

胃肠道肿瘤人工智能诊疗系统：公共数据库建模与真实世界临床验证的整合

韩怡婷^{1*#}, 田莲莲², 郝建清¹, 黄 婧¹

¹陇东学院医学院, 甘肃 庆阳

²第四军医大学, 西京消化病医院, 国家消化系统疾病临床医学研究中心和消化系肿瘤整合防治全国重点实验室, 陕西 西安

收稿日期: 2026年7月18日; 录用日期: 2026年8月11日; 发布日期: 2026年8月20日

摘 要

胃肠道肿瘤具有发病隐匿、病理异质性强、治疗路径长等特点。临床实践中, 单一内镜、影像、病理或组学信息往往难以覆盖筛查、诊断、分期、疗效预测及随访管理的连续需求。公共数据库为模型初步构建提供了规模化材料, 真实世界数据则决定算法能否承受设备差异、病种谱变化与流程约束。本文围绕胃肠道肿瘤智能预测诊断系统的真实世界应用, 综述TCGA、GEO、cBioPortal、TCIA等公共数据库的价值与边界, 梳理内镜、放射影像、数字病理和多组学融合模型在胃癌、结直肠癌、食管癌等场景中的研究进展, 并讨论从公共数据库建模到单中心回顾性验证、多中心外部验证、前瞻性静默试验和临床影响评价的转化路径。现有研究提示, 多模态融合可在一定程度上提高对淋巴结转移、腹膜转移、微卫星不稳定性、病理完全缓解及预后的判别能力, 但模型性能高度依赖数据质量、标注一致性、目标定义、分布偏移控制和可解释输出。未来研究应从“单点算法”转向“可审计的临床系统”, 在数据治理、模型更新、伦理监管和医生-算法协同机制下形成可复现、可验证、可追踪的应用闭环。

关键词

胃肠道肿瘤, 人工智能, 多模态融合, 真实世界研究, 临床验证, 预测模型

AI-Based Diagnostic and Therapeutic Systems for Gastrointestinal Tumors: Integrating Public Database Modeling with Real-World Clinical Validation

Yiting Han^{1*#}, Lianlian Tian², Jianqing Hao¹, Jing Huang¹

*第一作者。

#通讯作者。

文章引用: 韩怡婷, 田莲莲, 郝建清, 黄婧. 胃肠道肿瘤人工智能诊疗系统: 公共数据库建模与真实世界临床验证的整合[J]. 临床个性化医学, 2026, 5(4): 358-367. DOI: 10.12677/jcpm.2026.54260

¹School of Medicine, Longdong University, Qingyang Gansu

²National Clinical Research Center for Digestive System Diseases and National Key Laboratory for Integrated Prevention and Treatment of Digestive System Tumors, Xijing Hospital of Digestive Diseases, Fourth Military Medical University, Xi'an Shaanxi

Received: July 18, 2026; accepted: August 11, 2026; published: August 20, 2026

Abstract

Gastrointestinal tumours are characterised by occult onset, marked biological heterogeneity and a long continuum of clinical management. Single-source information from endoscopy, radiology, pathology or omics is usually insufficient for screening, diagnosis, staging, treatment response prediction and follow-up. Public databases provide scalable materials for model exploration, whereas real-world data determine whether algorithms can tolerate variations in devices, disease spectrum and clinical workflow. This review summarises the value and limitations of public resources such as TCGA, GEO, cBioPortal and TCIA, discusses intelligent models based on endoscopy, radiomics, digital pathology and multi-omics, and proposes a translational pathway from public data mining to retrospective modelling, multicentre external validation, prospective silent testing and impact evaluation. The available evidence suggests that multimodal fusion may improve the identification of lymph node metastasis, peritoneal metastasis, microsatellite instability, pathological complete response and prognosis, but its reliability depends on data governance, annotation consistency, outcome definition, domain shift control and interpretable outputs. Future studies should move from isolated algorithms to auditable clinical systems.

Keywords

Gastrointestinal Tumour, Artificial Intelligence, Multimodal Fusion, Real-World Study, Clinical Validation, Prediction Model

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. 背景与研究意义

胃肠道肿瘤包括食管癌、胃癌、结直肠癌及肝胆系统部分肿瘤，是我国和全球恶性肿瘤负担的重要组成部分。GLOBOCAN 2022 数据显示，全球癌症新发病例接近 2000 万，死亡病例接近 1000 万，其中结直肠癌、胃癌和食管癌均位于主要癌种之列[1]。在我国，胃肠道肿瘤的高发与人口老龄化、筛查覆盖不均、生活方式变化以及早诊早治资源分布不平衡有关。与部分实体瘤相比，胃肠道肿瘤的病灶形态、解剖部位、组织学亚型和分子分型差异较大，单一检查结果往往难以充分解释疾病进展和治疗反应。

