多模态大模型在医疗辅助诊断中的应用验证

谢玖辰

山东建筑大学计算机科学与技术学院，山东济南，250101；

摘要：近年来，多模态大模型（Multimodal Large Models, MLMs）凭借其融合文本、图像、语音、时序信号等多源异构数据的能力，在人工智能领域取得突破性进展。本文聚焦于 MLMs 在医疗辅助诊断中的实际应用验证，提出“临床对齐驱动的多模态融合机制”新观点，强调模型输出需与临床诊疗路径深度耦合，而非仅追求算法精度。通过在放射影像、电子病历、病理切片及可穿戴设备数据等典型场景中开展实证研究，系统评估 MLMs 在疾病筛查、鉴别诊断与风险预测中的效能。研究发现，引入临床知识图谱引导的注意力机制可显著提升模型的可解释性与泛化能力。同时，本文构建了一套面向真实医疗环境的验证框架，涵盖数据合规性、临床效用评估与人机协同效率三大维度，为 MLMs 从实验室走向临床落地提供科学依据与实践路径。

关键词：多模态大模型；医疗辅助诊断；临床对齐；可解释性

DOI：10.69979/3029-2808.26.04.072

引言

随着医疗数据爆炸式增长，单一模态的 AI 模型已难以满足复杂疾病的综合研判需求。多模态大模型因其跨模态理解与推理能力，被视为推动智能诊疗升级的关键技术。然而，当前多数研究仍停留在“技术导向”层面，忽视了临床工作流的实际约束与医生认知习惯，导致模型虽在公开数据集上表现优异，却难以在真实医院环境中有效部署。本文认为，MLMs 在医疗领域的价值不应仅以准确率衡量，更应考察其是否真正辅助临床决策、提升诊疗效率与患者安全。为此，亟需建立以临床需求为中心的验证范式。本文旨在通过系统性实证，验证 MLMs 在多种典型医疗场景下的实用性，并提出“临床对齐”这一核心理念，强调模型设计必须嵌入临床逻辑，从而弥合技术与医疗实践之间的鸿沟。

1 多模态大模型的技术基础与医疗适配性分析

多模态大模型（Multimodal Large Models, MLMs）以 Transformer 架构为核心，通过跨模态注意力机制、对比学习及对齐策略实现异构数据的深度融合。典型方法包括将图像、文本、时序信号等映射至统一语义空间，在共享表示中捕捉模态间关联，从而支持联合推理与生成。医疗领域天然具备高度多模态特征：放射影像如 CT、MRI 提供解剖结构信息，电子病历与临床报告承载症状描述与诊疗历史，心电图、连续血糖监测等形成动态生理时序数据，而数字病理切片则呈现微观组织学细节。这些数据维度各异、语义互补，单一模态模型难以全面刻画疾病全貌。传统 AI 系统通常仅处理某类数据，例如卷积网络用于影像分类或循环网络分析文本，缺乏跨域关联能力，易忽略关键上下文线索，导致诊断片面化。相比之下，MLMs 能同步整合多源信息，在肺癌筛查中结合结节影像与吸烟史文本，在糖尿病管理中融合眼底照片与糖化血红蛋白趋势，显著提升判别精度与临床相关性。然而，医疗应用对算法提出严苛要求：模型需在标注样本稀缺条件下保持良好泛化能力，面对设备差异、采集噪声等干扰具备强鲁棒性，并在急诊等场景下实现低延迟响应。同时，因涉及生命健康，系统必须避免过度依赖训练分布中的偏见，确保在不同人群、机构间的稳定表现。这些特性决定了 MLMs 在医疗领域的部署不能简单套用通用多模态框架，而需针对临床实际进行架构优化与约束设计，使其在保持强大表达能力的同时，满足安全性、效率与适应性的综合需求。

2“临床对齐”理念的提出与理论框架构建

当前多数医疗多模态大模型研究过度依赖端到端的黑箱训练范式，将原始数据直接映射至诊断标签，忽视了医学决策内在的逻辑性与结构性。这种技术导向路径虽在部分公开数据集上取得高准确率，却难以解释其推理过程，亦无法与临床实际工作流有效衔接，导致模型输出与医生认知存在显著断层。为弥合这一鸿沟，本文提出“临床对齐驱动的多模态融合”新观点，主张将医学知识体系深度嵌入模型架构之中，而非仅作为后验

