

人工智能赋能多模态影像学在炎症性肠病诊疗中的应用进展

康青青, 杨 佳, 王海鑫

延安大学延安医学院, 陕西 延安

收稿日期: 2026年6月21日; 录用日期: 2026年7月15日; 发布日期: 2026年7月22日

摘 要

炎症性肠病(IBD)的精准医疗高度依赖于影像学与病理学评估, 然而, 传统评估手段存在主观性强、重复性差以及信息割裂等局限。近些年来, 人工智能(AI), 尤其是深度学习技术, 为整合内镜、CT/MRI、肠道超声及组织病理学等多模态影像数据提供了全新范式。本文系统梳理近五年AI赋能多模态影像学在IBD中的应用进展。众多研究成果均表明, 单模态AI模型在内镜Mayo评分、CTE图像识别及组织学炎症分级中已展现出接近专家的诊断性能(准确率80%~97%, AUC 0.74~0.99); 而多模态融合策略通过整合异构数据, 在治疗反应预测(AUC达0.858)与组织学活动度评估(AUC达0.874)方面的准确性, 显著优于单一模态方法。虽然多模态融合前景广阔, 但实际中仍面临数据异质性、标注标准不一、模型泛化性不足及缺乏前瞻性验证等多种问题。未来我们应着力于构建标准化多模态数据库、发展可解释性大模型, 并通过多学科协作来推动临床转化, 以实现IBD从“经验决策”向“数据驱动的精准确管理”的进阶。

关键词

炎症性肠病, 人工智能, 多模态影像学, 深度学习, 溃疡性结肠炎, 克罗恩病, 精准医学

The Application Progress of Artificial Intelligence Empowering Multimodal Imaging in the Diagnosis and Treatment of Inflammatory Bowel Disease

Qingqing Kang, Jia Yang, Haixin Wang

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

Received: June 21, 2026; accepted: July 15, 2026; published: July 22, 2026

文章引用: 康青青, 杨佳, 王海鑫. 人工智能赋能多模态影像学在炎症性肠病诊疗中的应用进展[J]. 临床医学进展, 2026, 16(7): 2019-2029. DOI: 10.12677/acm.2026.1672729

Abstract

The precision medicine of inflammatory bowel disease (IBD) is highly dependent on imaging and pathological assessment. However, traditional assessment methods have limitations such as strong subjectivity, poor repeatability, and information fragmentation. In recent years, artificial intelligence (AI), especially deep learning technology, has provided a brand-new paradigm for integrating multimodal imaging data such as endoscopy, CT/MRI, intestinal ultrasound and histopathology. This article systematically reviews the application progress of AI-enabled multimodal imaging in IBD over the past five years. Numerous research results have shown that single-modal AI models have demonstrated diagnostic performance close to that of experts in endoscopic Mayo scoring, CTE image recognition, and histological inflammation grading (accuracy rate 80% to 97%, AUC 0.74 to 0.99). The multimodal fusion strategy, by integrating heterogeneous data, is significantly more accurate than the single-modal method in terms of treatment response prediction (AUC up to 0.858) and histological activity assessment (AUC up to 0.874). Although multimodal fusion has broad prospects, in practice, it still faces various problems such as data heterogeneity, inconsistent annotation standards, insufficient model generalization, and lack of prospective verification. In the future, we should focus on building standardized multimodal databases, developing large interpretable models, and promoting clinical transformation through multidisciplinary collaboration, so as to achieve the advancement of IBD from “experience-based decision-making” to “data-driven precise management”.

Keywords

Inflammatory Bowel Disease, Artificial Intelligence, Multimodal Imaging, Deep Learning, Ulcerative Colitis, Crohn's Disease, Precision Medicine

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

1.1. 炎症性肠病的临床诊疗负担与现有诊疗困境

炎症性肠病(Inflammatory Bowel Disease, IBD)是一类以迁延、反复的肠道炎症为特征的消化系统疾病,临床上主要分为溃疡性结肠炎(UC)和克罗恩病(CD)两大类型。近年来,IBD 在全球范围内的患病比例呈持续增长趋势,其严重危害了患者的日常生活质量,而且对社会医疗体系带来了巨大的经济负担[1]。IBD 的疾病活动度、并发症以及治疗反应的精准评估,是临床医生对其进行临床管理所高度依赖的,然而传统的评估手段存在诸多短板[1]。例如,作为评估黏膜炎症的“金标准”——内镜检查,其各类评分系统,如 Mayo 评分、SES-CD 在不同医师间的判读一致性欠佳;而组织病理学的评估结果也会因为活检样本代表性不足及判读标准不统一等问题产生误差[2]。除此之外,IBD 存在明显的个体表型差异,早期阶段很难被准确筛查确诊,同时该病症的症状与肠结核、肠道淋巴瘤等多种疾病较为相似,临床鉴别难度较大,这也进一步增加了诊疗工作的难度[3]。因此,临床上亟需建立一种新型的客观、可重复、无创或微创且能整合多维度信息的评估体系,以突破当前诊疗实践中的困境。