临床诊疗正在从经验驱动转向证据驱动与数据驱动相结合。内镜可直接观察黏膜表面细微结构，CT、MRI 和 PET/CT 可反映肿瘤局部浸润及转移状况，病理切片能够提供组织结构和细胞形态学依据，基因组、转录组和蛋白组数据则揭示潜在生物学机制。问题在于，这些信息通常分散在不同系统、不同时间点和不同格式中，诊断结论依赖医生对多源证据的整合能力。智能预测诊断系统的价值，不在于替代医生作出单一判断，而在于以可验证的方式把异质数据转化为风险分层、诊疗提示和随访决策支持。

目前许多胃肠道肿瘤智能模型仍停留在算法性能比较阶段。公共数据库模型在论文中常可取得较高

的 AUC 或准确率,但进入临床场景后,模型会遇到样本来源变化、设备参数不一致、病种谱偏移、结局定义差异、缺失数据以及医生工作流约束等问题。由此可见,真实世界应用并非模型训练后的附属环节,而应被置于研究设计的起点:模型需回答临床上确有意义的问题,训练数据需能映射目标人群,输出形式需与诊疗流程相容,验证过程需覆盖外部人群和前瞻性运行。

本文所称“融合模型”,主要指在数据层、特征层、决策层或知识层整合内镜图像、放射影像、数字病理、组学数据和临床变量的智能模型。与单模态模型相比,融合模型能够更接近临床诊疗的真实信息结构,但同时也引入了数据同步、模态缺失、批次效应、可解释性下降和监管边界不清等新问题。基于此,本文从公共数据库、关键应用场景、融合建模策略和临床验证路径四个方面进行综述,为胃肠道肿瘤智能预测诊断系统的真实世界转化提供参考。

2. 公共数据库与真实世界数据基础

2.1. 公共数据库的价值与边界

公共数据库是胃肠道肿瘤智能模型研究的重要起点。TCGA 对胃腺癌、结直肠癌和食管癌等进行了系统的分子图谱描绘,提出胃癌分子分型和结直肠癌多组学特征,为后续基于组学的预后预测、免疫治疗获益推断和分子标志物筛选奠定了基础[2]-[4]。GEO 提供大量芯片和测序数据,适合进行差异表达分析、外部验证和队列合并[5]。cBioPortal 则通过可视化平台整合基因突变、拷贝数变异、表达谱与临床信息,便于研究者快速探索分子事件与临床结局之间的关系[6]。TCIA 作为医学影像公共仓储,为影像组学和深度学习研究提供了可重复使用的影像资源[7]。

公共数据的优势在于样本可获得性、研究可复现性和跨团队比较便利,但其局限同样明显。首先,公共数据库中的病例多来自特定研究项目,采集流程、入组标准和随访完整性与日常临床存在差异;其次,临床变量常不完整,治疗细节、并发症、用药剂量和影像扫描参数缺失较多;再次,不同数据库之间的测序平台、病理制片流程和影像参数并不一致,直接合并可能引入不可忽视的批次效应。若忽视这些问题,模型可能学习到机构、平台或样本处理方式的差异,而不是疾病本身的规律。

因此,公共数据库更适合作为“假设生成”和“初步模型构建”的资源,而不宜被视为临床应用有效性的直接证据。对临床转化而言,研究者需要明确公共数据与目标应用场景之间的距离,包括目标人群是否一致、结局定义是否一致、检查设备是否可比、治疗策略是否具有时代差异等。只有将公共数据库发现与本地真实世界数据、外部中心数据和前瞻性运行数据连续衔接,模型才可能从统计相关走向临床可用。

2.2. 真实世界数据的获取与治理

真实世界数据来源于电子病历、内镜系统、PACS、病理信息系统、手术麻醉系统、检验系统、随访平台和肿瘤登记数据库。与标准化研究队列相比,真实世界数据更能反映诊疗流程中的复杂性,例如患者合并症、检查间隔不固定、治疗方案调整、影像质量参差不齐以及随访失访。对于智能预测诊断系统而言,这些“噪声”并非全然需要剔除,而是模型能否进入临床的压力测试。

真实世界数据治理的核心包括数据标准化、隐私保护、标注质量控制和可追溯管理。内镜图像需记录检查部位、清洁度、放大模式和染色方式;影像数据需保存扫描层厚、增强期相、重建算法和设备型号;数字病理应明确染色方案、扫描倍率和切片质量;组学数据则要说明样本来源、测序平台、质控阈值和批次信息。临床变量应尽可能采用结构化定义,如 TNM 分期、病理完全缓解、无病生存期、总生存期和治疗相关不良反应。

数据标注往往是影响模型上限的关键环节。胃肠道肿瘤的病灶边界、浸润深度、淋巴结状态和肿瘤

退缩分级均存在观察者差异。较为稳妥的做法是建立双人独立标注、专家仲裁和抽样复核机制，并在数据集中保留标注者信息、标注工具版本和争议样本记录。对真实世界部署而言，模型不应只输出一个概率值，还应输出其适用范围、置信程度、主要证据区域和可能的失败情形。

