

DOI:10.11918/202510003

# 多模态大模型在重大疾病领域的研究综述

陈书晴<sup>1</sup>, 郭熙铜<sup>2</sup>

(1. 哈尔滨工业大学 医学与健康学院, 哈尔滨 150001; 2. 哈尔滨工业大学 经济与管理学院, 哈尔滨 150001)

**摘要:** 癌症、心脑血管疾病、神经退行性疾病等重大疾病的防控是现代医学的核心挑战,其精准诊疗高度依赖对医学影像、电子病历、基因组学等多源异构信息的综合研判。然而,传统单模态分析方法存在信息孤岛,难以全面刻画疾病的复杂生物学机制与临床表型。为应对此挑战,本文系统综述了多模态大模型在重大疾病防控中的研究进展。文中对多模态大模型在重大疾病领域的研究进展进行了全面综述。首先,概括了以 Transformer 为核心的技术范式,阐明其融合多模态医疗数据的底层架构与协同机制。其次,系统介绍了该模型在早期诊断、精准分型、预后预测等核心临床场景中的应用现状,并深入剖析其技术潜力与实证价值;进而,归纳总结了当前面临的数据异构性、模型“黑箱”问题、伦理法规与数据安全等共性挑战。最后,展望了未来发展趋势,重点提出了面向临床任务的专用模型优化、因果推理与可解释性增强、联邦学习与隐私计算以及人机协同智能诊疗等关键突破口。本综述旨在为科研人员、临床医生与政策制定者提供系统性参考,推动多模态大模型在重大疾病防治中的转化落地,赋能精准医疗高质量发展。

**关键词:** 重大疾病;多模态大模型;精准医疗;可解释性;高质量发展

中图分类号: TP391.7

文献标志码: A

文章编号: 0367-6234(2025)12-0156-09

## A review of multimodal large models in the field of major diseases

CHEN Shuqing<sup>1</sup>, GUO Xitong<sup>2</sup>

(1. School of Medicine and Health, Harbin Institute of Technology, Harbin 150001, China;

2. School of Management, Harbin Institute of Technology, Harbin 150001, China)

**Abstract:** The prevention and control of major diseases, such as cancer, cardiovascular and cerebrovascular diseases, and neurodegenerative disorders, remain core challenges in modern medicine. Their precise diagnosis and treatment critically rely on the integrative analysis of heterogeneous multi-source data, including medical imaging, electronic health records, and genomics. Traditional unimodal approaches, however, suffer from information silos and struggle to comprehensively characterize the complex biological mechanisms and clinical phenotypes of diseases. In response to this challenge, this paper systematically reviews the progress of multimodal large models (MLM) in major disease prevention and control. First, we summarize the transformer-centered technical paradigm, elucidating the underlying architecture and synergistic mechanisms that enable fusion of multimodal medical data. Second, we systematically survey applications across core clinical scenarios-early diagnosis, precise subtyping, and prognostic prediction, while deeply analyzing its technical potential and empirical value. Furthermore, we summarize common challenges encountered in practice, including data heterogeneity, the model “black box” problem, and ethical, legal, and data security issues. Finally, we outlook future development trends and propose key breakthrough directions, emphasizing clinically task-oriented model optimization, causal reasoning and enhanced interpretability, federated learning and privacy-preserving computation, and human-AI collaborative intelligent diagnostics. This review aims to provide a systematic reference for researchers, clinicians, and policymakers, promoting the clinical translation of multimodal large models in the prevention and treatment of major diseases, thereby empowering the high-quality development of precision medicine.

**Keywords:** major diseases; multimodal large models; precision medicine; interpretability; high-quality development

在全球人口老龄化与生活方式变革的双重驱动下,癌症、心脑血管疾病及神经退行性疾病等重大疾

病已成为威胁人类健康和社会可持续发展的重大公共卫生问题<sup>[1]</sup>。其高发病率、高死亡率和高社会经

收稿日期: 2025-10-07;录用日期: 2025-11-16;网络首发日期: 2025-11-19

网络首发地址: <https://link.cnki.net/urlid/23.1235.T.20251118.1206.008>

基金项目: 国家自然科学基金(72201076, 72125001, 72431004, 72293584, 72441024); 香江学者计划(XJ2024004); 中国博士后科学基金(2022M710947, 2024T171147); 黑龙江省博士后科学基金(LBH-Z22125)

作者简介: 陈书晴(1990—),女,副研究员,硕士生导师;郭熙铜(1983—),男,教授,博士生导师

通信作者: 郭熙铜, xitongguo@hit.edu.cn

济负担,加之复杂的多因素致病机制与显著的临床异质性,导致传统诊疗体系面临早期识别难、干预窗口短、个体化不足等瓶颈,导致患者 5 年生存率普遍低于 45%,陷入“健康-贫困-疾病恶化”的恶性循环<sup>[2-3]</sup>中。因此,实现重大疾病的早期筛查、精准干预与动态预后评估,是现代医学亟需攻克的核心目标。

然而,传统的医疗数据分析范式正面临根本性瓶颈。其“单点式”的分析思维,过度依赖单一模态数据(如影像或病理)和医生经验,难以整合多源信息,导致在捕捉疾病动态演变、应对个体差异方面能力不足<sup>[4]</sup>。这种信息利用的片面性,不仅限制了诊断的敏感性与特异性,也严重制约了精准医疗策略的临床转化。

人工智能(artificial intelligent, AI)技术的飞速发展,特别是以 Transformer 架构为基础的多模态大模型的崛起,为摆脱上述困境带来了曙光。这标志着医疗 AI 从“单模态感知”向“多模态认知”的范式跃迁。其核心优势在于,能够通过自注意力机制,深度融合医学影像、电子病历、基因组学等多源异构数据,模拟临床多学科团队的综合研判过程<sup>[5-6]</sup>。例如,在肺癌诊疗中,通过整合计算机断层扫描(computed tomography, CT)影像、病理切片与基因数据,多模态模型不仅实现了亚型的精细化分类,更将靶向治疗响应预测的准确度提升至 0.96 以上,显著超越了任何单一模态方法<sup>[7-8]</sup>。更重要的是,多模态大模型的应用潜力远不止于提升诊断精度。其有望突破单一模态的性能瓶颈,挖掘出传统方法难以察觉的跨模态生物标志物,从而揭示疾病背后更深层的生物学机制。这不仅能够推动临床决策从经验驱动向数据驱动转型,更有望重构重大疾病的诊疗流程与科研范式,实现医学实践的系统性变革。