1.2. 多模态影像学在炎症性肠病诊疗中的核心作用

多模态影像学已经逐渐成为 IBD 综合评估的重要手段。通过整合内镜、横断面影像(CT/MRI)、肠道

超声及组织病理学图像等多项资料,帮助临床医生从宏观的解剖结构、功能状态乃至微观的分子层面全方位描绘疾病特征[1][2]。其结合了各种辅助检查的优势:内镜可以直观评估黏膜病变,并进行靶向活检;CT 小肠造影(CTE)和磁共振小肠造影(MRE)能够精确评估肠壁的增厚、瘘管及脓肿等穿透性并发症;肠道超声凭借其无创、便捷的优势,在长期动态监测疾病活动方面展现出独特优势;而数字病理技术的发展则使组织学炎症的量化分析成为可能[4][5]。多模态数据的融合,有助于推动诊疗模式从“症状导向”向“机制导向”转变,实现更精准的疾病分型与预后判断[6]。但大量的异构数据的整合,会大幅增加临床医师的判读与分析压力,因此临床工作中亟待智能辅助工具辅助其决策。

1.3. 人工智能与多模态影像学结合的研究背景与近 5 年研究概况

人工智能(Artificial Intelligence, AI),特别是深度学习算法的迅猛发展,为处理多模态影像数据的复杂性提供了强大的技术支撑。近五年来,相关研究呈爆发式增长,广泛覆盖内镜自动评分、CT/MRI 瘘管识别、病理切片炎症分级及多模态数据融合预测等多个方向[5][6]。研究表明,卷积神经网络(CNN)等算法在 IBD 相关任务中表现优异,诊断准确率可达 80%~97% [7]。国内学者亦取得瞩目进展:徐昶等基于深度卷积神经网络构建了 UC Mayo 内镜评分模型,为内镜病变的自动化量化评估提供工具[8];李裕等人进一步开发了融合临床知识的可解释性模型,经外部队列验证后准确率达 0.885,优于高年资医师(0.870) [9]。在病情复发预测方面,AI 辅助评估黏膜血管的愈合状态被证实能够显著区分远期复发风险[7]。这些成果充分表明,AI 不仅可以提升 IBD 的诊断效率与一致性,同时在预判疾病进展规律、评估临床治疗应答效果方面,已展现出可观的临床应用潜力。

2. 人工智能赋能多模态影像学的技术基础与近 5 年算法进展

2.1. 人工智能医学影像分析的核心算法发展概述

AI 在医学影像分析中的核心算法包括传统机器学习和深度学习两大类。在 IBD 多模态分析中,卷积神经网络(CNN)、自注意力机制及多头注意力架构被广泛用于图像特征提取和分类[5][10]。除此之外,自然语言处理技术也逐渐应用于内镜报告的结构化信息提取[5]。现有数据表明,AI 模型在 IBD 相关影像任务中的诊断准确性可达 80%~97%, AUC 范围为.74~.99,判别能力优异[7]。与此同时,可解释性人工智能(XAI)技术的逐步引入,通过特征归因与可视化架构显著增强模型透明度,可为临床医师建立技术信任、制定个性化管理决策提供重要支撑[11]。

2.2. 单模态影像学 AI 模型的技术优化进展

单模态 AI 模型能够实现不同影像源的特异性优化。在内镜领域,李裕团队通过收集的 2251 张内镜图像构建的可解释模型,外部验证准确率达 0.885,且通过 Grad-CAM 热力图技术实现了推理过程的可视化[9];徐昶等利用 2400 张图片构建的 Mayo 评分模型同样表现稳健[8]。在 CTE 影像中,王府进等采用改良 YOLO-V5 模型识别 IBD,回顾性纳入 192 例患病受试者、103 例对照受试者,为断层影像学智能辅助诊断积累了实践依据[12]。在病理领域,Cai 等利用南京鼓楼医院和南方医科大学珠江医院共 756 张有效全切片图像(WSI)构建了自动化分级模型,内部及外部测试集 AUC 分别达 0.863 和 0.947,诊断能力可与病理医师持平[13]。多项研究结果表明,通过任务导向的架构设计与高质量标注训练,单模态 AI 已经在特定场景下趋近资深临床专家。

2.3. 多模态数据整合的 AI 算法框架进展

多模态融合是提升诊断全面性的关键路径。在技术策略层面,当前主流的多模态融合框架可细分为三类:早期融合(特征级拼接)、中期融合(共享表示学习)和晚期融合(决策级投票)。针对 IBD 的异构数据