3. 智能预测诊断系统的关键应用场景

3.1. 内镜辅助识别与早诊早筛

内镜是胃肠道肿瘤早期诊断的重要入口。结直肠癌筛查中，腺瘤检出率与后续癌症风险密切相关，实时计算机辅助检测系统已在随机对照研究中显示出提高息肉和腺瘤检出率的潜力[8][9]。类似思路也被用于胃早癌、食管早癌、Barrett 食管相关病变及炎症性肠病相关肿瘤的辅助识别。与静态图片分类相比，真实内镜场景包含运动模糊、气泡、黏液、反光、视野遮挡和操作习惯差异，系统必须在实时性、低误报和低漏报之间取得平衡。

内镜 AI 系统的临床难点主要有三点。第一，病灶标注粒度不同会影响训练目标：以图片级标签训练的模型可能只学到背景差异，而像素级或框级标注虽然更精确，却需要大量专家工作量。第二，模型在高质量示教图片上表现良好，不代表能适应基层医院、急诊检查或准备不充分患者。第三，系统提示会影响内镜医师的注意力分配，过多误报可能造成疲劳，过强提示也可能诱发自动化依赖。因而，内镜 AI 的临床评价不仅应报告病灶检出性能，还应评价检查时间、撤镜质量、误报负担、医师学习曲线和患者层面的净获益。

3.2. 影像组学用于分期、转移与疗效评估

影像组学的基本思想是从医学影像中提取可量化特征，并与临床结局或分子状态建立联系[10]-[12]。随着特征提取工具和标准化倡议的发展[13][14]，影像组学已用于胃癌、直肠癌和食管癌的淋巴结转移、远处转移、肿瘤退缩和生存预测研究。其中，胃癌淋巴结转移和腹膜转移是较具代表性的场景：传统影像对微小转移和隐匿腹膜转移的敏感性有限，若模型能够在术前提供更精细的风险提示，可辅助决定腹腔镜探查、围手术期治疗和手术范围[15][16]。

直肠癌新辅助放化疗或全程新辅助治疗后的疗效判断，是另一类具有临床价值的影像 AI 任务。多参数 MRI 可反映肿瘤形态、扩散受限和周围组织关系，影像组学模型被用于预测病理完全缓解、肿瘤退缩分级和器官保留可能性[17]-[19]。然而，疗效预测模型若要影响“等待观察”或手术策略，必须具备极高的安全边界。仅凭单中心回顾性 AUC 不足以支撑临床改变，需进一步验证模型在不同 MRI 设备、扫描方案、治疗方案和病理评估体系中的稳定性。

影像组学研究中常见问题包括图像分割不一致、特征冗余、样本量不足、过拟合和外部验证缺失。传统手工特征模型可解释性相对较好，但对分割和预处理敏感；深度学习模型能够自动学习高阶特征，却更容易受到域偏移影响。较合理的方向是将影像模型与临床变量、实验室指标、内镜特征和病理信息联合建模，并在报告中说明数据预处理、特征筛选、交叉验证、校准曲线和决策曲线分析。

3.3. 数字病理与分子标志物推断

数字病理为胃肠道肿瘤智能诊断提供了高信息密度的数据基础。全视野切片包含腺体结构、细胞核形态、间质反应、淋巴细胞浸润和坏死等多层次信息，这些形态学特征与分子状态和预后之间存在潜在联系。研究显示，深度学习可从常规 HE 切片中预测结直肠癌和其他胃肠道肿瘤的微卫星不稳定性状态[20]-[22]，也可用于推断关键分子通路和突变状态[23][24]。这类研究提示，传统病理图像中可能蕴含尚未被人工显微观察充分利用的隐含信息。

数字病理模型的优势在于可与现有病理流程衔接,并可能降低分子检测前的筛选成本。但应注意,形态学推断并不能直接替代标准分子检测。以 MSI 为例,病理 AI 可作为预筛工具,提示需要进一步免疫组化或分子检测的病例;若将其作为最终检测结果,必须证明其在目标人群中的敏感性、特异性和临床安全性。全视野切片体量巨大,弱监督学习和多实例学习为切片级标签训练提供了可行路径[25],但模型关注区域是否具有病理学合理性仍需专家验证。

病理 AI 还可与基因组和临床结局联合用于预后预测。卷积神经网络结合组织图像和基因组数据预测肿瘤结局的研究表明,形态信息与分子信息之间存在互补价值[26]。近年也有研究尝试引入生物学知识约束,使深度学习模型同时服务于预后预测和免疫治疗反应评估[27]。对胃肠道肿瘤而言,这类知识引导模型有望缓解“黑箱”问题,但仍需在更大规模、多中心和前瞻性队列中验证。

3.4. 预后分层和治疗获益预测

胃肠道肿瘤治疗路径复杂,预后受到肿瘤分期、分子亚型、切缘状态、淋巴结清扫质量、治疗强度、营养状态和合并症等多因素影响。传统预后模型多基于 TNM 分期和少数临床变量,能够提供总体风险估计,但对个体化治疗获益的解释有限。融合模型通过联合临床、影像、病理和组学数据,有望将患者分为复发风险不同、治疗敏感性不同和随访需求不同的亚群。