尽管前景广阔,但多模态大模型在重大疾病领域的应用仍处于早期探索阶段,面临数据异构与隐私、模型可解释性(“黑箱”困境)、高昂的标注成本以及亟待完善的伦理法规等多重挑战。鉴于此,本文旨在系统性地梳理多模态大模型在重大疾病领域的最新研究进展,深入剖析其核心技术范式、关键应用场景、面临的挑战以及未来发展方向,以期为该领域的学术研究与临床转化提供参考。

## 1 多模态大模型的核心技术与医疗数据模态

多模态大模型在重大疾病领域的成功应用,源于其卓越的跨模态表征学习与对齐能力。这种能力由其创新的架构基础、对关键医疗数据模态的深刻理解及模态融合策略三者共同决定。本文将系统阐

述该范式的 3 个核心支柱,揭示其支撑高性能医学人工智能的内在机制。

### 1.1 基于 Transformer 的多模态数据表征

多模态大模型的主流架构基于 Transformer<sup>[9]</sup>。其核心创新在于自注意力机制,它克服了传统循环神经网络和卷积神经网络的局限,能够直接捕捉序列内任意元素间的长距离依赖关系。这一特性在医疗场景中至关重要,例如,模型可同时关联患者 10 年前的过敏史与当前用药,或同时捕捉病理图像中的微观细胞异型性与宏观组织浸润模式,实现跨空间尺度的协同分析。

更重要的是,Transformer 为跨模态语义对齐提供了统一框架。通过将不同模态的数据(如图像块、文本词元、基因序列片段)线性投影至共享的嵌入空间,自注意力机制可动态计算并学习模态间元素的关联权重。例如,模型能够自动建立肺部 CT 中“磨玻璃结节”的视觉特征、放射报告中“ground-glass opacity”的文本描述,以及基因突变的分子信号之间的隐式语义关联,从而构建一致、多维的疾病表征。

基于 Transformer 框架,衍生出两种主流的多模态实现范式。

1) 双塔模型。使用独立的编码器处理不同模态,通过对比学习在表征空间进行对齐,其优势在于推理效率高,尤其适用于检索类任务<sup>[10]</sup>。

2) 融合编码器模型。将多模态输入拼接后送入统一 Transformer,通过交叉注意力机制实现细粒度交互,虽计算成本较高,但可支持更复杂的联合推理。

在医疗实践中,研究人员常根据任务需求灵活组合这两种范式。例如,先利用预训练双塔模型快速检索候选病例,再调用融合编码器进行精细化诊断推理,兼顾效率与精度。

### 1.2 医疗多模态数据

重大疾病的复杂性决定了单一数据源难以全面刻画其病理本质。多模态大模型的核心价值,正在于其能够整合来自不同维度、不同尺度的异构医疗数据,构建患者的“多维信息图谱”,见图 1。

1) 电子病历与临床文本。电子健康记录(electronic health record, EHR)包含患者的人口统计学信息、主诉、病史、实验室检查结果、医嘱等结构化与非结构化数据。其核心特性在于信息的高度非结构化、语义稀疏性以及上下文依赖性强;同时,数据缺失、术语不规范以及录入错误等问题普遍存在,构成了主要的处理难点<sup>[11-12]</sup>。临床病历(如出院小结、影像报告),是医生诊断思维和临床决策的直接体现。利用基于 Transformer 的大型语言模型,可从

中提取疾病实体、症状时序、因果关系关键语义,为模型提供深层上下文理解能力<sup>[6]</sup>。

2) 医学影像。包括放射影像(X-ray、CT、磁共振成像(magnetic resonance imaging, MRI)、正电子发射型计算机断层显像(positron emission computed tomography, PET))、病理图像、内窥镜,以及皮肤镜图像等,提供从器官到细胞层面的形态与功能信息。例如, MRI 能够揭示脑肿瘤的位置、大小及其对周围脑组织的侵袭情况,而病理图像则能展示肿瘤细胞的核分裂象、异型性等关键病理特征<sup>[13]</sup>。此类数据具有高维度、高分辨率、模态间物理机制差异大等特点,且病灶尺度变化剧烈、背景噪声复杂,标注成本极高(尤其像素级任务)<sup>[14]</sup>。视觉 Transformer 等架构可有效将其转化为富含语义的向量表征。

3) 基因组学与多组学数据。作为揭示疾病分子机制的“金标准”,包括全基因组测序、转录组、蛋白质组、代谢组等。其典型特征是超高维(数万至百万级特征)、强稀疏性、样本量有限,易导致模型过拟合;同时技术噪声显著,需有效降噪以提取生物学意义信号<sup>[15]</sup>。多模态模型可通过注意力机制或图神经网络,建模基因-基因、基因-疾病间的复杂调控网络,将抽象分子信号转化为可计算表征。

4) 其他新兴模态。随着技术发展,可穿戴设备(心率、血氧、步态等动态生理信号)、医学知识图谱(结构化的疾病-药物-症状关系),以及实时手术视频等新型数据源正被纳入分析体系,为疾病的动态监测、个性化干预和手术智能化提供了前所未有的数据支持<sup>[16-17]</sup>。

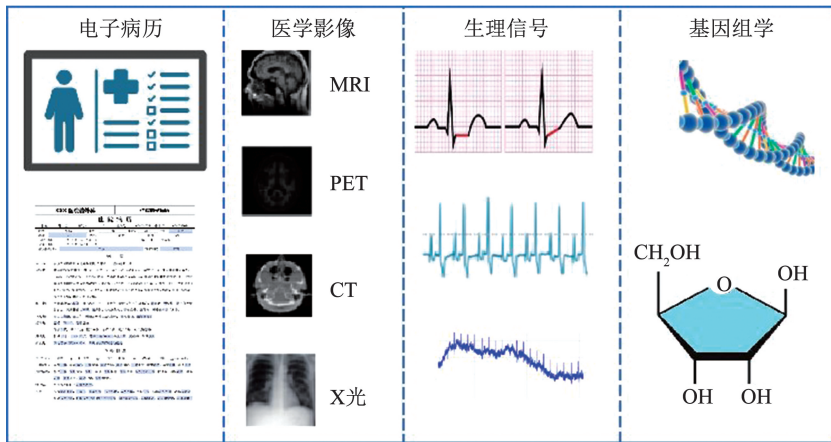


图 1 医学多模态数据

Fig. 1 Medical multimodal data

### 1.3 多模态数据融合策略

多模态融合策略经历了从浅层到深层、从静态到动态的演进。多模态数据融合方法主要可分为 2 大类:1)传统融合方法,包括早期融合、特征级融合与晚期融合(图 2(a));2)更为先进的混合融合方法(图 2(b))。

