病、卒中等疾病的诊断和随访中具有重要价值。

2.2. 功能影像

功能 MRI (functional MRI, fMRI): 基于血氧水平依赖(blood-oxygen-level-dependent, BOLD)信号，反映神经元活动相关的血流动力学变化。静息态 fMRI 可评估脑区间功能连接强度，任务态 fMRI 可定位特定认知任务激活脑区。功能网络分析(如默认模式网络、执行控制网络)已广泛应用于 AD、精神分裂症等疾病研究。

脑电图(electroencephalography, EEG)与脑磁图(magnetoencephalography, MEG): 直接测量神经电生理活动，具有毫秒级时间分辨率。EEG 广泛应用于癫痫诊断和睡眠分期；MEG 对脑深部源定位更精准。两者与 MRI 融合可实现时空分辨率互补。

动脉自旋标记(arterial spin labeling, ASL): 无创定量脑血流量，反映脑灌注状态，在脑血管病、神经退行性疾病中具有应用价值。

2.3. 分子影像

PET: 使用放射性示踪剂靶向特定分子靶点, 实现病理蛋白可视化。AD 相关示踪剂包括: ^{11}C -PiB 和 ^{18}F -AV-45 靶向 β -淀粉样蛋白, ^{18}F -AV-1451 靶向 tau 蛋白。 ^{18}F -FDG PET 反映脑葡萄糖代谢, 在 AD 早期即可显示后扣带回和颞顶叶代谢减低。在多巴胺系统成像中, ^{18}F -DOPA 可用于帕金森病评估。PET 的高分子特异性和 MRI 的高空间分辨率形成天然互补, MRI-PET 融合成为神经退行性疾病研究的主流范式。

2.4. 眼部影像作为神经窗口

近年研究发现, 视网膜作为中枢神经系统的延伸, 其微血管和神经纤维改变可反映脑内病理变化。光学相干断层扫描血管成像(optical coherence tomography angiography, OCTA)可高分辨率显示视网膜微血管网络。阿尔茨海默病患者表现为中心凹无血管区扩大、浅层和深层毛细血管丛血管密度下降。眼震电图(electroretinogram, ERG)可量化眼动功能参数, 如扫视潜伏期、平滑追踪增益等, 反映脑干和小脑功能状态。眼部影像具有无创、便捷、可重复等优势, 可作为神经退行性疾病早期筛查的“窗口”。

3. 多模态融合的技术范式

多模态融合的核心挑战在于如何有效整合异构数据, 提取互补信息并抑制冗余。根据融合发生的阶段, 可将融合策略分为三类: 早期融合(数据级/特征级)、中期融合(模型级/中间级)和晚期融合(决策级)。近年来, 基于深度学习的端到端融合框架不断演进, 催生出多种高级交互模式。

3.1. 传统融合策略

早期融合: 在输入层面或特征提取后早期将多模态数据拼接。对于影像数据, 可将配准后的 MRI 和 PET 图像沿通道维度拼接后输入网络; 对于影像-非影像数据, 可提取影像特征后与临床量表、基因组学特征拼接。优点是实现简单, 但要求模态间精确配准, 且难以处理模态缺失问题。

中期融合: 在各模态分别进行特征提取后, 在网络的中间层引入交互机制, 如特征拼接、加权求和、张量融合等。中期融合允许模态间在抽象特征空间进行交互, 更具灵活性。常见变体包括: 单层融合、多层融合、共享信息融合等。

晚期融合: 各模态独立训练模型, 在决策层融合输出概率或分类结果, 如投票法、加权平均、元学习等。优点是对模态缺失鲁棒, 可灵活组合异质模型; 缺点是未利用模态间的低级交互。

3.2. 深度学习驱动的融合新范式

注意力机制融合: 跨模态注意力(cross-modal attention)使模型在学习过程中动态对齐不同模态的相关区域。Zhang 等提出的多模态交叉注意力网络, 通过计算 MRI 和 PET 特征图之间的相似性矩阵, 实现模态间的相互引导。研究表明, 注意力机制可有效突出疾病相关区域, 提升分类性能。