治疗获益预测不同于单纯预后预测。预后模型回答“患者结局可能如何”,而治疗获益模型回答“某项治疗是否改变结局”。若研究只在接受同一治疗的队列中训练模型,很难区分患者本身风险与治疗敏感性。真实世界研究应尽量引入治疗暴露信息、时间依赖变量和混杂控制策略,并采用外部验证或前瞻性研究确认模型对治疗选择的实际增益。对于免疫治疗、靶向治疗和新辅助治疗等场景,模型输出应与可检测、可干预的临床问题对应,而不是停留在抽象风险评分。

4. 融合模型构建策略与关键技术

4.1. 融合层次：从数据拼接到知识约束

多模态融合可按融合位置分为早期融合、中间融合和后期融合。早期融合直接将不同模态的特征或原始数据拼接输入模型,形式简单,但对数据完整性、尺度统一和噪声控制要求较高。中间融合在各模态建立子网络或特征提取器后,通过注意力机制、图神经网络或交叉模态交互层进行信息整合,适合处理图像、文本和组学等结构差异较大的数据。后期融合则在各模态模型分别输出预测结果后再进行加权或集成,便于解释和维护,适合临床部署早期阶段。

对于胃肠道肿瘤而言,融合策略应服从临床问题。例如,术前淋巴结转移风险预测可联合 CT 影像、内镜肿瘤部位、活检病理类型、肿瘤标志物和炎症指标;新辅助治疗反应预测可联合基线 MRI、治疗中影像变化、病理分化程度和血液学指标;MSI 预筛可联合 HE 切片、年龄、肿瘤部位、组织学亚型和免疫组化结果。融合不是模态越多越好,关键在于新增模态是否提供独立信息、是否可在目标场景稳定获得、是否会增加不可接受的成本和等待时间。

近年来,知识引导建模受到关注。与纯数据驱动模型相比,知识约束可将肿瘤生物学通路、TNM 分期规则、治疗指南或病理诊断逻辑嵌入模型结构或损失函数中,使模型输出更容易被临床解释[27]。不过,知识约束并非天然可靠。若知识图谱不完整、指南更新滞后或规则过度简化,模型可能固化已有偏见。因此,知识引导模型仍需接受外部数据和临床结局的检验。

4.2. 模态缺失、批次效应与域偏移

真实世界中,多模态数据很少完整同步。患者可能有 CT 而无 MRI,有病理切片而无基因测序,有内镜图像而缺少长期随访。若模型训练时只纳入资料完整者,容易形成选择偏倚,使模型更适用于资源

充足、病情较轻或随访较好的患者。解决模态缺失可采用缺失指示变量、模态丢弃训练、生成式补全、多任务学习或后期融合框架，但更重要的是在研究方案中说明缺失机制，并比较完整病例与缺失病例的基线差异。

批次效应是融合模型的常见隐患。不同医院的内镜品牌、影像协议、病理扫描仪、染色深浅和测序平台均可能成为模型识别中心来源的线索。若训练集和测试集按随机样本划分，模型可能在同一中心分布内表现良好，却无法推广到其他医院。外部验证应以中心或时间为单位进行，而不是仅在同一数据库中随机拆分。对影像和病理模型，可采用标准化预处理、颜色归一化、重采样、域自适应和模型校准等方法，但这些技术不能替代真实外部验证。

域偏移还可能来自诊疗策略变化。例如，胃癌围手术期治疗、直肠癌全程新辅助治疗和免疫治疗适应证的变化，会改变模型输入与结局之间的关系。一个在旧治疗时代训练的预后模型，未必适用于新的治疗路径。临床部署后应持续监测模型性能，包括总体判别能力、校准程度、亚组公平性和临床行为改变，并预设模型更新和再验证机制。

4.3. 可解释性与临床可读性

智能预测诊断系统若不能解释其判断依据，临床采纳会受到限制。可解释性并不等同于简单模型，而是指系统能够以医生可理解、可核验的形式展示证据。内镜和病理图像模型可通过热图、病灶框、关注区域和关键视野回放显示证据位置；影像组学模型可报告主要特征类别和与临床变量的相对贡献；多组学模型可提供通路层面解释。对于融合模型，除全局变量重要性外，还应说明不同模态在个体预测中的贡献，避免出现某一低质量模态主导结果而医生无从察觉的情况。此类解释应服务于医生-算法协同，而不是把模型分数包装成不可质疑的结论[28]。

临床可读性还包括输出阈值、风险分层和建议文本。AUC 虽便于比较模型优劣，但医生更关心某一风险阈值下的漏诊率、误诊率、需要追加检查人数和可能避免的不必要治疗。模型报告应同时提供校准曲线、决策曲线、混淆矩阵和关键亚组表现。对于高风险决策，如器官保留、免疫治疗选择或减少手术范围，模型应作为辅助证据，而非唯一依据。