特点(如内镜图像的空间结构、CTE 的断层纹理、病理 WSI 的细胞级细节、临床时序指标),多数前沿研究采用基于注意力机制的中期融合,即利用跨模态注意力(Cross-modal Attention)或协同注意力(Co-attention)自适应地学习各模态特征间的相关性权重,避免简单拼接导致的信息冗余或冲突。例如,王宇等构建的模型并没有直接将影像特征与临床参数拼接,而是设计了一个多模态对齐模块,先分别提取病理 WSI 的局部形态特征(通过多实例学习)和 CTE 的全局肠道壁特征(通过 3D CNN),再利用多头注意力计算两者与临床指标(C 反应蛋白、白蛋白等)的交互权重,使模型能动态聚焦于最具预测价值的跨模态信号组合。张丽辉等则采用协同注意力机制,将白光内镜的黏膜颜色纹理与超声内镜的肠壁分层信息进行逐层特征交互,实现了对组织学活动度的协同判别。此外,多实例学习(MIL)被广泛用于处理病理 WSI 这类弱标注数据,通过聚合局部块级预测得到全局级结果,而对比学习范式则被用于拉近同一患者不同模态间的特征距离,从而提升融合表征的鲁棒性。国内诸多科研团队已在该研究领域取得实质性创新成果:王宇等构建了融合内镜活检病理 WSI、CTE 深层影像特征及临床指标的多模态深度学习模型,用于预测 CD 患者英夫利西单抗(IFX)原发无应答,内部与外部验证 AUC 分别为 0.852 和 0.858,显著优于单一模态算法模型[14]。张丽辉等开发了白光内镜与超声内镜融合模型,该模型评估对 UC 组织学活动度的 AUC 可达 0.874 [10]。还有相关研究依托多模态超声影像组学特征构建 UC 严重程度分级评估模型,该模型在训练样本集中的诊断准确率达到 85.33% [15]。综合现有各项研究成果能够发现,结合注意力机制与多实例学习算法的新型融合手段,已然成为优化多模态人工智能诊断性能的核心技术路径。

2.4. 近 5 年主流算法性能变化汇总

近五年来,AI 算法的综合识别与诊断性能稳步提升。以往研究大多基于传统机器学习处理结构化数据,如今深度学习已成为医学图像智能分析的主流手段。当前模型诊断准确率稳定在 80%~97%之间[7]。在内镜量化评分中,可解释模型的准确率可达 0.885 [9];在组织学评估中,AI 的 AUC 可达 0.947 [13];在多模态数据的预测研究中,AUC 可达 0.858 [14];在超声评估中,其准确度达 85.33% [15]。从整体研究发展趋势可以看出,随着自注意力机制、对比学习等架构逐步完善成熟,模型在复杂预后预测任务中的优势愈发明显,但泛化能力仍受限于高质量标注数据的可获得性[16]。

3. AI 赋能单模态影像学在炎症性肠病诊疗中的应用进展

3.1. 人工智能内镜影像:疾病诊断与活动度评估

AI 目前在内镜影像分析中的应用最为成熟。徐昶等利用深度 CNN 模型对 UC Mayo 内镜评分进行分类,效果显著[8];李裕等构建的可解释模型不仅准确率高,且通过可视化技术增强了临床可信度[9]。中山六院吴小剑团队研发的 INTELCAPE 系统,可基于胶囊内镜视频实现小肠区域自动分割以及对 CD 的辅助诊断,诊断精度可媲美专科医生,且处理速度提升 10 倍[17]。在疾病复发预测方面,AI 辅助评估的血管愈合状态被证实与远期临床复发显著相关[7]。《肠镜人工智能系统临床应用专家共识(2023,武汉)》(以下简称《肠镜 AI 共识》)同样肯定了 AI 在肠镜质量控制与辅助诊断中的重要作用[18]。综上,AI 内镜应用显著提升了疾病活动度评估的客观性,也为临床试验的患者分层提供了可靠工具[7]。

3.2. 人工智能放射影像(CT/MRI):瘘管/并发症识别与分型

在放射影像领域,AI 已逐步用于 CD 并发症的识别与分型。王府进等利用改良 YOLO-V5 模型分析 CTE 图像,为 IBD 的影像智能化识别提供了高效的技术手段[12]。深度学习算法还被用于客观量化肠壁增厚、脂肪沉积等各类肠道炎症影像学征象,能够有效减少观察者间差异[4]。尽管多数模型现阶段仍处于验证阶段,但其在自动识别瘘管路径、区分活动性狭窄与纤维性狭窄方面已经展现出超越传统人工阅

片的潜力。但数据标准化程度较低与适配性能不足仍是制约其广泛应用的主要阻碍[16]。

3.3. 人工智能肠道超声：无创疾病活动监测

肠道超声作为无创的监测手段，日益受到重视，而 AI 的引入进一步提升了其定量分析能力。国内研究基于多模态超声影像组学构建了 UC 严重程度的智能评估模型，筛选出达峰时间、肠壁厚度等核心影像学指标，其训练集准确度达 85.33% [15]。张丽辉等研发的超声内镜融合模型在组织学活动度的评估中同样表现出良好性能[10]。尽管该方向的相关研究数量仍较为有限，但现有证据表明，AI 驱动的超声分析能够降低检查结果对操作人员经验的依赖，可支持长期随访中的动态监测。

3.4. 人工智能病理影像：炎症分级与病变特征识别

在组织病理层面，AI 系统已取得实质性进展。Cai 等开发的自动化分级方法在内外部队列中均展现出与专业病理医师相当的诊断水准[13]。除此之外，国内研发的 IBDAIM 模型在辅助鉴别正常肠黏膜、CD 和 UC 组织方面表现优异[3]。AI 除了可以自动化完成传统组织学评分，还能深度挖掘镜下细微特征，有效降低人为判断带来的主观误差。但现阶段，病理 AI 的临床落地仍面临着全切片扫描设备普及率低、标注标准不统一等现实障碍[16]。