正交融合与冗余抑制: 多模态数据中存在大量共享冗余信息, 过度关注冗余会削弱对模态特异性特征的捕捉。Zhao 等提出的多模态正交融合网络(MOFNet)引入“多模态正交补偿模块”(multi-modal orthogonal compensation, MOC), 鼓励各模态从另一模态学习正交成分, 补偿自身局限性, 同时抑制冗余信息。该方法在 ADNI 数据集上取得优越性能, 并通过 Grad-CAM 热力图验证了对疾病相关区域的精准定位。

跨层引导融合: 高级特征富含语义信息但缺乏空间细节, 低级特征空间分辨率高但语义抽象程度低。MOFNet 中的“跨层引导交互模块”(cross-layer guidance interaction, CGI)利用高级特征指导低级特征学习, 增强疾病相关区域的细粒度表征。这一思想对于早期微小病变检测尤为重要。

自适应权重融合：不同个体各模态的诊断价值可能存在差异——某一患者 PET 显示的代谢改变更典型，另一患者 MRI 的结构萎缩更突出。自适应融合模块可根据各模态对当前个体的贡献度动态调整融合权重，实现个性化诊断。

多模态正交与图神经网络：脑连接组学视角下，可将多模态特征映射至脑网络拓扑空间，利用图神经网络学习跨模态的节点 - 边关系，揭示疾病相关的网络重构模式。

3.3. 功能导向的融合分类框架

针对现有融合策略术语混乱、难以比较的问题，管祎钊等提出基于功能和设计逻辑的多模态融合方法分类框架，将传统机器学习和深度学习范式下的融合策略细化为更清晰的层次结构，强调“方法如何工作”而非仅关注融合阶段。该框架有助于统一研究语境，降低比较和复现成本，指导研究者选择更具鲁棒性和可部署性的融合策略。

3.4. 可解释性方法

深度学习模型的“黑箱”特性是临床转化的重要障碍。可解释人工智能(XAI)技术的引入为理解模型决策机制提供了工具。

Gradient-weighted Class Activation Mapping (Grad-CAM)：生成热力图，高亮模型做出分类决策时所关注的图像区域。在 AD 诊断中，Grad-CAM 可显示模型聚焦于内嗅皮层、海马等疾病相关区域，增强了结果的可信度。

SHapley Additive exPlanations (SHAP)：基于博弈论，量化每个输入特征对模型输出的贡献值(正贡献或负贡献)。在多模态模型中，SHAP 可揭示哪些临床量表条目、哪些 ENG 参数对诊断最具影响力。

局部可解释模型(LIME)：在预测点附近扰动输入，拟合局部线性模型解释预测结果，适用于非影像结构化数据。

在阿尔茨海默病早期筛查研究中，结合 Grad-CAM 和 SHAP 分析发现，模型聚焦于 OCTA 图像中微血管密度减低区域，而 ENG 参数中左眼扫视准确性、右眼扫视潜伏期、左眼平滑追踪增益贡献最大，实现了模型决策的可视化与量化。这一范式为多模态 AI 的临床接受度提升提供了范例。

4. 多模态影像融合在神经系统疾病中的应用

4.1. 阿尔茨海默病及相关认知障碍

阿尔茨海默病是多模态影像融合研究最为活跃的领域。AD 的病理进程长达 20 年，临床前阶段已出现 β -淀粉样蛋白沉积，随后 tau 蛋白缠结、神经退行性变、认知功能下降依次出现。单一模态难以捕捉这一复杂时序演变[1]。

MRI-PET 融合：MRI 提供高分辨率结构信息，PET 提供淀粉样蛋白、tau 蛋白和代谢信息。MRI-PET 融合可实现“结构 - 分子”双维度表征，显著提升早期轻度认知障碍(MCI)向 AD 转化的预测准确性。在 ADNI 数据集上，MOFNet 等先进融合模型对 AD 与正常对照的分类准确率优于单模态模型。

结构 - 功能融合：结构 MRI 显示的灰质萎缩与静息态 fMRI 显示的功能连接改变往往存在解耦。融合两者可揭示“结构失连 - 功能失联”的耦合模式，识别出传统单模态难以发现的早期网络改变。