多模态数据融合是提升急危重症 AI 预测性能的重要途径，根据融合发生的阶段可分为早期融合、中期融合、晚期融合和混合融合四种策略。表 1 汇总了各类策略在医疗 AI 中的典型应用、性能表现及优缺点。当前急危重症 AI 研究普遍存在以下局限性：1) 多数模型基于单中心回顾性数据(如 MIMIC)，外部泛化能力未经充分验证；2) 临床数据缺失模式多为非随机缺失(MNAR)，常规缺失处理方法可能引入系统性偏倚；3) 回顾性设计无法完全模拟实时预警场景中的传感器故障、记录延迟等复杂因素，可能高估真实世界性能；4) 多模态融合策略的选择缺乏统一标准，不同研究的结论可比性有限；5) 仅有极少数模型进入前瞻性随机对照试验，预测性能改善是否转化为患者结局获益尚待证实；6) SHAP 等事后解释方法提供的是近似解释，不同方法可能给出不一致的特征重要性排序，其临床可信度有待系统评估。未来研究应遵循 TRIPOD-AI 等标准化报告框架，加强多中心前瞻性验证，并在融合策略选择上建立系统比较范式。

Table 1. Comparison of multimodal fusion strategies in medical imaging applications
表 1. 多模态融合策略在医学影像中的应用比较

融合策略	典型应用案例	多模态组合	性能表现	优势	局限性
早期融合	肝癌超声多模态诊断	B-mode、SWE、CEUS 与临床数据	AUC 达 0.9854	实现方式相对简单，能够捕捉不同模态之间的低层级交互信息	容易导致特征维度过高，对缺失模态较为敏感

续表

中期融合	脑肿瘤分割	MRI 多序列, 包括 T1、T1c、T2 和 FLAIR	分割精度较高, 模型鲁棒性较好	融合方式灵活, 可在不同网络层级选择性整合多模态特征	网络结构较复杂, 需要进行精细化参数调节
晚期融合	卵巢癌风险分层	CT 影像组学特征与全切片图像	多模态模型整体表现优于单模态模型	各模态可独立建模, 结果解释性较好	对模态间交互信息的利用相对有限
混合融合	肝癌超声多模态诊断中的早期 - 晚期组合融合	B-mode、SWE、CEUS 与临床数据	AUC 达 0.9854, 较单模态模型提升 13.8%~25.88%	兼顾低层特征交互与高层语义互补, 能够充分整合多源信息	模型架构最为复杂, 计算开销较大

5. 从公共数据库到临床验证：转化路径

胃肠道肿瘤智能预测诊断系统的转化可分为五个连续阶段。第一阶段是公共数据库探索，目标是明确候选特征、分子通路或影像表型与临床结局之间的关系。此阶段应重视数据清洗、重复样本处理和外部数据集验证，避免过度依赖单一公共队列。第二阶段是本地单中心回顾性建模，目标是将公共数据库发现迁移到本院实际数据，明确数据采集流程和标签定义。第三阶段是多中心外部验证，重点考察模型在不同设备、医生、患者构成和治疗策略中的稳定性。

第四阶段是前瞻性静默试验，即模型在真实工作流中运行，但预测结果暂不影响临床决策。该阶段可评估数据接口、响应时间、失败率、异常样本比例和实时性能，并比较模型预测与最终诊断或结局之间的一致性。第五阶段是临床影响评价，可采用前瞻性队列、随机对照试验或阶梯楔形设计，评估系统是否改善诊断效率、风险分层、治疗选择、患者结局或医疗资源利用。人工智能干预的临床研究报告可参考 TRIPOD、CLAIM、SPIRIT-AI、CONSORT-AI、DECIDE-AI 和 PROBAST + AI 等指南或工具[29]-[34]。

人工智能在急危重症护理中的研究成果向临床转化，需与当前主流医疗器械监管框架有效衔接。美国 FDA 将 AI 软件视为软件即医疗器械(SaMD)，依据信息对临床决策的影响程度和疾病严重程度进行风险分级。急危重症预警及预测类 AI 多涉及“危急”病情且输出驱动临床决策，通常归入较高风险类别，需通过 510(k)、De Novo 或上市前批准(PMA)路径申报。FDA 于 2024 年批准的 Prenosis Sepsis ImmunoScore 即为脓毒症 AI 预测工具通过 De Novo 路径上市的标志性案例。对于持续学习的算法，FDA 要求提交预定变更控制计划(PCCP)，明确上市后可允许的算法更新范围及验证方法，避免每次迭代均需重新审批。中国 NMPA 已建立由 19 项指导原则构成的人工智能医疗器械监管体系，明确辅助决策类软件多数按第三类医疗器械管理，截至 2025 年已批准超过 110 款深度学习辅助决策软件。研究者应在概念验证阶段即遵循良好机器学习实践(GMLP)，在前瞻性验证阶段主动与监管机构沟通(如 FDA 的 Q-Submission)，在上市后建立真实世界性能监测体系。将研发各阶段与监管要求提前映射，可显著缩短转化周期，提升申报成功率。