4. 人工智能整合多模态影像学在炎症性肠病精准诊疗中的应用进展

4.1. 多模态数据整合辅助炎症性肠病的早期诊断与鉴别诊断

多模态融合策略可有效弥补单一模态数据信息片面的固有缺陷。昆明医科大学缪应雷教授团队集成多种影像模态与内镜图谱，创建了适用于 IBD 鉴别诊断的新技术体系[7]。《炎症性肠病实验室诊断专家共识(2026 版)》的发布，进一步规范了 IBD 筛查与鉴别诊断路径[19]。AI 在内镜图像中实现对 UC 与 CD 的实时鉴别，其准确率已经超过 92% [7]。这种多模态协同分析，不仅提升了诊断标准化水平，也为高危易感群体的早期筛查工作开辟了新思路。

4.2. 多模态 AI 模型对疾病活动度与严重程度的精准评估

多模态模型在肠道疾病活动度的量化评估中展现出优异性能。张丽辉团队搭建的白光 - 超声内镜双模态模型对 UC 组织学缓解状态的判断 AUC 可达 0.874 [10]；多模态超声影像组学模型的准确度亦达 85.33% [15]。除此之外，AI 辅助的血管愈合复合评估指标在预测复发风险方面优于单一内镜评分[7]。这些证据充分表明，多模态 AI 能够标准化疾病评估流程，规范疾病全程评估流程，大幅提升多中心临床研究中结局指标判定的统一度与可靠性。

4.3. 多模态 AI 对疾病复发、并发症及治疗应答的预测

在治疗效果预测方面，王宇等整合病理 WSI、CTE 与临床数据，构建了多模态深度学习模型，该模型预测 IFX 原发无应答的 AUC 达 0.858，且可解释性分析显示模型主要关注中性粒细胞浸润区域及隐窝结构异常[14]。该研究结果充分体现了多模态融合在指导个体化用药方面的巨大潜力。尽管多数模型仍处于研发与验证阶段，但已能提前识别高危患者，为临床诊疗方案的制定提供科学参考[7]。

4.4. 多模态 AI 驱动的炎症性肠病精准管理与决策支持

多模态 AI 正逐步演变为 IBD 全流程管理的核心支撑手段。除了能够自动评分与病变识别外，AI 还可以结合患者既往的治疗史与生物标志物，推荐适配个体的最优治疗方案。《医疗健康行业智能体应用技术要求 第 2 部分：炎症性肠病 AI 医生》标准的启动研制，标志着我国在该领域的应用标准化建设取得了突破性进展[20]。未来需进一步整合多学科知识，后续研究可进一步融合多学科理论与技术体系，搭

建以患者个体化需求为核心的一体化智能诊疗闭环体系。

5. AI 多模态影像学模型性能的近 5 年证据汇总分析

5.1. 关键研究信息汇总表

为系统梳理现有证据，本文分别归纳单模态与多模态代表性研究的核心设计、性能及验证方式，见表 1、表 2。

Table 1. Summary of representative studies of monomodal ai models in the diagnosis and treatment of IBD
表 1. 单模态 AI 模型在 IBD 诊疗中的代表性研究汇总

研究 (第一作者/ 团队)	应用 场景	研究设计 (数据 来源)	样本量 (图像/患者)	所用 模态	AI 模型架构	关键性能 指标(内部 验证)	关键性能指标 (外部验证)	验证方式
徐昶等[8] (2024)	UC Mayo 内镜评分 (四分类)	多中心 回顾性	训练集和验证 集共 2400 张(苏 州大学附一院 + HyperKva- sir); 测试集 200 张(江苏大 学附属金坛医 院)	白光 内镜	MobileNetV2、 ResNetV2、 Xception 及 Ef- ficientNetV2S (均基于 ImageNet 预训 练权重进行迁 移学习)	准确率 0.830; 优于低年资医 师(0.785), 略低 于高年资医师 (0.870)	独立外部验证	内部验证 (苏大附一 院 + Hy- perkvasir) + 外部验证 (江苏大学 附属金坛医 院)
李裕等[9] (2025)	UC Mayo 内镜评分 (可解释性)	多中心 回顾性	2251 张内镜 图像(苏州大学 附一院, 训练 + 内部验证)	白光 内镜	基于 EfficientNet 及 5 项临床特征融 合的可解释 CNN (含 Grad- CAM 可视化)	基于 Efficient- Net/Mo- bileNet/Res- Net/Xception 的 可解释模型内 部验证准确率 分别为 0.889、 0.813、0.837、 0.857 (均优于对 应黑盒模型)	EfficientNet 可 解释模型准确 率 0.885, 优于 低年资(0.805)及 高年资医师 (0.870), 其余 3 个模型外部验 证准确率分别 为 0.765、 0.800、0.830	内部验证 (苏州大学 附一院) + 外部验证 (苏州大学 附属常熟 医院)
王府进等 [12] (2024)	CTE 图像 识别 IBD (CD/UC 分类)	单中心 回顾性	192 例 IBD (103 例 CD + 89 例 UC) + 103 例 对照; CTE 图 像 3511 张; 测 试组 8 例(5CD + 3UC, 25 幅)	CTE	改良 YOLO-V5 (5 个子模型: YOLO- V5n/s/m/l/x)	YOLO-V5l 在训 练集 + 验证集 中整体 Pr0.97、 Rc 0.93、 mAP_0.5 0.96	独立测试集: YOLO-V5x 识 别 CD 的 mAP_0.5:0.95 达 0.92, YOLO-V5l 识别 UC 的 mAP_0.5:0.95 达 0.91	内部验证 (9:1 划分) + 独立测试集 (8 例患 者, 25 幅 图像), 未 进行多中心 外部验证
Cai 等[13] (2024)	UC 组织学 炎症分级 (二分类 + 多分类)	多中心回 顾性(南京 鼓楼 + 珠江医院)	570 张 WSI (训 练 + 验证 + 测试); 186 张 WSI 外部测试 (珠江医院 ZJH)	病理 WSI	ResNet50 + 多实例学习 (MIL) + 自注意 力机制(SAMIL)	二分类: AUC 0.863, ACC 0.873, 敏感度 0.913, 特异度 0.816; 多分 类: Macro- AUC0.827, Mi- cro- AUC0.816, ACC 0.769	二分类: AUC 0.947, ACC 0.887, 敏感度 0.889, 特异度 0.858; 多分 类: Macro- AUC 0.908, Micro-AUC 0.898, ACC 0.700	10 折蒙特 卡洛交叉验 证(训练:验 证:测试 = 7:1:2); 独 立外部测试 集(珠江医 院 186 张 WSI); 与 5 年经验病理 医师水平 对比