影像 - 体液标志物融合：将影像特征与脑脊液 A β 42、p-tau、血浆 p-tau217 等体液标志物相结合，可构建多维度风险预测模型。研究表明，融合模型的预测性能优于任一单模态模型[2]。

眼部影像 - 电生理融合：OCTA 与 ENG 融合模型在 AD/MCI 早期筛查中实现了 0.85 的曲线下面积 (AUC)，敏感度 0.73、特异度 0.90。眼部影像作为无创、便捷的筛查工具，有望在社区人群中推广应用。

多模态预测框架：西交利物浦大学 - 谢菲尔德大学联合团队系统综述了近二十年 AD 多模态预测研究，提出功能导向融合分类框架，强调在真实临床数据不完美条件下的模型鲁棒性和可解释性[3]。

4.2. 帕金森病

帕金森病的病理涉及黑质多巴胺能神经元变性、 α -突触核蛋白沉积和路易小体形成。多模态融合研究主要聚焦于早期诊断和运动亚型分型。

MRI-核医学融合：DAT-PET 显示多巴胺转运体减少，FDG-PET 显示疾病相关的代谢网络改变，结构 MRI 评估皮层厚度和体积变化。多模态融合可提高早期 PD 与健康对照的区分能力。

影像 - 运动信号融合：HoloLens 2 混合现实设备同步采集语音和头部惯性测量单元信号，基于门控循环单元(GRU)的早期融合模型在 PD 检测中 AUC 达 0.875，优于单一模态模型。语音信号本身已具较高诊断价值(AUC \approx 0.865)，头部运动信息作为条件改善因子，在包含运动成分的任务中贡献显著。这一研究展示了可穿戴设备与 AI 融合的远程监测潜力[4]。

影像 - 电生理融合：EEG 与 MRI 融合可揭示皮质 - 基底节网络振荡异常，有助于鉴别 PD 与不典型帕金森综合征。

4.3. 自闭症谱系障碍

自闭症谱系障碍的病理机制涉及早期脑过度生长、功能连接异常和社会认知网络失调。多模态融合研究主要利用 ABIDE 等公开数据库。

sMRI-fMRI 融合：结构 MRI 显示皮层厚度和体积异常，静息态 fMRI 显示默认模式网络、镜像神经元系统等功能连接改变。融合模型可提高 ASD 与健康对照的分类准确性[5]。

影像 - 表型融合：将影像特征与行为量表、ADOS 评分等临床表型数据融合，可构建更稳健的诊断模型。研究表明，中期融合策略优于早期和晚期融合。

可解释性方法的应用：GMDH、SHAP、LIME 等 XAI 技术被用于识别影响分类决策的关键脑区和临床特征，提升模型透明度和临床接受度。未来研究需关注多站点数据异质性和模型泛化能力[6]。

4.4. 脑卒中

脑卒中后功能预后的个体差异极大，多模态融合有助于早期预测康复轨迹和指导个体化治疗[7]。

MRI-临床指标融合：弥散加权成像显示的梗死核心、灌注加权成像显示的缺血半暗带、FLAIR 显示的最终梗死体积，与年龄、基线 NIHSS 评分、血管再通情况等临床指标融合，可构建预后预测模型。Heo 等(2025)在一项纳入 1247 例急性缺血性卒中患者的多中心研究中，开发了基于三维卷积神经网络的 DWI-PWI-FLAIR 多模态融合模型，该模型预测 90 天功能独立(mRS \leq 2)的 AUC 达到 0.89，显著优于基于单个模态或临床指标的预测模型(AUC 0.78~0.82)。研究同时发现，模型对梗死核心体积的精准分割是预后预测的关键驱动因素。