真实世界验证中，应区分技术性能、临床性能和系统性能。技术性能关注模型是否准确；临床性能关注模型是否改善医生判断和患者管理；系统性能则关注其能否稳定嵌入信息系统、是否造成额外工作负担、是否便于审计和持续更新。很多模型停留在技术性能阶段，是因为研究设计没有回答临床使用中最关键的问题：谁在何时看到模型输出，如何处理低置信预测，模型建议与医生判断不一致时由谁承担责任。

6. 真实世界部署中的风险控制

智能系统进入临床后，风险并不只来自算法错误，还来自使用方式。首先是适用人群风险。模型训练对象若主要为三级医院手术患者，就不应直接用于社区筛查或晚期姑息治疗患者。其次是自动化依赖风险。医生可能因系统提示而忽视自身判断，尤其在模型界面突出显示高风险或低风险标签时更明显。系统应保留原始证据、提供不确定性提示，并允许医生记录不同意见。

其次是公平性和可及性问题。不同年龄、性别、民族、地区、经济状况和医院等级患者的数据质量和检查机会不同，模型可能在某些亚组中性能下降。胃肠道肿瘤筛查和诊断本已存在资源差异，若智能系统只在高质量数据环境中表现良好，反而可能扩大差距。研究报告应尽量提供关键亚组的判别、校准和净获益分析，而不是仅报告总体指标。

再次是监管和责任边界。预测诊断系统若被用于辅助临床决策，应明确软件版本、训练数据范围、性能指标、更新策略和不良事件报告机制。模型更新不是简单重新训练，而是新的医疗软件版本管理过程，需要再验证和文档记录。临床机构也应建立本地化监测机制，包括运行日志、模型漂移、接口失败、异常输出和医生反馈。

最后是隐私与数据安全。多中心训练和模型共享需满足数据脱敏、访问控制、审计追踪和伦理审批要求。联邦学习、隐私计算和安全多方计算为跨机构协作提供了技术可能，但其有效性仍取决于数据标准、样本质量和治理流程。对患者而言，重要的不只是数据是否被加密，还包括其数据被用于何种模型、模型输出如何影响诊疗、出现错误时如何追溯和纠正。

7. 展望

未来胃肠道肿瘤智能预测诊断系统的发展，可能呈现三个方向。第一，从单任务模型转向诊疗连续体模型。患者从筛查、确诊、分期、治疗到随访的过程中会不断产生新数据，模型应支持动态更新风险，而不是在某一时间点给出固定结论。第二，从封闭数据集竞争转向可复现的真实世界评估。公开代码、数据字典、预处理流程、模型卡和外部验证报告将成为判断研究可信度的重要依据。第三，从黑箱预测转向机制相关和知识约束的融合建模。结合病理形态、影像表型、免疫微环境和分子通路，有望提高模型解释力并形成可检验的生物学假设。

同时，应避免对人工智能作过度承诺。胃肠道肿瘤诊疗中的许多关键问题，如晚期患者治疗序贯、复发风险管理、器官保留选择和老年患者综合评估，并不能仅靠模型分数解决。智能系统的合理定位是帮助医生更早发现异常、更系统整合证据、更稳定执行随访和更及时识别高风险患者。其最终价值应由患者获益、医疗质量和资源效率来衡量，而不是由单一算法指标决定。

8. 结语

从公共数据库到临床验证，是胃肠道肿瘤智能预测诊断系统走向真实世界应用的必经路径。公共数据库提供了疾病分子图谱和多源数据探索空间，真实世界数据则检验模型是否能够适应临床复杂性。内镜、影像、病理、组学和临床变量的融合建模，有望提升早诊早筛、分期评估、疗效预测和预后分层的精度，但其临床价值取决于严谨的数据治理、外部验证、前瞻性评价和持续监管。未来研究应减少只追求算法指标的倾向，将模型设计、系统部署和临床影响评价纳入同一研究框架，推动胃肠道肿瘤智能诊疗从“可发表模型”走向“可验证系统”。