续表

INTELCAPE 系统(Fan 等) [17] (2026)	胶囊内镜 辅助 CD 诊断(小肠 分割 + 病灶检测+ 视频级诊断)	多中心回 顾性(2 家 中国医 院: 中山 六院 + 瑞金医院)	872 例患者(Co- hort 1: 757 例, Cohort 2: 115 例)	胶囊 内镜 视频	ResNet + Transformer + EfficientNet 多任务分层架 构	小肠分割 IoU 94.82% (Co- hort1)/ 96.87% (Cohort2); 病 灶检测 AUC 0.993/0.980 (准 确率 99.33%, 媲美专科 医生 97.83%)	CD 诊断 AUC: 0.982 (Cohort1)/ 0.984 (Cohort2), 诊断准确率 90%, 媲美专科 医生(93.33%), 速度提升 10 倍; 辅助医生 准确率从 76.7% 提升至 94.8%		
							多中心外部 验证 (Cohort 2)		

Table 2. Summary of representative studies on multimodal AI models in the diagnosis and treatment of IBD
表 2. 多模态 AI 模型在 IBD 诊疗中的代表性研究汇总

研究 (第一作者 /团队)	应用场景	研究设计 (数据来源)	样本量(患 者/病灶)	融合模态	融合 策略	AI 模型架 构	关键性能 指标(内部 验证)	关键性能 指标(外部 验证)	与单模态 对比	验证 方式
王宇等 [14] (2026)	预测 CD 患者 IFX 原发 无应答	多中心 回顾性 (中山 一院、 深圳大学 等)	188 名患 者(内部训 练组 93 名, 内部 验证组 38 名, 外部 测试组 57 名)	病理 WSI + CTE 影 像 + 临 床指标	特征级融 合 + 中 期融合(跨 模态注意 力对齐)	TabNet + 多实例学 习(MIL)融 合框架	AUC 0.852	AUC 0.858	单模态 AUC 仅 0.539, 多模态 显著提升	外部验证 (多中心)
张丽辉等 [10] (2026)	评估 UC 组织学 活动度	单中心 前瞻性	62 例 UC 患者, 195 个病灶(白 光 607 张 + 超声 1055 张)	白光内镜 + 超声内 镜	多模态 特征 融合	Efficient- Net-B2 双 模态模型	组织学缓 解任务: AUC0.874 ± 0.068, 准确度 0.838, 敏 感度 0.867, 特 异度 0.822	白光单模 态 AUC 0.851, 差 异无统计 学 意义 (P = 0.295)	组织学缓 解任务中 优于白光 单模态 (AUC 0.851 vs 0.874, P = 0.295)	5 折交叉 验证重复 3 次 (共 15 次 独立 测试)
李涛鹏 [15] (2026)	UC 严重 程度分级 (改良 Mayo 评分)	单中心 前瞻性	75 例训练 集 + 12 例验证集	多模态 超声(灰阶 + 彩色 多普勒 + 超微血流 成像 + 超声造影)	特征级融 合(多模态 超声影像 组学 + LASSO + ANN)	影像组学 + 人工神 经网络	内部训练 准确率 85.33% (Kappa = 0.753) (概 念上讲非 内部验证 集)	验证集中 达 75.00% (Kappa = 0.791)	未报告单 模态对比	独立 验证集 (12 例)