1) 早期融合。也称为数据层融合。在输入端直接拼接或交织不同模态的原始特征。其优点是保留原始信息完整性,理论上可学习最底层的跨模态关联;但缺点显著:模态间数据分布与尺度差异巨大,易引发表征失衡,且对模型容量与训练数据量要求极高,在医疗数据稀缺场景下难以收敛<sup>[18]</sup>。

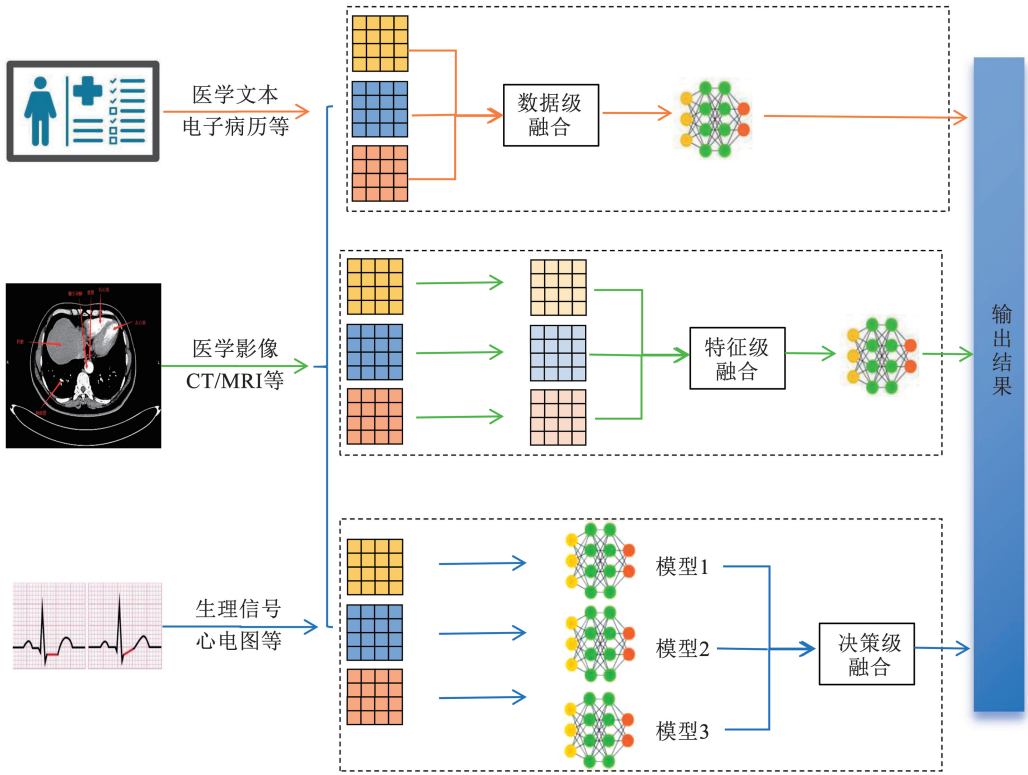
2) 特征级融合。也称为中间层融合。在各模态数据经过独立编码器提取出高层次特征后,在中间层进行对齐、交互与融合,再送入后续网络进行决策<sup>[19-20]</sup>。该策略在性能和复杂度间取得良好平衡,既能实现较深层次的跨模态理解,又对资源需求相对可控,因此在医学多模态分析中被广泛采用。

3) 晚期融合。也称为决策层融合。该方法为每个模态分别训练一个独立的专家模型,通过投票、加权平均或一个小型分类器来整合预测结果。其优势在于灵活性强、易于并行训练,且可复用各模态最优模型。但其本质上是“分而治之”,各模态之间缺乏深层次的信息交互,无法捕捉到模态间细微而复杂的互补与协同关系,融合效果存在上限。

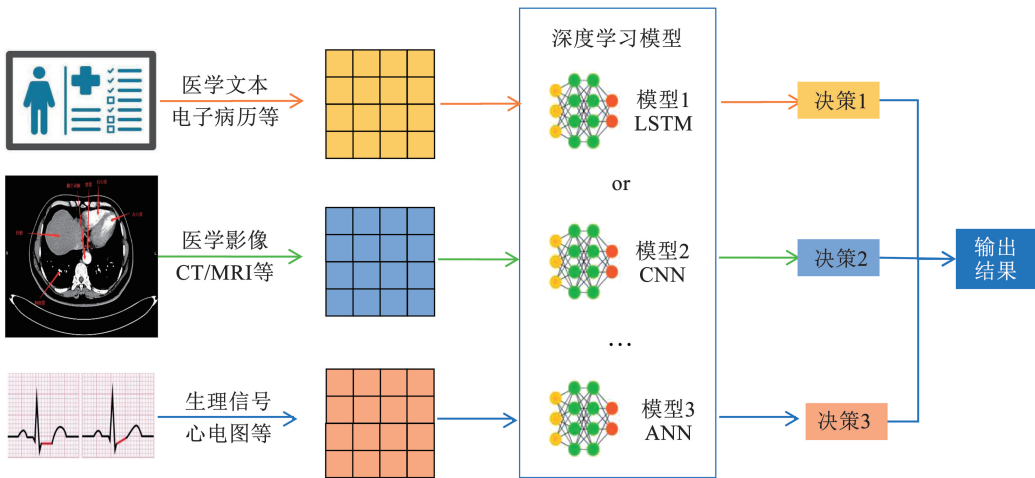
4) 混合/深度融合。结合早期与晚期融合优势,通过精巧设计在网络不同层级实现动态、双向的跨模态交互,其典型技术是跨模态注意力机制<sup>[21]</sup>。例如,在肺癌诊疗中:模型可通过“文本查询图像”机制,使 EHR 中“吸烟史”、“咳血”等文本表征引导 CT 编码器聚焦可疑结节区域;反之,“图像查询文本”机制可让 CT 发现的“肿瘤大小”、“淋巴结转移”等视觉特征强化 EHR 编码器对“晚期”、“预后不良”等概念的感知。这种基于内容的双向推理,使模型超越简单的“看”与“读”,模拟人类专家在多源信息间进行联想、验证与综合判断的过程,从而显著提升诊断的准确性与鲁棒性<sup>[22]</sup>。