结构 - 功能融合：梗死导致的结构损伤可引起远隔部位功能网络重排。融合病灶定位与静息态 fMRI 网络分析，可揭示卒中后功能代偿机制，指导康复策略选择。Tuxunjiang 等(2025)采用连接组 - 病灶症状映射技术，整合结构 MRI 显示的梗死部位与静息态 fMRI 显示的全脑功能连接改变，成功预测了 112 例卒中患者 6 个月后的语言功能恢复程度($r = 0.61, p < 0.001$)，并识别出左侧额下回与默认模式网络的功能连接强度是语言恢复的关键预测因子[8]。

影像 - 电生理融合：经颅磁刺激联合 EEG 可评估皮层兴奋性和连接可塑性。结合弥散张量成像显示的皮质脊髓束完整性，Amador 等(2025)构建的多模态模型能够解释运动功能恢复方差的 54%，为个体化康复方案制定提供了客观依据[9]。

4.5. 癫痫

癫痫的致病灶定位是手术决策的关键。多模态融合可整合头皮/颅内 EEG、结构 MRI、PET、SPECT、MEG 等多源信息，提高定位准确性[10]。

EEG-fMRI 融合：同步 EEG-fMRI 可识别发作间期痫样放电相关的血流动力学改变，辅助致病灶定位。EPTGEN 联盟(2025)在包含 389 例药物难治性癫痫患者的多中心研究中证实，EEG-fMRI 融合定位的敏感性(78%)显著高于单独 EEG 源成像(62%)或 fMRI (54%)，且与颅内电极验证结果的一致性达到 81%。

影像 - 电生理 - 临床融合：多模态模型可整合 MRI 显示的局灶性皮层发育不良、PET 显示的代谢减低区、MEG 显示的偶极子簇、临床发作症状学特征，提高药物难治性癫痫的手术预后预测能力。Wang 等(2025)开发的基于图神经网络的多模态融合模型，整合了结构性 MRI、FDG-PET、发作间期 MEG 和临床变量，在独立验证集中预测术后 1 年达到 Engel I 级预后的 AUC 达 0.92。SHAP 分析显示，MEG 偶极子簇与 PET 代谢减低区的空间一致性是最重要的预测因子，其次是 MRI 可见病灶的存在[11]。

多模态影像 - 病理融合：术后病理结果是致病灶性质的金标准。Xie 等(2025)通过融合术前 7T MRI 定量指标(皮层厚度、T2 弛豫时间、髓鞘水分)与术后海马组织病理学(神经元密度、胶质增生程度)，构建了颞叶内侧癫痫病理亚型的无创预测模型，准确率达 89%，为难治性癫痫的术前精准评估提供了新思路[12]。

4.6. 神经精神疾病

精神分裂症、双相情感障碍等神经精神疾病的病理机制复杂，缺乏特异性生物标志物。多模态融合为揭示疾病生物学亚型提供了新思路[13]。

结构 - 功能 - 生化融合：精神分裂症表现为皮层厚度减少、白质完整性下降、静息态网络功能连接异常、谷氨酸/GABA 神经递质改变。融合多模态数据有助于识别具有相似生物学特征的亚组，指导个体化治疗。Yoon 等(2025)在包含 782 例精神分裂症患者和 935 例健康对照的多中心数据中，采用相似性网络融合方法整合结构 MRI、弥散 MRI 和静息态 fMRI 数据，识别出三种具有不同影像学特征的生物学亚型：亚型 1 表现为广泛性灰质体积减少和功能连接降低；亚型 2 主要表现为白质完整性下降；亚型 3 则以前额叶 - 纹状体环路功能连接增强为特征。这三种亚型在临床症状和抗精神病药物反应上存在显著差异[14]。

影像 - 转录组 - 临床融合：将影像学表型与脑组织转录组数据关联，可揭示分子层面的病理机制。利用 AHBA 人脑转录图谱数据库，Zhao 等(2025)发现精神分裂症患者的功能连接异常模式与突触相关基因(如 SNAP25、SYT1)的表达空间分布显著相关($r=0.53$, FDR 校正 $p<0.05$)，理解精神分裂症的突触病理学提供了在体证据[15]。