5.2. 单模态与多模态模型的性能对比及横向批判性分析

目前国内外各项相关研究均证实多模态模型性能优于单模态。王宇等人的研究中，单模态模型的预测能力并不理想，AUC 仅为 0.539，而多模态模型 AUC 可提升至 0.858 [14]；张丽辉等同样验证了这一结论，其构建的双模态模型 AUC 为 0.874，显著优于单模态[10]。多模态整合通过提供互补信息，有效提升了模型的判别能力，同时极大增强了其临床决策支持价值。

在横向比较中, 需注意以下关键差异因素:

- 数据来源与样本异质性: 王宇等采用多中心数据, 外部验证 AUC 保持稳定(0.858), 而多数单中心研究(如徐昶、王府进等)未进行外部验证或外部验证时性能显著下降, 提示单中心模型过拟合风险较高。
- 模态组合与融合深度: 简单早期融合(如超声影像组学拼接)虽操作简便, 但性能提升有限; 而引入注意力机制的深度中期融合(王宇、张丽辉)能更有效地捕获跨模态互补信号, 是性能跃升的关键。
- 标注质量与任务难度: 内镜评分(宏观级)和组织学分级(微观级)的标注一致性差异较大, Cai 等的外部验证 AUC 达 0.947, 部分归因于病理 WSI 有相对标准的量化指标(如 Nancy 指数), 而内镜 Mayo 评分的主观性更强, 导致模型上限受限。
- 验证强度: 仅有少数研究(王宇、Cai、李裕)进行了独立外部验证, 其余多为内部验证或交叉验证, 其报告的性能指标可能存在乐观偏倚。即便在多模态研究中, 融合策略的“增益”并非总能转化为临床净获益。例如, 王宇等的模型虽 AUC 达 0.858, 但其决策阈值、校准度及临床决策曲线分析(DCA)尚未充分报告, 高 AUC 未必等同于临床实用。此外, 多模态模型常因输入模态的缺失(如部分患者无 CTE 或病理)而无法在真实世界中稳定运行, 当前研究普遍采用“完整数据”分析, 忽略了模态缺失的实际情况, 这可能导致选择偏倚。未来需引入模态缺失鲁棒性训练或插补策略, 以提升模型的真实世界适用性。

5.3. 不同研究间模型性能差异的影响因素分析

各类智能模型出现性能参差不齐的核心原因可归结为几方面: 数据质量的高低与样本异质性问题, 回顾性、单中心试验研究的局限、人工标注的主观性、简单拼接与先进对齐机制的算法架构差异, 以及多数模型缺乏真实世界的前瞻性验证[16]。后续研究可通过统一数据采集规范、开展大样本多中心外部验证等方式, 有效缩小不同模型之间的性能差距。

6. 当前研究的局限性与临床转化争议

6.1. 数据层面: 数据异质性与标注质量问题及潜在解决方案

数据异质性与标注质量是限制 AI 模型发展的核心阻碍。各级医疗机构的检查设备参数(如 CT 扫描层厚、内镜光源强度、病理切片染色批次)存在显著差异, 导致图像分布偏移(distribution shift), 使模型难以泛化。人工标注工作容易受到主观判断影响, 尤其在内镜评分和病理分级上, 观察者间一致性(kappa 值)常仅为 0.5~0.7, 标注噪声会直接降低模型的上限性能。

针对这些问题, 学术界已提出多项建设性解决方案:

- 数据标准化流程: 参照《肠镜 AI 共识》[18]和《IBD AI 医生》标准[20], 建立跨中心的影像采集、存储和标注统一规范, 涵盖内镜视频的帧率、CTE 的窗宽窗位、WSI 的扫描分辨率等硬性参数。同时, 开发自动化的质量控制系统, 实时剔除模糊、曝光不足或伪影严重的图像。
- 联邦学习(Federated Learning): 允许多个医疗机构在无需共享原始患者数据的前提下, 共同训练全局模型, 既可扩大数据多样性, 又保护患者隐私。已有初步研究显示, 联邦学习在医学影像分类中可达到与集中式训练相近的性能, 并能显著提高外部泛化能力。
- 半监督与自监督学习: 利用大量无标注影像数据预训练模型(如对比学习、掩码自编码器), 再以少量精细标注数据进行微调, 可缓解标注人力不足的问题, 同时提升模型对异质数据的适应力。
- 主动学习: 设计不确定性采样策略, 让模型主动筛选出最具信息量的样本供专家标注, 从而以最小标注成本换取最大性能增益。

6.2. 模型层面：泛化性不足与算法偏倚及应对策略

现有多数模型在外部测试中都会出现明显的性能下降。王宇等的外部验证结果(AUC 0.858)是目前表现较为优异的少数研究案例[14]。算法偏倚主要源于训练数据对结肠镜的过度依赖，该类数据占比极高，忽视了超声及 MRI 等影像模态，限制了模型对全疾病谱的适用性[16]。此外，模型常因训练集和测试集的人口学特征(如种族、年龄、疾病严重程度分布)差异而产生公平性偏倚，导致在特定亚组中表现不佳。