综上所述,多模态大模型的技术范式,核心在于以 Transformer 为统一架构,对医学影像、临床文本、基因组学等多源异构数据进行深度整合与动态协

同。这一强大的技术能力,为其在重大疾病的早期筛查、精准分型、治疗评估等核心应用中展现巨大潜力奠定了坚实基础。



(a) 传统多模态数据融合方法示意



(b) 基于深度学习的多模态数据融合方法示意

图 2 多模态数据融合方法

Fig. 2 Multimodal data fusion methods

## 2 多模态大模型在重大疾病领域的核心应用

多模态大模型的技术价值,最终体现在其能否切实解决重大疾病防治中的关键临床难题。其优势在于,通过整合多源异构数据,构建超越人类

专家单一视角的“全景式”疾病理解,从而在关键诊疗环节实现范式革新<sup>[23]</sup>。本文将围绕多模态大模型在重大疾病中的四大核心应用场景,结合具体研究案例,深入剖析多模态大模型的应用价值与科学内涵。基于重大疾病的多模态大模型见图 3。

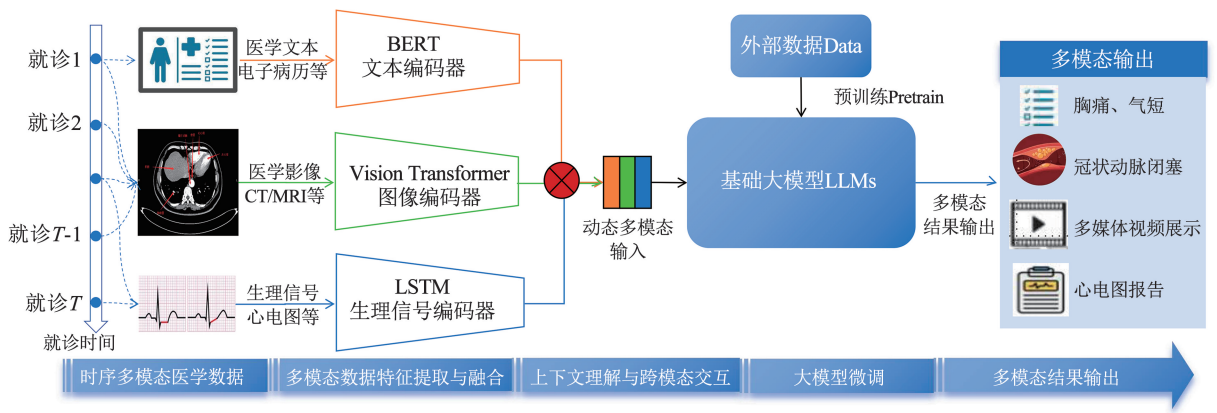


图 3 基于重大疾病的动态时序多模态大模型构建

Fig. 3 Construction of a dynamic temporal multimodal large model based on major diseases

## 2.1 多模态大模型助力重大疾病的早期筛查与诊断

重大疾病的早期诊断是改善患者预后的关键。然而,早期疾病信号往往微弱且散布于不同模态,极易被忽略或误判。多模态大模型通过构建“综合证据链”,可显著提升早期筛查的敏感性和特异性。

1) 癌症的早期预警。传统筛查依赖单一指标,而多模态模型能融合低剂量 CT 影像、临床风险因素与基因组信息等<sup>[24-25]</sup>。例如,一项肺癌研究融合了 CT 视觉特征、吸烟史和循环肿瘤细胞(circulating-tumor DNA, ctDNA)甲基化谱,其识别早期恶性结节的准确率高达 0.96,显著超越单一模态和放射科医生<sup>[8]</sup>。其创新在于,模型能够学习到肉眼难以辨识的“高危纹理”与特定基因改变的关联,实现形态学改变前的预警。

2) 神经退行性疾病的亚临床期识别。对于阿尔茨海默病等神经退行性疾病,临床症状出现时常已错过最佳干预窗口<sup>[26-27]</sup>。多模态模型通过整合结构/功能 MRI、PET 及认知量表,可识别大脑结构、功能与代谢间的细微失衡。相关研究在认知正常人群中,预测出未来 3~5 a 内可能发展为轻度认知障碍或阿尔茨海默症的高风险人群<sup>[28]</sup>。其科学逻辑在于,模型能够捕捉到病理生理过程(如 A $\beta$  沉积)引起的跨模态“微弱信号”,实现超早期识别。

## 2.2 多模态大模型助力重大疾病精准分型与预后预测

1) 传统疾病分型(如肿瘤的 TNM 分期)主要基于解剖学和形态学特征,已无法满足精准医疗时代对个体化治疗决策的需求。多模态大模型通过整合基因组学、病理组学和临床信息,可推动疾病的分型向更深层次、更具预测性的维度演进。

2) 癌症的整合分型。以胶质瘤为例,其诊断已从形态学演进到整合分子标志物的“整合诊断”。

多模态模型将此推向新高度<sup>[29]</sup>。例如,一项研究融合了数字病理、全外显子测序和临床信息,不仅复现了现有分型,更发现了一种具有独特通路激活和免疫表型的“潜在亚型”<sup>[30]</sup>。该亚型患者预后显著不同,为开发新靶向策略提供了线索,实现了从“验证已知”到“探索未知”的跨越。

3) 个体化预后风险评估。传统预后模型难以捕捉风险因素间的复杂交互。多模态模型通过构建动态、多维的风险评估模型解决了此问题<sup>[31]</sup>。例如,在心血管疾病中,一个融合了 CT 血管斑块特征、心电图、血液指标和 EHR 合并症的模型,能精准预测患者出院后的不良心血管事件风险<sup>[7]</sup>。其创新在于能捕捉非线性交互(如特定斑块在糖尿病患者中风险倍增),输出更精细的个性化风险概率。

## 2.3 多模态大模型有助于提升重大疾病的治疗响应评估与新药研发

如何快速、准确地评估患者对治疗的响应,并加速新药研发进程,是重大疾病管理的两大难题。多模态大模型通过建立宏观疗效指标与微观分子机制之间的桥梁,为这两大难题提供了创新解决方案。

1) 精准疗效评估与预测。在肿瘤免疫治疗中,假性/超进展现象给传统疗效标准带来挑战。多模态模型通过纵向整合影像、外周血免疫动态和 ctDNA 变化,能够更早、更准确地判断真实疗效<sup>[32]</sup>。例如,有研究通过分析治疗早期的影像浸润淋巴细胞模式和外周血 T 细胞克隆扩增,成功预测了患者的长期临床获益<sup>[33]</sup>。这种“早期应答预测”能力有助于及时调整无效方案,避免患者承受不必要的负担。