多模态纵向预测：预测精神病临床高危人群的转化风险是早期干预的关键。NAPLS-3 联盟(2025)结合结构 MRI、弥散 MRI 和临床量表，开发了基于多核学习的高危人群转化预测模型，在独立验证集中 2 年内转化精神病的 AUC 达到 0.84，敏感度 0.81、特异度 0.76，其中前扣带回皮层厚度、纹状体各向异性分数和基线阳性症状评分是最重要的预测因子。

5. 大规模多模态数据库与临床转化

近年来，多模态神经影像数据库的快速发展为模型训练和验证提供了数据基础：

阿尔茨海默病神经影像倡议(ADNI)：包含 MRI、PET、体液标志物、临床认知评估的纵向数据，是多模态融合研究的金标准数据集。

UK Biobank：计划采集 10 万志愿者的多模态影像数据(MRI、fMRI、dMRI)，是目前规模最大的公开

数据库。

青少年脑认知发展研究(ABCD): 追踪 11,000 余名儿童的脑发育轨迹, 包含结构 MRI、fMRI、dMRI 及多维度行为数据。

人类连接组计划(HCP): 提供 1200 余名健康成人的高分辨率结构、功能、弥散成像数据。

ABIDE: 自闭症领域最大规模多中心数据库, 包含 sMRI 和 fMRI 数据。

这些数据库的开放共享促进了算法的快速迭代和比较验证, 推动了多模态融合研究从单中心小样本向多中心大样本的转变。

6. 挑战与未来方向

尽管多模态影像融合研究取得了显著进展, 但要实现真正的临床转化仍面临多重挑战。

6.1. 数据层面挑战

模态缺失: 真实临床场景中, 受患者配合度、设备可用性、禁忌症等因素影响, 部分模态可能缺失。如何在不完整数据条件下保持模型性能, 是实用化的重要课题。

多中心异质性: 不同机构间扫描设备、采集协议、人群构成存在差异, 导致数据分布偏移。模型在单一数据集表现优异, 跨中心泛化时性能显著下降。

数据规模与不平衡: 疾病早期和罕见病例样本量有限, 训练数据存在类别不平衡。迁移学习、自监督学习、数据增强等技术可部分缓解。

6.2. 模型层面挑战

可解释性不足: 复杂深度模型的决策过程难以直观理解, 影响临床信任度和接受度。XAI 技术需进一步发展和标准化。

模态配准误差: 不同模态影像间的精确配准是实现有效融合的前提。配准误差会引入噪声, 降低融合性能。

融合策略选择: 尚无统一标准指导如何针对特定任务选择最优融合策略。功能导向分类框架有助于系统比较, 但仍需更多实证研究。

6.3. 临床转化挑战

验证体系缺失: 多数模型停留于回顾性研究阶段, 缺乏前瞻性、多中心、真实世界的临床验证。

监管与伦理: AI 辅助诊断软件的审批监管框架尚未完全成熟, 数据隐私、算法偏见、责任归属等伦理问题需明确。

工作流整合: 如何将 AI 模型无缝嵌入现有临床工作流程, 而不增加医生额外负担, 是落地应用的关键。在实际部署中, 多模态 AI 模型面临多重工程化障碍: ① 接口兼容性问题: 现有医院的 PACS (医学影像存档与通讯系统) 和 HIS (医院信息系统) 多采用 HL7/FHIR 等医疗数据交换标准, 而 AI 模型通常需要特定格式的输入数据(如 NIFTI、DICOM 元数据解析), 需开发中间件实现数据格式转换、元数据映射和质量控制; ② 实时算力需求与基础设施不匹配: 复杂的多模态深度学习模型(如基于 Transformer 的融合网络)推理需要 GPU 算力支持, 而多数医院的 PACS 系统仅配备 CPU 服务器, 导致推理延迟过高(可达分钟级), 无法满足急诊卒中、癫痫持续状态等场景的实时性要求; ③ 数据治理与隐私合规: 多模态融合需要汇集影像、电生理、临床文本等多源数据, 涉及跨科室、跨系统调用, 需建立统一的数据治理框架, 确保符合《个人信息保护法》和 HIPAA 等法规要求; ④ 临床决策支持系统集成: AI 输出结果(如热力图、概率分数)需要以直观、可交互的方式嵌入放射科医师和神经科医师的阅片界面, 需与现有报告系统(如结