基金项目

1) 自然科学院东学院科技创新基金项目。项目编号：XYZK2502。

2) 自然科学基金项目。项目编号: 2025RC1016。

参考文献

- [1] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., *et al.* (2024) Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, **74**, 229-263. <https://doi.org/10.3322/caac.21834>
- [2] The Cancer Genome Atlas Research Network (2014) Comprehensive Molecular Characterization of Gastric Adenocarcinoma. *Nature*, **513**, 202-209. <https://doi.org/10.1038/nature13480>
- [3] The Cancer Genome Atlas Network (2012) Comprehensive Molecular Characterization of Human Colon and Rectal Cancer. *Nature*, **487**, 330-337. <https://doi.org/10.1038/nature11252>
- [4] The Cancer Genome Atlas Research Network (2017) Integrated Genomic Characterization of Oesophageal Carcinoma. *Nature*, **541**, 169-175. <https://doi.org/10.1038/nature20805>
- [5] Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., *et al.* (2013) NCBI GEO: Archive for Functional Genomics Data Sets—Update. *Nucleic Acids Research*, **41**, D991-D995. <https://doi.org/10.1093/nar/gks1193>
- [6] Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., *et al.* (2013) Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the CBioPortal. *Science Signaling*, **6**, pl1. <https://doi.org/10.1126/scisignal.2004088>
- [7] Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., *et al.* (2013) The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *Journal of Digital Imaging*, **26**, 1045-1057. <https://doi.org/10.1007/s10278-013-9622-7>
- [8] Wang, P., Berzin, T.M., Glissen Brown, J.R., Bharadwaj, S., Becq, A., Xiao, X., *et al.* (2019) Real-Time Automatic Detection System Increases Colonoscopic Polyp and Adenoma Detection Rates: A Prospective Randomised Controlled Study. *Gut*, **68**, 1813-1819. <https://doi.org/10.1136/gutjnl-2018-317500>
- [9] Repici, A., Badalamenti, M., Maselli, R., Correale, L., Radaelli, F., Rondonotti, E., *et al.* (2020) Efficacy of Real-Time Computer-Aided Detection of Colorectal Neoplasia in a Randomized Trial. *Gastroenterology*, **159**, 512-520.e7. <https://doi.org/10.1053/j.gastro.2020.04.062>
- [10] Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., van Stiphout, R.G.P.M., Granton, P., *et al.* (2012) Radiomics: Extracting More Information from Medical Images Using Advanced Feature Analysis. *European Journal of Cancer*, **48**, 441-446. <https://doi.org/10.1016/j.ejca.2011.11.036>
- [11] Gillies, R.J., Kinahan, P.E. and Hricak, H. (2016) Radiomics: Images Are More than Pictures, They Are Data. *Radiology*, **278**, 563-577. <https://doi.org/10.1148/radiol.2015151169>
- [12] Aerts, H.J.W.L., Velazquez, E.R., Leijenaar, R.T.H., Parmar, C., Grossmann, P., Carvalho, S., *et al.* (2014) Correction: Corrigendum: Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nature Communications*, **5**, Article No. 4006. <https://doi.org/10.1038/ncomms5644>
- [13] van Griethuysen, J.J.M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., *et al.* (2017) Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research*, **77**, e104-e107. <https://doi.org/10.1158/0008-5472.can-17-0339>
- [14] Zwanenburg, A., Vallières, M., Abdalah, M.A., Aerts, H.J.W.L., Andrearczyk, V., Apte, A., *et al.* (2020) The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-Based Phenotyping. *Radiology*, **295**, 328-338. <https://doi.org/10.1148/radiol.2020191145>
- [15] Dong, D., Fang, M.-J., Tang, L., Shan, X.-H., Gao, J.-B., Giganti, F., *et al.* (2020) Deep Learning Radiomic Nomogram Can Predict the Number of Lymph Node Metastasis in Locally Advanced Gastric Cancer: An International Multicenter Study. *Annals of Oncology*, **31**, 912-920. <https://doi.org/10.1016/j.annonc.2020.04.003>
- [16] Dong, D., Tang, L., Li, Z.-Y., Fang, M.-J., Gao, J.-B., Shan, X.-H., *et al.* (2019) Development and Validation of an Individualized Nomogram to Identify Occult Peritoneal Metastasis in Patients with Advanced Gastric Cancer. *Annals of Oncology*, **30**, 431-438. <https://doi.org/10.1093/annonc/mdz001>
- [17] Nie, K., Shi, L., Chen, Q., Hu, X., Jabbour, S.K., Yue, N., *et al.* (2016) Rectal Cancer: Assessment of Neoadjuvant Chemoradiation Outcome Based on Radiomics of Multiparametric MRI. *Clinical Cancer Research*, **22**, 5256-5264. <https://doi.org/10.1158/1078-0432.ccr-15-2997>
- [18] Liu, Z., Zhang, X.Y., Shi, Y.J., Wang, L., Zhu, H., Tang, Z., *et al.* (2017) Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Clinical Cancer Research*, **23**, 7253-7262. <https://doi.org/10.1158/1078-0432.ccr-17-1038>