2) 赋能新药研发与精准用药。多模态大模型正在重塑新药研发流程。在靶点发现阶段,可整合文献、基因组、蛋白质网络等数据,识别新靶点<sup>[34]</sup>。在临床试验阶段,可筛选最可能获益的“富集人

群”,提高成功率。在药物重定位方面,可分析药物结构、靶点通路和疾病图谱,预测老药新用<sup>[35]</sup>。这标志着 AI 正从“数据分析工具”向“科学发现伙伴”转变。

### 3 面临的挑战与局限性

尽管多模态大模型在重大疾病研究中展现出前所未有的潜力,其从实验室走向临床实践仍面临来自数据、信任与伦理三大维度的系统性障碍。这些挑战已成为制约其规模化应用的关键瓶颈,亟需系统地剖析与应对。

#### 3.1 医疗数据异质性与标准化瓶颈

多模态大模型的性能高度依赖于海量、高质量、标注一致的训练数据。然而,医疗数据的固有复杂性使其成为模型构建中最根本的制约因素,主要体现在以下 3 个方面。

1) 医疗数据模态的内在异质性。如前所述,医学影像(高维空间)、基因组学(离散序列)与电子病历(稀疏文本)等数据,在格式、尺度和语义上存在巨大差异,导致模型在跨模态对齐时极易产生信息损失或虚假关联<sup>[36]</sup>。

2) 医疗数据孤岛与标准化缺失。数据被严格隔离在不同机构,且内部缺乏统一标准,如不同设备影像参数、不同测序平台流程、不同术语编码习惯。这种“碎片化”现状严重阻碍了大规模、多中心数据集的构建,削弱了模型的泛化能力<sup>[2, 37]</sup>。

3) 医疗高质量数据标注的稀缺与昂贵。精准的监督学习依赖专家标注,但为病理图像进行像素级标注或为复杂病例进行综合诊断,成本极高且耗时。标注的主观性也引入了额外噪声,限制了模型训练规模和可信基准的构建<sup>[30]</sup>。

#### 3.2 临床信任与模型可解释性困境

在高风险的医疗决策场景中,模型的可解释性是保障安全、建立信任、实现人机协同的前提。然而,当前主流多模态大模型(尤其是深度神经网络),因其高度非线性与复杂性,呈现出典型的“黑箱”特性,严重阻碍其临床落地。

1) 医疗决策过程的不可追溯性。当一个多模态模型做出“该患者为肺癌高风险”的判断时,临床医生无法知晓其决策依据。是 CT 影像上的某个微小结节? 是基因测序中的某个突变位点? 还是 EHR 中记录的某种家族史? 模型究竟是如何权衡这些不同来源的证据的? 这种决策过程的“不透明性”,使得医生难以验证、理解和信任模型的输出,更无法在模型出错时进行有效的干预和修正,这与循证医学的基本原则相悖<sup>[38]</sup>。

2) 解释的“伪相关性”与误导风险。当前的可解释性人工智能方法,如注意力机制可视化、特征归因图等,虽能够在一定程度上揭示模型“关注”的区域,但这些解释本身可能存在误导。其揭示的往往是模型学到的统计相关性,而非真正的因果关系。模型可能因训练数据偏差而将“标记笔痕迹”误判为癌症预测因子,这种误导性解释可能带来灾难性后果<sup>[6]</sup>。

3) 缺乏与临床认知的对齐。多数解释是技术性的(像素级、特征级),难以用医生熟悉的病理生理学概念和语言呈现。模型无法解释“为何”某个脑区变化与认知下降相关,导致解释无法真正融入临床思维,难以辅助决策,削弱了其辅助价值<sup>[39]</sup>。

#### 3.3 医疗伦理法规与数据安全

多模态大模型涉及高度敏感的个人健康信息,其发展必须在技术创新与权利保障之间寻求审慎平衡。

1) 患者隐私与数据安全。融合多源数据显著扩大了数据暴露面,使隐私泄露风险呈指数级增长。一旦模型或系统遭到攻击,患者的遗传信息、传染病史等核心隐私可能被泄露,引发严重伦理后果。因此,如何在充分释放医疗数据潜在价值的同时,确保其在采集、存储、传输及模型训练全生命周期中的机密性、完整性和可追溯性,已成为亟待解决的关键技术挑战与法律合规难题<sup>[40]</sup>。

2) 算法偏见与健康公平。模型会继承并放大训练数据中的系统性偏见(如人种、地域、经济水平偏差),导致在少数或弱势群体中性能下降,形成“算法歧视”,加剧医疗不平等。确保模型的公平性和普惠性,是其在公共卫生价值实现的伦理底线<sup>[41]</sup>。

3) 责任归属与监管的滞后。当 AI 参与诊疗并导致不良后果时,责任归属(开发者、医院、医生)尚无明确法律界定。此外,针对此类复杂 AI 系统的审批、监管框架也尚在探索,监管的滞后性为技术的合规化落地带来了巨大不确定性<sup>[42]</sup>。

## 4 面向重大疾病临床落地的多模态大模型前沿方向

文中系统梳理了多模态大模型在重大疾病防控中的潜力与挑战。本文将围绕上述关键挑战,提出四大战略性突破方向,旨在构建一条从“技术可行性”向“临床可用性”转化的演进路径,推动多模态大模型从实验室走向真实世界医疗场景,实现对精准医疗体系可持续发展的实质性赋能。

#### 4.1 面向临床任务的专用模型优化

当前许多研究依赖于通用领域预训练的大模型

进行微调,尽管这些模型具有强大的基础表征能力,但在处理特定临床任务时往往效率不高且性能受限<sup>[43]</sup>。未来的重点应放在构建“预训练-微调-再强化”三位一体的专用化模型优化路径上。

1) 构建医学领域的“基石模型”。利用海量高质量的多模态医疗数据(如数百万级别的影像-报告对、全基因组测序数据、结构化电子病历)进行从零开始的预训练,打造真正属于医学领域的基石模型。这类模型将内化医学先验知识,对疾病相关的特征更加敏感,为下游任务提供更优质的起点。

2) 任务导向的参数高效微调。针对具体的临床任务(如肺癌早期筛查、阿尔茨海默病预后预测),探索更为高效的参数微调技术,如适配器调优、低秩适应和提示调优。这些方法仅需训练少量新增参数,即可快速适应特定任务,降低计算成本和过拟合风险,尤其适合数据稀缺的罕见病研究。