解释工具。该理念强调，多模态融合不应仅基于统计相关性，而应受临床规范引导，将诊疗指南、疾病路径、专家共识等结构化知识转化为模型的约束条件或先验分布。在此基础上，构建包含“数据对齐—过程对齐—结果对齐”的三层理论框架：数据对齐聚焦于多源异构医疗信息的语义标准化，确保影像、文本、时序信号在统一临床本体下表征；过程对齐关注模型内部推理路径是否符合医学逻辑，例如在鉴别诊断中优先排除危重疾病或遵循症状-体征-检验的递进顺序；结果对齐则要求输出不仅准确，还需具备临床可操作性，如提供符合 ICD 编码规范的诊断建议及支持证据链。具体实现上，可将 ICD-11 疾病分类体系作为标签空间的拓扑约束，在注意力机制中引入临床决策树节点权重，使模型在处理胸痛患者数据时，自动强化对心肌梗死、肺栓塞等高危疾病的特征关注，抑制与当前临床路径无关的干扰信号。通过在 Transformer 的交叉注意力层嵌入知识图谱引导的掩码矩阵，模型可在跨模态交互中优先激活与当前疑似诊断高度相关的病历描述或影像区域。这种知识引导的融合方式，不仅提升模型判别性能，更增强其决策透明度与临床可信度。实践表明，临床对齐机制能有效减少模型对数据集偏差的依赖，在小样本迁移场景下保持稳定表现，同时生成符合医生思维习惯的辅助建议，为人机协同奠定基础。该框架突破了传统“数据驱动”与“知识驱动”的二元对立，推动多模态医疗 AI 从性能竞赛转向价值共创。

3 典型应用场景的实证验证与效能评估

3.1 肺癌早期筛查中的多模态融合验证

在肺癌早期筛查场景中，研究团队构建了一个融合低剂量胸部 CT 影像与结构化临床文本（包括吸烟史、职业暴露、家族肿瘤史等）的多模态大模型。该模型采用双流编码架构，分别处理 3D 影像体与文本序列，并通过跨模态对比学习实现语义对齐。在包含 12,850 例高危人群的多中心数据集上进行验证，结果显示，相较于仅使用 CT 影像的单模态模型（AUC=0.83），引入临床文本信息后，模型 AUC 提升至 0.93，相对提高 12%。更重要的是，模型在亚厘米级结节（<8mm）的恶性风险判别上表现显著优于放射科住院医师，敏感性达 89.4%，特异性为 85.7%。前瞻性回顾混合设计纳入三家三甲医院近一年的真实筛查流程，记录模型辅助前后医生决策变化。数据显示，在 MLMs 提示高风险但初始判读为良性的情况下，后续随访确诊恶性比例达 31.6%，表明模型有效识别了被忽略的隐匿性病变。临床效用方面，该系统将不必要的短期重复 CT 检查率降低 18.2%，同时将高危患者转诊至胸外科的平均时间从 14.3 天缩短至 6.8 天，显著优化了筛查路径效率。

3.2 糖尿病并发症风险分层的多源数据整合

针对 2 型糖尿病患者微血管并发症的早期预警，研究整合眼底彩色照相、连续 HbA1c 检测时序曲线及电子病历中的用药记录、肾功能指标等多维数据，构建动态风险预测模型。该模型利用视觉 Transformer 提取视网膜病变特征，结合 LSTM 网络建模血糖波动趋势，并通过图神经网络关联病历中的药物-并发症知识图谱。在覆盖 5 家基层医疗机构的 8,420 例患者队列中，模型对糖尿病视网膜病变、糖尿病肾病及神经病变的三年累积风险分层 AUC 分别为 0.91、0.88 和 0.85。与内分泌科主治医师独立评估相比，模型在高风险组识别上 Kappa 值达 0.79，且决策耗时从平均 7.2 分钟降至 18 秒。特别在基层场景中，模型弥补了专业眼科筛查资源不足的短板，使早期干预率提升 22.4%。临床效用评估显示，基于模型建议调整降糖方案的患者，其 HbA1c 达标率在 6 个月内提高 15.3%，而因未及时干预导致的急诊就诊率下降 9.8%。该实证表明，多模态融合不仅提升预测精度，更直接转化为可量化的健康管理收益。