应对策略包括：

- 对抗性训练(Adversarial Training)：在模型训练中引入领域判别器(domain discriminator)，迫使特征提取器学习与数据来源(中心、设备)无关的疾病相关表征，从而削弱领域偏移的影响。
- 域自适应(Domain Adaptation)：利用无监督域自适应技术(如对抗性域对齐、统计矩匹配)将源域(训练中心)的知识迁移至目标域(新中心)，无需目标域标签即可提升泛化性能。
- 集成学习与不确定性估计：通过多个模型的集成或贝叶斯神经网络，输出预测置信区间，当不确定性过高时触发人工复核，从而在临床决策中提供安全缓冲。

6.3. 临床转化层面：可解释性不足与伦理规范问题

AI 算法的“黑箱”特性阻碍了临床信任。虽然李裕等通过 Grad-CAM 进行了可视化尝试[9]，但目前 XAI 的整体应用仍处初级阶段。在医学伦理维度，患者数据隐私保护、诊疗事故责任界定以及行业监管规范等内容仍缺乏统一界定标准，即便适配该领域的行业规范已进入启动研制阶段[20]。当前 XAI 方法(如注意力热图)往往只能揭示模型“关注何处”，而非“为何如此决策”，这远未达到临床对因果推理的要求。前沿研究正探索概念级解释(concept-based explanation)和反事实解释(counterfactual explanation)，试图以临床语义(如“隐窝结构破坏”、“中性粒细胞浸润密度”)而非像素级热图向医生解释，此类方法值得在 IBD 领域推广。

6.4. 真实世界临床验证证据的缺乏

当前研究中，高质量前瞻性、多中心随机对照试验证据极为匮乏。几乎所有模型均未在真实临床 workflow 中经受大规模、长时间的随访验证[16]。上述高质量临床证据的缺失，已然成为人工智能技术从科研试验落地至常规临床诊疗的核心阻碍。

7. 未来研究方向与技术路径

7.1. 高质量标准化多模态影像数据库的构建

未来研究工作亟需建立涵盖内镜、CT/MRI、超声、病理及多组学数据的标准化数据库。数据库建设需要严格参照统一技术规范，同时严格恪守医学伦理相关准则。《肠镜 AI 共识》[18]及《IBD AI 医生》标准[20]为这一建设提供了框架指引。具体实施中，应推动多中心前瞻性数据采集，引入数据脱敏和隐私计算技术，并建立动态更新的数据治理机制。

7.2. 可解释 AI 与多模态大模型的开发方向

发展 XAI 技术以打破“黑箱”是该领域的必然趋势。李裕等的可视化探索提供了可靠参考范式[9]。后续研发应聚焦打造能同步处理图像、文本、时序数据的多模态大模型，实现针对疾病全程发展过程的动态智能建模分析。特别地，应探索基于 Transformer 或扩散模型(diffusion model)的统一架构，将内镜视频帧、断层切片、病理全扫描及临床时序数据统一编码为令牌序列，利用大规模预训练和提示学习(prompt learning)来适应下游多项 IBD 任务。

7.3. 多学科协作推动 AI 模型的真实世界验证

AI 落地离不开消化内科、医学影像、临床病理、生物信息学以及医学伦理等多领域专家的深度交叉合作。研究人员需设计多中心、前瞻性随机对照试验,量化分析 AI 技术对疾病复发、患者住院等核心临床结局的改善效果。与此同时,还需加强临床从业人员的 AI 应用能力培养[16]。应优先选择 AI 辅助决策与标准诊疗的头对头比较,并以患者层面的远期结局(如 1 年手术率、住院率、生活质量评分)作为主要终点,而非仅以 AUC 等代理指标评价。

7.4. 面向临床落地的整合型诊疗系统构建

成熟的智能诊疗系统应能实时融合多源数据,自动生成活动度评分、并发症风险预警以及个体化干预方案推荐等一系列工作。《IBD AI 医生》标准[20]为一体化诊疗系统的构建提供了技术规范。未来研发需注重模块化与互操作性,最终建立以患者为中心的 AI 驱动精准管理平台。

8. 总结

近五年来, AI 赋能多模态影像学在 IBD 诊疗中取得了令人瞩目的研究成果。AI 通过整合内镜成像、放射影像、超声检查与病理切片等多维度临床资料,有效解决了传统评估方式主观性较强的问题,推动疾病评价体系走向客观化与规范化。国内学者构建的多模态预测模型(AUC 0.858) [14]与融合评估模型(AUC 0.874) [10]充分展示了多模态融合相较单模态的显著优势。《肠镜 AI 共识》[18]与《IBD AI 医生》标准[20]的相继出台,为 AI 技术的规范化应用奠定了基础。即便如此,样本数据异质性、模型泛化性不足及真实场景临床验证欠缺仍是现阶段该领域发展的主要瓶颈[16]。未来研究应聚焦构建高质量数据库、开发可解释大模型与协同多学科验证等工作,推动 AI 从理论层面的可行性突破,真正落地服务于临床,最终有效优化炎症性肠病患者的长期预后,全面提升患者日常生活质量。