构化报告模板)深度整合,而非独立运行。

6.4. 未来方向

鲁棒性融合: 发展对模态缺失、噪声、分布偏移鲁棒的融合算法,如基于变分自编码器的缺失模态生成、域泛化和域适应技术。

先验知识嵌入: 将医学先验知识(如脑图谱、纤维束走行、疾病传播模式)融入模型设计,提升可解释性和泛化能力。

大语言模型整合: 利用大语言模型理解非结构化的临床文本信息(病历、影像报告、出院小结),实现影像-文本多模态融合。

基础模型与迁移学习: 基于大规模健康人群数据预训练的脑影像基础模型,通过微调适应下游疾病诊断任务,缓解小样本问题。

多时间点纵向融合: 将疾病视为时间演进过程,利用循环神经网络、Transformer 等建模纵向影像序列,实现动态预测。

联邦学习与隐私保护: 在多中心协作中应用联邦学习,在不共享原始数据的前提下联合训练模型,兼顾数据隐私和模型性能。

标准化评估框架: 建立多模态 AI 模型的标准化评估流程,包括数据划分、性能指标、可解释性度量、外部验证要求等,推动领域规范发展。

7. 总结

多模态影像融合技术正推动神经系统疾病诊断从“单一模态描述”向“系统生物学表征”的深刻转型。MRI、PET、fMRI、dMRI、EEG 等不同模态提供了解剖、功能、分子、电生理等多维度的互补信息,融合这些信息可更全面地刻画疾病全貌,提升早期诊断准确性和预后预测能力。

近年来,深度学习驱动的融合范式不断创新:从特征拼接走向基于注意力机制、正交补偿、跨层引导的高级交互模式;可解释人工智能技术的引入为模型决策提供了可视化依据;大规模多模态数据库的开放共享为算法验证和比较提供了坚实基础。阿尔茨海默病、帕金森病、自闭症谱系障碍、脑卒中、癫痫等疾病的融合研究均取得了显著进展,部分成果已接近临床应用门槛。

然而,从研究走向临床仍面临数据异质性、模态缺失、模型泛化、可解释性、监管伦理等多重挑战。未来的发展方向将聚焦于鲁棒融合算法、先验知识嵌入、大语言模型整合、联邦学习隐私保护等前沿领域,并建立标准化的评估验证体系。

可以预见,随着技术的持续进步和跨学科协作的深化,多模态影像融合有望成为神经系统疾病精准医疗的核心工具,为早期干预、个体化治疗和预后管理提供科学依据和决策支持。

参考文献

- [1] Zhu, Z., Zhang, X., Xu, C. and Shen, Y. (2026) Construction and Interpretability of a Multimodal Deep Learning Model of Electronystagmography-Optical Coherence Tomography Angiography for Early Screening of Alzheimer's Disease. *American Journal of Alzheimer's Disease & Other Dementias*, **41**. <https://doi.org/10.1177/15333175261422037>
- [2] Zhang, R., Sheng, J., Zhang, Q., Wang, J. and Wang, B. (2025) A Review of Multimodal Fusion-Based Deep Learning for Alzheimer's Disease. *Neuroscience*, **576**, 80-95. <https://doi.org/10.1016/j.neuroscience.2025.04.035>
- [3] Guan, Y., Wang, W., Chen, J., Yang, P., Xu, J. and Qi, J. (2026) A Survey of Multimodal Fusion for Alzheimer's Disease Prediction: A New Taxonomy and Trends. *Information Fusion*, **131**, 104098. <https://doi.org/10.1016/j.inffus.2025.104098>
- [4] Hemmerling, D., Dudek, M., Krzywdziak, J., Żbik, M., Szecowka, W., Daniol, M., et al. (2026) Gru-Based Deep Multimodal Fusion of Speech and Head-IMU Signals in Mixed Reality for Parkinson's Disease Detection. *Sensors*, **26**, Article 269. <https://doi.org/10.3390/s26010269>