3) 临床反馈驱动的强化学习对齐。构建基于人类反馈的强化学习框架,将资深临床医生的诊断逻辑、治疗偏好和经验判断作为奖励信号,对模型进行迭代校准,保证模型输出既具有数据准确性,又符合临床合理性,进而提升模型的临床实用性与可信度。

#### 4.2 因果推理与可解释性增强

可解释性是 AI 赢得临床信任的关键,而现有关联性分析方法导致了“黑箱”困境<sup>[44]</sup>。未来的研究必须从发现相关性转向探索因果关系,构建兼具预测能力和解释能力的因果性多模态大模型。

1) 融合因果发现与因果推断。探索将因果图模型(如结构方程模型、因果贝叶斯网络)与深度学习架构深度融合的新方法,不仅能从多模态数据中发现潜在因果关系,还能推断给定干预后的可能临床结局,对于理解疾病机制和制定个性化治疗方案至关重要。

2) 发展多模态归因与反事实解释。未来的可解释性技术需超越单一模态的热力图,提供跨模态的联合归因。开发能够清晰指出诊断依据的多模态归因方法,并可生成反事实解释,如“如果基因未突变,则肺癌概率降低 70%”,这种解释方式符合人类因果认知,极具临床说服力。

3) 构建面向临床实践的可解释性交互界面。开发面向临床医生的交互式可视化解释系统,支持以自然语言进行语义化提问,并由模型生成基于多模态证据链的可理解反馈,包括影像关键区域的视觉高亮、显著性基因变异的标注以及相关 EHR 条目的关联呈现。此类基于证据驱动的交互式解释范式,不仅增强了模型决策过程的透明度,更有助于建立医生对模型输出的认知信任与临床可接受性,是

推动人工智能系统融入真实医疗工作流的关键路径。

#### 4.3 联邦学习与隐私计算

数据孤岛现象与隐私安全约束成为制约多模态大模型发展的主要障碍<sup>[45]</sup>。为解决这些问题,未来研究需深入贯彻“数据不动模型动”的协同计算范式,重点发展以联邦学习为代表的隐私增强计算技术,为多模态医学模型的分布式训练与临床转化提供基础设施支撑。

1) 面向医疗异构性的高效联邦学习算法。设计个性化联邦学习和解耦表示学习,允许各参与方在保留本地数据特性的同时,协同学习一个全局共享的、更具泛化性的模型。此外,设计跨模态联邦学习框架,支持不同医院拥有不同数据模态(如 A 医院有影像, B 医院有基因数据)的联合建模。

2) 融合多种隐私计算技术,构建纵深防御体系。构建一个融合联邦学习、差分隐私、同态加密和安全多方计算的纵深防御体系。例如,在模型参数传输过程中应用差分隐私添加噪声,防止模型反演;在聚合服务器端使用同态加密,确保即使在服务器被攻破的情况下,原始参数也不会泄露。这种组合拳式的隐私保护,为敏感医疗数据的共享利用提供安全保障。

3) 建立标准化的联邦学习医疗协作平台与规范。推动建立医疗联邦学习的行业标准,包括数据接口标准、模型评估协议、隐私审计流程等。同时,应鼓励构建开源的、可插拔的联邦学习平台,降低医院和研究机构的参与门槛,形成一个开放、共享、共赢的医疗 AI 研发生态。

#### 4.4 人机协同的智能诊疗新模式

多模态大模型的终极目标是成为医生的“超级智能助手”,重塑未来的临床工作流<sup>[46]</sup>。未来的研究应聚焦于设计以人为中心的人机协同智能系统,实现 AI 与医生的优势互补。

1) 构建“AI-医生”双回路决策支持系统。未来的诊疗系统将是一个闭环反馈的协同体。AI 模型负责处理海量数据,提供快速、全面的量化分析,如病灶分割、生存概率预测、药物敏感性排序;医生则利用其临床经验、人文关怀和伦理判断,对 AI 的建议进行最终审核、决策和与患者的沟通。系统应能记录医生的修正和反馈,并用于模型的持续迭代优化,形成“AI 辅助医生,医生提升 AI”的良性循环。

2) 探索生成式 AI 在临床文书与患者教育中的应用。以大型语言模型为代表的生成式 AI,可以在人机协同中扮演重要角色。例如,自动根据多模态检查结果生成结构化、规范化的出院小结和病程记

录,将医生从烦琐的文书工作中解放出来。此外,可以根据患者的具体情况(年龄、文化程度、病情),生成个性化的、易于理解的疾病科普材料和康复指导,提升患者依从性和就医体验。

3) 前瞻性布局数字孪生与虚拟临床试验。人机协同的高阶形态之一在于构建患者的“数字孪生”(digital twin)。整合个体多模态动态数据,构建患者的数字孪生,在虚拟空间模拟治疗效果、优化个体化策略。进一步,由大规模孪生体构成虚拟临床试验平台,高效评估新药,推动精准医疗从被动响应向主动预测与干预跃迁。这是 AI 深度融合未来医疗的核心范式。

## 5 结 论

重大疾病的防治是现代医学的核心挑战,其高度的异质性与复杂性,正驱动着研究范式从传统的单模态分析向多模态系统性整合跃迁。本文系统考察了多模态大模型在这一变革中的关键作用,并得出以下核心结论。

1) 多模态大模型为理解重大疾病提供了全新的理论框架。基于 Transformer 架构的跨模态表征与对齐能力,将医学影像、电子病历、基因组学等异构数据融合于一个统一的深层语义空间,实现了从“单模态感知”到“多模态认知”的范式转变,使医生得以在更接近疾病本质的维度上进行全景式洞察。

2) 该模型已在关键临床环节展现出颠覆性应用潜力。从提升早期筛查与诊断的精准度,到驱动精准分型与预后预测,再到赋能新药研发,其价值已得到初步验证。其不仅是突破现有临床瓶颈的技术引擎,更有望催生全新的诊疗标准与科研范式,为实现真正的个体化精准医疗铺平道路。

3) 通往临床转化的道路依然充满挑战。当前领域面临数据、算法与社会伦理 3 大核心瓶颈:数据的异构性与标准化难题、算法的“黑箱”与可解释性困境,以及伦理法规与数据安全风险。这些挑战相互交织,其解决迫切需要临床医学、计算机科学、伦理学及政策制定者等多学科的协同攻关。