- [19] Shaish, H., Aukerman, A., Vanguri, R., Spinelli, A., Armenta, P., Jambawalikar, S., *et al.* (2020) Radiomics of MRI for Pretreatment Prediction of Pathologic Complete Response, Tumor Regression Grade, and Neoadjuvant Rectal Score in Patients with Locally Advanced Rectal Cancer Undergoing Neoadjuvant Chemoradiation: An International Multicenter Study. *European Radiology*, **30**, 6263-6273. <https://doi.org/10.1007/s00330-020-06968-6>
- [20] Kather, J.N., Pearson, A.T., Halama, N., Jäger, D., Krause, J., Loosen, S.H., *et al.* (2019) Deep Learning Can Predict Microsatellite Instability Directly from Histology in Gastrointestinal Cancer. *Nature Medicine*, **25**, 1054-1056. <https://doi.org/10.1038/s41591-019-0462-y>
- [21] Echle, A., Grabsch, H.I., Quirke, P., van den Brandt, P.A., West, N.P., Hutchins, G.G.A., *et al.* (2020) Clinical-Grade Detection of Microsatellite Instability in Colorectal Tumors by Deep Learning. *Gastroenterology*, **159**, 1406-1416.e11. <https://doi.org/10.1053/j.gastro.2020.06.021>
- [22] Yamashita, R., Long, J., Longacre, T., Peng, L., Berry, G., Martin, B., *et al.* (2021) Deep Learning Model for the Prediction of Microsatellite Instability in Colorectal Cancer: A Diagnostic Study. *The Lancet Oncology*, **22**, 132-141. [https://doi.org/10.1016/s1470-2045\(20\)30535-0](https://doi.org/10.1016/s1470-2045(20)30535-0)
- [23] Bilal, M., Raza, S.E.A., Azam, A., Graham, S., Ilyas, M., Cree, I.A., *et al.* (2021) Development and Validation of a Weakly Supervised Deep Learning Framework to Predict the Status of Molecular Pathways and Key Mutations in Colorectal Cancer from Routine Histology Images: A Retrospective Study. *The Lancet Digital Health*, **3**, e763-e772. [https://doi.org/10.1016/s2589-7500\(21\)00180-1](https://doi.org/10.1016/s2589-7500(21)00180-1)
- [24] Kather, J.N., Heij, L.R., Grabsch, H.I., Loeffler, C., Echle, A., Muti, H.S., *et al.* (2020) Pan-Cancer Image-Based Detection of Clinically Actionable Genetic Alterations. *Nature Cancer*, **1**, 789-799. <https://doi.org/10.1038/s43018-020-0087-6>
- [25] Lu, M.Y., Williamson, D.F.K., Chen, T.Y., Chen, R.J., Barbieri, M. and Mahmood, F. (2021) Data-Efficient and Weakly Supervised Computational Pathology on Whole-Slide Images. *Nature Biomedical Engineering*, **5**, 555-570. <https://doi.org/10.1038/s41551-020-00682-w>
- [26] Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D.A., Barnholtz-Sloan, J.S., Velázquez Vega, J.E., *et al.* (2018) Predicting Cancer Outcomes from Histology and Genomics Using Convolutional Networks. *Proceedings of the National Academy of Sciences*, **115**, E2970-E2979. <https://doi.org/10.1073/pnas.1717139115>
- [27] Jiang, Y., Zhang, Z., Wang, W., Huang, W., Chen, C., Xi, S., *et al.* (2023) Biology-Guided Deep Learning Predicts Prognosis and Cancer Immunotherapy Response. *Nature Communications*, **14**, Article No. 5135. <https://doi.org/10.1038/s41467-023-40890-x>
- [28] Topol, E.J. (2019) High-Performance Medicine: The Convergence of Human and Artificial Intelligence. *Nature Medicine*, **25**, 44-56. <https://doi.org/10.1038/s41591-018-0300-7>
- [29] Collins, G.S., Reitsma, J.B., Altman, D.G. and Moons, K.G.M. (2015) Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *Annals of Internal Medicine*, **162**, 55-63. <https://doi.org/10.7326/m14-0697>
- [30] Mongan, J., Moy, L. and Kahn, C.E. (2020) Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. *Radiology: Artificial Intelligence*, **2**, e200029. <https://doi.org/10.1148/ryai.2020200029>
- [31] Cruz Rivera, S., Liu, X., Chan, A., Denniston, A.K., Calvert, M.J., Darzi, A., *et al.* (2020) Guidelines for Clinical Trial Protocols for Interventions Involving Artificial Intelligence: The SPIRIT-AI Extension. *Nature Medicine*, **26**, 1351-1363. <https://doi.org/10.1038/s41591-020-1037-7>
- [32] Liu, X., Cruz Rivera, S., Moher, D., Calvert, M.J., Denniston, A.K., Chan, A., *et al.* (2020) Reporting Guidelines for Clinical Trial Reports for Interventions Involving Artificial Intelligence: The CONSORT-AI Extension. *Nature Medicine*, **26**, 1364-1374. <https://doi.org/10.1038/s41591-020-1034-x>
- [33] Vasey, B., Nagendran, M., Campbell, B., Clifton, D.A., Collins, G.S., Denaxas, S., *et al.* (2022) Reporting Guideline for the Early-Stage Clinical Evaluation of Decision Support Systems Driven by Artificial Intelligence: Decide-AI. *Nature Medicine*, **28**, 924-933. <https://doi.org/10.1038/s41591-022-01772-9>
- [34] Moons, K.G.M., Damen, J.A.A., Kaul, T., Hooft, L., Andaur Navarro, C., Dhiman, P., *et al.* (2025) PROBAST + AI: An Updated Quality, Risk of Bias, and Applicability Assessment Tool for Prediction Models Using Regression or Artificial Intelligence Methods. *BMJ*, **388**, e082505. <https://doi.org/10.1136/bmj-2024-082505>