-
- [5] Huang, W. and Shu, N. (2025) AI-Powered Integration of Multimodal Imaging in Precision Medicine for Neuropsychiatric Disorders. *Cell Reports Medicine*, **6**, Article ID: 102132. <https://doi.org/10.1016/j.xcrm.2025.102132>
- [6] El-Askary, N.S., Gawish, M., Morsey, M.M., Mahmoud, A.M., Aref, M. and El-Arif, T.I. (2025) Towards Explainable Multi-Modal Fusion Strategies for ASD Detection: A Review. 2025 *Twelfth International Conference on Intelligent Computing and Information Systems (ICICIS)*, Cairo, 25-27 November 2025, 560-567. <https://doi.org/10.1109/icicis66182.2025.11313196>
- [7] Viswan, V., Shaffi, N., Malathy, E., Chemmalar Selvi, G., Kavitha, B.R., Abdesselam, A., *et al.* (2026) Multimodal Fusion and Explainability of Artificial Intelligence Models in Alzheimer's Disease Detection. *Brain Informatics*, **13**, Article No. 5. <https://doi.org/10.1186/s40708-025-00291-w>
- [8] Tuxunjiang, P., Huang, C., Zhou, Z., Zhao, W., Han, B., Tan, W., *et al.* (2025) Prediction of NIHSS Scores and Acute Ischemic Stroke Severity Using a Cross-Attention Vision Transformer Model with Multimodal MRI. *Academic Radiology*, **32**, 5453-5467. <https://doi.org/10.1016/j.acra.2025.05.031>
- [9] Amador, K., Winder, A.J., Fiehler, J., Barber, P.A., Wilms, M. and Forkert, N.D. (2025) A Multimodal Multitask Deep Learning Model for Predicting Stroke Lesion and Functional Outcomes Using 4D CTP Imaging and Clinical Metadata. *Scientific Reports*, **15**, Article No. 38136. <https://doi.org/10.1038/s41598-025-21945-z>
- [10] Xiao, L., Zheng, Q., Li, S., Wei, Y., Si, W. and Pan, Y. (2025) Integration of Spatiotemporal Dynamics and Structural Connectivity for Automated Epileptogenic Zone Localization in Temporal Lobe Epilepsy. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **33**, 3065-3075. <https://doi.org/10.1109/tnsre.2025.3595906>
- [11] Zhang, L., Sheng, S., Wang, X., Gao, J., Sun, Y., Xiao, K., *et al.* (2025) CrossConvPyramid: Deep Multimodal Fusion for Epileptic Magnetoencephalography Spike Detection. *IEEE Journal of Biomedical and Health Informatics*, **29**, 3194-3205. <https://doi.org/10.1109/jbhi.2025.3538582>
- [12] Xie, K., Sahlas, E., Ngo, A., Chen, J., Arafat, T., Royer, J., *et al.* (2025) Personalized Biomarkers of Multiscale Functional Alterations in Temporal Lobe Epilepsy. *Nature Communications*, **16**, 10145. <https://doi.org/10.1038/s41467-025-65042-1>
- [13] Rokham, H., Falakshahi, H., Pearlson, G.D. and Calhoun, V.D. (2025) Neuroimaging Data Informed Mood and Psychosis Diagnosis Using an Ensemble Deep Multimodal Framework. *Human Brain Mapping*, **46**, e70347. <https://doi.org/10.1002/hbm.70347>
- [14] Yoon, V., Kim, S., Park, J., Lee, H., Choi, M., Kang, D., *et al.* (2025) Multimodal Meta-Analytically Anchored Clustering Reveals Five Distinct Schizophrenia Spectrum Disorder Subgroups with Divergent Brain-Symptom-Genetics Profiles. medRxiv.
- [15] Zhao, J., Wang, Y., Liu, M., Guo, B., Wang, Y., Gao, B., *et al.* (2025) Deciphering the Molecular Tapestry of Schizophrenia: Integrating Transcriptomics, Neuroimaging, and Clinical Data for Precision Medicine. *Translational Psychiatry*, **15**, Article No. 489. <https://doi.org/10.1038/s41398-025-03692-x>