面向未来,研究焦点必须从“技术可行”迈向“临床可用”。为此,本文提出四大关键突破口。

1) 模型专科化。发展面向特定疾病的预训练架构与微调机制,推动通用大模型向“专科小专家”演进。

2) 决策因果化。将可解释性提升至因果推理层面,构建机制透明的“因果性白箱”模型,以赢得临床信任。

3) 数据协作化。依托联邦学习等技术,构建跨

机构的“数据联邦”,在保障安全的前提下释放多中心数据价值。

4) 诊疗协同化。探索人机协同新模式,构建以患者为中心的动态“数字孪生”系统,支持个性化干预与长期健康管理。

综上所述,多模态大模型正处在一个机遇与挑战并存的关键节点。其不仅是 AI 前沿探索,更是推动精准医疗变革的核心引擎。通过医学、信息科学与伦理学等多学科的持续协作,该技术有望在未来深刻重塑重大疾病的诊疗与管理模式,推动医疗服务向更个体化、预测性与可及性的方向演进,最终实现“为每位患者提供最适化诊疗”的终极目标。

## 参考文献

- [1] KEYL J, KEYL P, MONTAVON G, et al. Decoding pan-cancer treatment outcomes using multimodal real-world data and explainable artificial intelligence[J]. *Nature Cancer*, 2025, 6(2): 307. DOI: 10.1038/s43018-024-00891-1
- [2] 何霆, 吴雅婷, 王华珍, 等. 基于 EHR 的医疗知识图谱研究与应用综述[J]. *哈尔滨工业大学学报*, 2018, 50(11): 137. HE Ting, WU Yating, WANG Huazhen, et al. A survey of medical knowledge graph based on EHR[J]. *Journal of Harbin Institute of Technology*, 2018, 50(11): 137. DOI: 10.11918/j.issn.0367-6234.201806001
- [3] FENDLER A, SHEPHERD S T C, AU L, et al. Omicron neutralising antibodies after third COVID-19 vaccine dose in patients with cancer[J]. *Lancet*, 2022, 399(10328): 905. DOI: 10.1016/S0140-6736(22)00147-7
- [4] YATES J, VAN ALLEN E M. New horizons at the interface of artificial intelligence and translational cancer research[J]. *Cancer Cell*, 2025, 43(4): 708. DOI: 10.1016/j.ccell.2025.03.018
- [5] YIN Shukang, FU Chaoyou, ZHAO Sirui, et al. A survey on multimodal large language models[J]. *National Science Review*, 2024, 11(12): nwae403. DOI: 10.1093/nsr/nwae403
- [6] RAO V M, HLA M, MOOR M, et al. Multimodal generative AI for medical image interpretation[J]. *Nature*, 2025, 639(8056): 888. DOI: 10.1038/s41586-025-08675-y
- [7] XIANG Jinxi, WANG Xiyue, ZHANG Xiaoming, et al. A vision-language foundation model for precision oncology[J]. *Nature*, 2025, 638(8051): 769. DOI: 10.1038/s41586-024-08378-w
- [8] CAMPANELLA G, KUMAR N, NANDA S, et al. Real-world deployment of a fine-tuned pathology foundation model for lung cancer biomarker detection[J]. *Nature Medicine*, 2025, 31(9): 3002. DOI: 10.1038/s41591-025-03780-x
- [9] WU Jiayang, GAN Wensheng, CHEN Zefeng, et al. Multimodal large language models: A survey[C]//2023 IEEE International Conference on Big Data (BigData). Sorrento: IEEE, 2024: 2247. DOI: 10.1109/BigData59044.2023.10386743
- [10] ZHAO Zihao, LIU Yuxiao, WU Han, et al. Clip in medical imaging: A comprehensive survey[J]. *Medical Image Analysis*, 2025, 102: 103551. DOI: 10.1016/j.media.2025.103551
- [11] NEMESURE M D, HEINZ M V, HUANG R, et al. Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence[J]. *Scientific Reports*, 2021, 11(1): 1980. DOI: 10.1038/s41598-021-81368-4
- [12] SUN M, OLIWA T, PEEK M E, et al. Negative patient descriptors: documenting racial bias in the electronic health record: Study examines racial bias in the patient descriptors used in the electronic health record[J]. *Health Affairs*, 2022, 41(2): 203. DOI: 10.1377/hlthaff.2021.01423
- [13] WEN D, KHAN S M, JI XU A, et al. Characteristics of publicly