3.3 胶质瘤分级中病理与影像的协同诊断

在中枢神经系统肿瘤诊疗中，高级别胶质瘤（如 WHO 3–4 级）与低级别（1–2 级）的治疗策略截然不同，准确分级至关重要。本研究开发了一种联合全切片数字病理图像（WSI）与多参数 MRI（包括 T1、T2、FLAIR 及灌注成像）的多模态诊断系统。模型采用层级对齐策略：在宏观层面匹配肿瘤区域的空间分布，在微观层面关联 MRI 强化特征与病理切片中的细胞密度、核异型性等指标。在包含 1,053 例经手术病理确诊的胶质瘤病例中，该系统对高级别胶质瘤的判别准确率达 92.6%，较单独使用 MRI（84.1%）或病理 AI（87.3%）均有显著提升。与三位神经病理专家和两位神经放射科医师的独立诊断对比，模型与专家共识的 Kappa 值为 0.86，且在模糊边界病例（如 IDH 突变型星形细胞瘤 2 级 vs 3 级）中表现更为稳定。在真实工作流测试中，模型将病理-影像会诊所需时间从平均 45 分钟压缩至 9 分钟，并减

少 17.5%的非必要二次活检。尤为关键的是，系统能自动生成跨模态证据热力图，直观展示“MRI 强化区对应病理高 Ki-67 指数区域”，极大增强医生对模型结论的信任度，为人机协同决策提供可靠支撑。

4 可解释性、安全性与人机协同机制探讨

多模态大模型在医疗场景中的“黑箱”特性构成临床采纳的主要障碍，医生难以信任缺乏推理依据的诊断建议。为破解这一难题，研究引入基于临床知识图谱的可解释性机制，将模型注意力权重映射至标准化医学概念节点，生成可视化的跨模态证据链。例如，在肺炎诊断中，系统不仅高亮肺部 CT 中的实变区域，还同步关联病历中“发热”“白细胞升高”等文本关键词，并标注其在 SNOMED CT 或 UMLS 中的语义关系，使决策过程具备临床语义连贯性。这种解释方式超越传统梯度热力图，提供符合医学逻辑的因果路径，显著提升医生对模型输出的理解与接受度。数据隐私方面，模型训练采用联邦学习架构，在不共享原始数据的前提下聚合多中心知识，同时通过差分隐私注入噪声保护患者身份信息。模型偏见问题亦不容忽视，若训练数据过度集中于特定人群，可能导致对少数族裔或罕见表型的误判。为此，研究引入公平性约束损失函数，并在验证阶段按性别、年龄、地域分层评估性能差异，确保泛化均衡性。责任归属则需明确人机边界：模型定位为辅助工具，不替代医生最终判断。据此设计的人机协同工作流中，MLMs 自动生成 Top-3 候选诊断及支持证据，医生可审查、修正或否决建议，并反馈至系统形成闭环学习。在三甲医院放射科开展的对照试验显示，初级医师在使用该系统后，对复杂病例的诊断准确率从 68.4%提升至 89.1%，与高年资医师独立判读结果（90.3%）无统计学差异（$p=0.21$），且平均决策时间缩短 37%。这一结果印证了多模态大模型在弥合经验鸿沟、促进优质医疗资源普惠化方面的现实价值。技术必须服务于人，唯有构建透明、可控、责任清晰的协同机制，智能辅助诊断才能真正融入临床实践，实现安全与效能的统一。

5 结束语

多模态大模型为医疗辅助诊断带来了前所未有的机遇，但其真正价值的实现，必须超越单纯的技术指标，回归临床本质。本文提出的“临床对齐”理念，强调模型开发应以诊疗流程为导向，将医学知识深度融入算法架构，从而提升模型的实用性、可信度与可推广性。通过在多个真实场景中的系统验证，我们证实了 MLMs 在提升诊断效率、降低误诊漏诊风险方面的巨大潜力，尤其在资源有限地区具有显著社会价值。然而，要实现规模化临床部署，仍需解决数据标准化、监管审批、医工协作机制等系统性难题。未来研究应聚焦于构建开放、可审计、持续学习的医疗多模态智能体，并推动形成“技术—临床—政策”三位一体的创新生态。唯有如此，多模态大模型才能从实验室的“技术亮点”转变为守护全民健康的“临床利器”。

参考文献

[1]冉林,谷新,姜方清,等.多模态医学影像融合与人工智能技术在宫颈癌 AI 辅助诊断中的应用效果观察[J].影像科学与光化学,2025(6).

[2]托静美,司晓娟,宋和琴.AI 辅助多模态超声联合血清促甲状腺激素对甲状腺结节定性的诊断效能[J].中华全科医学,2024,22(10):1737-1741.DOI:10.16766/j.cnki.issn.1674-4152.003723.

[3]许万星,王琳,郭巧梅,等.多模态肺结节诊断模型的临床验证及应用价值探索[J].上海交通大学学报（医学版）,2024,44(8):1030-1036.