参考文献

- [1] 吴朴仙, 王红霞, 林婧, 等. 人工智能在炎症性肠病中的应用进展[J]. 胃肠病学和肝病学杂志, 2024, 33(10): 1377-1381.
- [2] 申越, 周子孺, 孙静, 等. 人工智能在炎症性肠病组织学评估中的应用进展[J]. 中华炎症肠病杂志(中英文), 2025, 9(6): 475-478.
- [3] 裴可, 丛春莉, 李星娇, 等. 人工智能与数字病理评估溃疡性结肠炎组织学愈合的研究进展[J]. 现代消化及介入诊疗, 2025, 30(7): 812-816.
- [4] 李广, 阿里旦·艾尔肯. 溃疡性结肠炎黏膜愈合无创监测的研究进展[J]. 现代消化及介入诊疗, 2025, 30(9): 903-909.
- [5] 熊梓力, 田力. 人工智能在炎症性肠病横断面成像技术中的应用进展[J]. 广州医科大学学报, 2026, 54(1): 43-52.
- [6] Labarile, N., Vitello, A., Sinagra, E., Nardone, O.M., Calabrese, G., Bonomo, F., *et al.* (2025) Artificial Intelligence in Advancing Inflammatory Bowel Disease Management: Setting New Standards. *Cancers*, **17**, Article No. 2337. <https://doi.org/10.3390/cancers17142337>
- [7] 江学良. 人工智能在炎症性肠病诊疗中的研究进展与应用展望[J]. 中华消化病与影像杂志(电子版), 2025, 15(6): 680.
- [8] 徐昶, 林嘉希, 王玉, 等. 基于深度学习的溃疡性结肠炎 Mayo 内镜评分模型的建立[J]. 中华炎症肠病杂志(中英文), 2024, 8(1): 71-76.
- [9] 李裕, 徐晓丹, 朱锦舟. 可解释性模型在溃疡性结肠炎 Mayo 内镜评分中的应用[J]. 中国医疗设备, 2025, 40(9): 20-25.
- [10] 张丽辉, 毛仁, 张宁, 等. 超声内镜融合白光内镜的深度学习模型评估溃疡性结肠炎组织学活动度: 一项单中心前瞻性研究[J]. 中华炎症肠病杂志(中英文), 2026, 10(3): 276-283.
- [11] Okpete, U.E. and Byeon, H. (2025) Explainable Artificial Intelligence for Personalized Management of Inflammatory

- Bowel Disease: A Minireview of Recent Advances. *World Journal of Gastroenterology*, **31**, Article 111033. <https://doi.org/10.3748/wjg.v31.i35.111033>
- [12] 王府进, 孟名柱, 王欣, 等. 改良 YOLO-V5 模型用于识别 CT 肠道造影所示炎症肠病[J]. 中国医学影像技术, 2024, 40(10): 1593-1598.
- [13] Cai, C.F., Shi, Q.Y., Li, J., *et al.* (2024) Pathologist-Level Diagnosis of Ulcerative Colitis Inflammatory Activity Level Using an Automated Histological Grading Method. *International Journal of Medical Informatics*, **192**, Article 105648. <https://doi.org/10.1016/j.ijmedinf.2024.105648>
- [14] Wang, Y., Wang, H.P., Wu, X.M., *et al.* (2026) Development and Validation of a Novel Multimodal Deep Learning Model Integrating Histopathology, Radiology, and Clinical Data to Predict Primary Non-Response to Infliximab in Patients with Crohn's Disease. *Journal of Crohn's and Colitis*, **20**, jjaf206. <https://doi.org/10.1093/ecco-jcc/jjaf206>
- [15] 李寿鹏, 王耀辉, 田霁松, 等. 多模态超声影像组学评分系统评估活动期溃疡性结肠炎严重程度[J]. 中国医学影像学杂志, 2026(1): 83-89+97.
- [16] Liu, X.X., Reigle, J., Prasath, V.B.S., *et al.* (2024) Artificial Intelligence Image-Based Prediction Models in IBD Exhibit High Risk of Bias: A Systematic Review. *Computers in Biology and Medicine*, **171**, Article 108093. <https://doi.org/10.1016/j.combiomed.2024.108093>
- [17] Fan, D.J., Mao, Y.Z., Liang, F., *et al.* (2026) INTELCAPE: A Deep Learning-Powered System for Automated, High-Accuracy Crohn's Disease Diagnosis via Capsule Endoscopy. *Clinical Gastroenterology and Hepatology*. <https://doi.org/10.1016/j.cgh.2026.02.026>
- [18] 中华医学会消化内镜学分会大数据协作组. 肠镜人工智能系统临床应用专家共识(2023, 武汉) [J]. 中华消化内镜杂志, 2024, 41(4): 253-262.
- [19] 上海中西医结合学会检验医学专业委员会, 中国中西医结合学会检验医学专业委员会. 炎症性肠病实验室诊断专家共识(2026 版) [J]. 中华预防医学杂志, 2026, 60(7): 1-17.
- [20] 互联网医疗健康产业联盟. 医疗健康行业智能体应用技术要求 第 2 部分: 炎症性肠病 AI 医生[S]. 北京: 中国信息通信研究院, 2026.