- available skin cancer image datasets: a systematic review[J]. *The Lancet Digital Health*, 2022, 4(1): e64. DOI: 10.1016/S2589-7500(21)00252-1
- [14] ZHANG Shaoting, METAXAS D. On the challenges and perspectives of foundation models for medical image analysis[J]. *Medical Image Analysis*, 2024, 91: 102996. DOI: 10.1016/j.media.2023.102996
- [15] BEHERA S, CATREUX S, ROSSI M, et al. Comprehensive genome analysis and variant detection at scale using DRAGEN[J]. *Nature Biotechnology*, 2025, 43(7): 1177. DOI: 10.1038/s41587-024-02382-1
- [16] CHEN Shuqing, GUO Xitong, WU Tianshi, et al. Exploring the online doctor-patient interaction on patient satisfaction based on text mining and empirical analysis[J]. *Information Processing & Management*, 2020, 57(5): 102253. DOI: 10.1016/j.ipm.2020.102253
- [17] 金志刚, 何晓勇, 岳顺民, 等. 融合知识图谱的医疗领域命名实体识别[J]. *哈尔滨工业大学学报*, 2023, 55(5): 50  
JIN Zhigang, HE Xiaoyong, YUE Shunmin, et al. Named entity recognition in medical domain combined with knowledge graph[J]. *Journal of Harbin Institute of Technology*, 2023, 55(5): 50. DOI: 10.11918/202201126
- [18] ALBAHRI A S, DUHAIM A M, FADHEL M A, et al. A systematic review of trustworthy and explainable artificial intelligence in healthcare: assessment of quality, bias risk, and data fusion[J]. *Information Fusion*, 2023, 96: 156. DOI: 10.1016/j.inffus.2023.03.008
- [19] NING Zhaolong, HU Hao, YI Ling, et al. A depression detection auxiliary decision system based on multi-modal feature-level fusion of EEG and speech[J]. *IEEE Transactions on Consumer Electronics*, 2024, 70(1): 3392. DOI: 10.1109/TCE.2024.3370310
- [20] YE Zhuang, YU Jianbo. Multi-level features fusion network-based feature learning for machinery fault diagnosis[J]. *Applied Soft Computing*, 2022, 122: 108900. DOI: 10.1016/j.asoc.2022.108900
- [21] ZHAO Fei, ZHANG Chengcui, GENG Baoheng. Deep multimodal data fusion[J]. *ACM Computing Surveys*, 2024, 56(9): 1. DOI: 10.1145/3649447
- [22] STEYAERT S, PIZURICA M, NAGARAJ D, et al. Multimodal data fusion for cancer biomarker discovery with deep learning[J]. *Nature Machine Intelligence*, 2023, 5(4): 351. DOI: 10.1038/s42256-023-00633-5
- [23] FAHRNER L J, CHEN E, TOPOL E, et al. The generative era of medical AI[J]. *Cell*, 2025, 188(14): 3648. DOI: 10.1016/j.cell.2025.05.018
- [24] BOEHM K M, KHOSRAVI P, VANGURI R, et al. Harnessing multimodal data integration to advance precision oncology[J]. *Nature Reviews Cancer*, 2022, 22(2): 114. DOI: 10.1038/s41568-021-00408-3
- [25] CHERUKURI S P, KAUR A, GOYAL B, et al. Artificial intelligence-enhanced liquid biopsy and radiomics in early-stage lung cancer detection; a precision oncology paradigm[J]. *Cancers*, 2025, 17(19): 3165. DOI: 10.3390/cancers17193165
- [26] VAN DER FLIER W M, DE VUGT M E, SMETS E M A, et al. Towards a future where Alzheimer's disease pathology is stopped before the onset of dementia[J]. *Nature Aging*, 2023, 3(5): 494. DOI: 10.1038/s43587-023-00404-2
- [27] HANSSON O. Biomarkers for neurodegenerative diseases[J]. *Nature Medicine*, 2021, 27(6): 954. DOI: 10.1038/s41591-021-01382-x
- [28] XUE Chonghua, KOWSHIK S S, LTEIF D, et al. AI-based differential diagnosis of dementia etiologies on multimodal data[J]. *Nature Medicine*, 2024, 30(10): 2977. DOI: 10.1038/s41591-024-03118-z
- [29] LIU Zaoqu, WU Yushuai, XU Hui, et al. Multimodal fusion of radio-pathology and proteogenomics identify integrated glioma subtypes with prognostic and therapeutic opportunities[J]. *Nature Communications*, 2025, 16(1): 3510. DOI: 10.1038/s41467-025-58675-9
- [30] ZHANG Li, JINDAL B, ALAA A, et al. Generative AI enables medical image segmentation in ultra low-data regimes[J]. *Nature Communications*, 2025, 16: 6486. DOI: 10.1038/s41467-025-61754-6
- [31] LOBATO-DELGADO B, PRIEGO-TORRES B, SANCHEZ-MORILLO D. Combining molecular, imaging, and clinical data analysis for predicting cancer prognosis[J]. *Cancers*, 2022, 14(13): 3215. DOI: 10.3390/cancers14133215
- [32] ASSAF Z J F, ZOU Wei, FINE A D, et al. A longitudinal circulating tumor DNA-based model associated with survival in metastatic non-small-cell lung cancer[J]. *Nature Medicine*, 2023, 29(4): 859. DOI: 10.1038/s41591-023-02226-6
- [33] DE JONG J, CUTCUTACHE I, PAGE M, et al. Towards realizing the vision of precision medicine: AI based prediction of clinical drug response[J]. *Brain*, 2021, 144(6): 1738. DOI: 10.1093/brain/awab108
- [34] LIU Xinyu, ZHANG Jia, WANG Xiaoran, et al. Application of artificial intelligence large language models in drug target discovery[J]. *Frontiers in Pharmacology*, 2025, 16: 1597351. DOI: 10.3389/fphar.2025.1597351
- [35] ZHANG Kang, YANG Xin, WANG Yifei, et al. Artificial intelligence in drug development[J]. *Nature Medicine*, 2025, 31(1): 45. DOI: 10.1038/s41591-024-03434-4
- [36] HU Can, XIA Yingda, ZHENG Zhilin, et al. AI-based large-scale screening of gastric cancer from noncontrast CT imaging[J]. *Nature Medicine*, 2025, 31(9): 3011. DOI: 10.1038/s41591-025-03785-6
- [37] CHEN Shuqing, LAI K H, GUO Xitong, et al. The influence of digital health technology on the allocation of regional medical resources in China[J]. *Health Policy and Technology*, 2025, 14(3): 101013. DOI: 10.1016/j.hlpt.2025.101013
- [38] ACOSTA J N, FALCONE G J, RAJPUKAR P, et al. Multimodal biomedical AI[J]. *Nature Medicine*, 2022, 28(9): 1773. DOI: 10.1038/s41591-022-01981-2
- [39] DECAMP M, TILBURT J C. Why we cannot trust artificial intelligence in medicine[J]. *The Lancet Digital Health*, 2019, 1(8): e390. DOI: 10.1016/S2589-7500(19)30197-9
- [40] PRICE W N, COHEN I G. Privacy in the age of medical big data[J]. *Nature Medicine*, 2019, 25(1): 37. DOI: 10.1038/s41591-018-0272-7
- [41] MCLENNAN S, FISKE A, TIGARD D, et al. Embedded ethics: a proposal for integrating ethics into the development of medical AI[J]. *BMC Medical Ethics*, 2022, 23(1): 6. DOI: 10.1186/s12910-022-00746-3
- [42] MESKÓ B, TOPOL E J. The imperative for regulatory oversight of large language models (or generative AI) in healthcare[J]. *NPJ Digital Medicine*, 2023, 6(1): 120. DOI: 10.1038/s41746-023-00873-0
- [43] WANG Zifeng, WANG Hanyin, DANEK B, et al. A perspective for adapting generalist AI to specialized medical AI applications and their challenges[J]. *NPJ Digital Medicine*, 2025, 8(1): 429. DOI: 10.1038/s41746-025-01789-7
- [44] DURÁN J M, JONGSMA K R. Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI[J]. *Journal of Medical Ethics*, 2021, 47(5): 329. DOI: 10.1136/medethics-2020-106820
- [45] ADNAN M, KALRA S, CRESSWELL J C, et al. Federated learning and differential privacy for medical image analysis[J]. *Scientific Reports*, 2022, 12: 1953. DOI: 10.1038/s41598-022-05539-7
- [46] PARK H J, KIM S H, CHOI J Y, et al. Human-machine cooperation meta-model for clinical diagnosis by adaptation to human expert's diagnostic characteristics[J]. *Scientific Reports*, 2023, 13: 16204. DOI: 10.1038/s41598-023-43291-8