



中国医学影像学杂志

Chinese Journal of Medical Imaging

ISSN 1005-5185, CN 11-3154/R

## 《中国医学影像学杂志》网络首发论文

题目：结直肠癌 CT 影像数据库构建及质量控制专家共识  
作者：周俊林, 洪楠, 张惠茅, 陈敏, 刘士远, 李昇霖, 袁隆, 卢婷  
收稿日期：2024-11-06  
网络首发日期：2024-12-31  
引用格式：周俊林, 洪楠, 张惠茅, 陈敏, 刘士远, 李昇霖, 袁隆, 卢婷. 结直肠癌 CT 影像数据库构建及质量控制专家共识[J/OL]. 中国医学影像学杂志.  
<https://link.cnki.net/urlid/11.3154.R.20241230.1331.002>



**网络首发：**在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

**出版确认：**纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188, CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

## 结直肠癌 CT 影像数据库构建及质量控制专家共识

中华医学会放射学分会医学影像人工智能工作组；\*通信作者 周俊林 [ery\\_zhoujl@lzu.edu.cn](mailto:ery_zhoujl@lzu.edu.cn)；洪楠 [hongnan@pkuph.edu.cn](mailto:hongnan@pkuph.edu.cn)；张惠茅 [huimao@jlu.edu.cn](mailto:huimao@jlu.edu.cn)；陈敏 [cjr.chenmin@vip.163.com](mailto:cjr.chenmin@vip.163.com)；刘士远 [cjr.liushi yuan@vip.163.com](mailto:cjr.liushi yuan@vip.163.com)

【基金项目】国家自然科学基金面上项目（82371914）；兰州大学第二医院“萃英科技创新”计划（CY2021-ZD-01）

【摘要】结直肠癌是临床最常见的消化系统恶性肿瘤之一。基于人工智能的结直肠癌早期检出及进一步辅助临床诊疗，从而实现患者长期受益具有重要临床意义。人工智能软件的开发与验证均依赖高质量、大数据量和有标注的结直肠癌影像数据集。本文以如何构建高质量结直肠癌 CT 数据库为例，从数据库描述、病灶标注与储存、数据库的质量评估与维护等方面，提出数据库搭建的完整流程，旨在为搭建结直肠癌 CT 影像数据库提供参考，保障数据库源头资料的高质量 and 可挖掘性，促进医疗影像人工智能行业生态可持续健康发展，加速结直肠癌 CT 影像人工智能相关产业的研发与应用。

【关键词】结直肠肿瘤；数据库；体层摄影术，X 线计算机；质量控制；专家共识

【中图分类号】R445.2；R735.3+5；R735.3+7 【DOI】10.3969/j.issn.1005-5185.2025.01.001

### Expert Consensus on CT Image Database Construction and Quality Control for Colorectal Cancer

Artificial Intelligence Working Group of Chinese Society of Radiology Chinese Medical Association; \*Address Correspondence to: ZHOU Junlin, E-mail: [ery\\_zhoujl@lzu.edu.cn](mailto:ery_zhoujl@lzu.edu.cn); HONG Nan, E-mail: [hongnan@pkuph.edu.cn](mailto:hongnan@pkuph.edu.cn); ZHANG Huimao, E-mail: [huimao@jlu.edu.cn](mailto:huimao@jlu.edu.cn); CHEN Min, E-mail: [cjr.chenmin@vip.163.com](mailto:cjr.chenmin@vip.163.com); LIU Shiyuan, E-mail: [cjr.liushi yuan@vip.163.com](mailto:cjr.liushi yuan@vip.163.com)

【Abstract】Colorectal cancer is one of the most common malignant tumors of the digestive system in clinical practice. The early detection of colorectal cancer based on artificial intelligence and its further assistance in clinical diagnosis and treatment hold significant clinical importance for achieving long-term benefits for patients. The development and validation of artificial intelligence software rely on high-quality, large-volume, and annotated colorectal cancer imaging datasets. This paper aims to provide a reference for constructing a high-quality colorectal cancer CT database, taking the construction of the database as an example. It discusses the complete process of database establishment, including database description, lesion annotation and storage, database quality evaluation and maintenance. The purpose is to ensure the high quality and exploitability of the source materials in the database, promote the sustainable and healthy development of the medical imaging artificial intelligence industry ecosystem, and accelerate the research, development, and application of industries related to artificial intelligence in colorectal cancer CT imaging.

【Key words】Colorectal neoplasms; Database; Tomography, X-ray computed; Quality control; Expert consensus

Chinese Journal of Medical Imaging, 2025, 33 (1):

结直肠癌（colorectal cancer, CRC）是国内外发病率率和死亡率均居前列的恶性肿瘤之一<sup>[1-2]</sup>，预计到2040年，全球新发病例将增加到320万，死亡人数将增加到160万<sup>[1]</sup>。2023年我国CRC新发病例约56万例，占所有癌症新发病例的11.7%；死亡病例约28万例，占所有癌症死亡病例的10.9%，居癌症死亡原因第5位<sup>[3]</sup>。国内CRC患者的筛查和诊断相对滞后，早期发现和精确诊断是提高患者生存率和生活质量的关键。随着医学影像的快速发展和应用，CT已经成为CRC临床指南中推荐的重要检查方法，也因此产生了大量医学影像数据<sup>[4-5]</sup>。在全球人工智能快速发展的背景下，我国的医学影像人工智能发展也颇受影响，然而优质的数据资源和图像质量是训练大模型的基础，如何控制

CT图像质量和规范构建CRC CT影像数据库，从而有效推进目前临床人工智能领域影像大模型的搭建，尚需要统一标准。

目前，基于腹部CRC CT图像构建的影像人工智能模型在CRC诊断、疗效和预后等方面具有良好的效果<sup>[6-8]</sup>，但均存在样本量较小、单中心研究以及图像参数、图像质量不同导致研究偏移等共性不足。由美国国家癌症研究所癌症成像项目资助建立的癌症医学影像档案库中具备230例CRC CT影像数据资料，还具有197例CRC肝转移的CT影像数据资料，且同时具备肝转移病灶的标注信息<sup>[9-10]</sup>。目前，国内缺乏第三方数据库，且影像数据量偏小、多中心合作机构有限，所拥有数据量和质量存在差异，造成数据研究分析过

程中具有偏倚和系统性偏差，导致CRC相关人工智能产品开发应用存在诸多困难。为此，中华医学会放射学分会医学影像人工智能工作组组织相关影像学专家，结合国内临床实际情况，并参考国内外最新指南和文献，以建设国内高质量CRC CT影像数据库的具体问题为导向，共同对CRC CT影像数据库的构建流程、数据采集和储存、质量控制与评价、运行与维护等内容达成共识，旨在为CRC CT影像标准数据库建设和质量管理给予示范及引导，促进医疗影像人工智能行业生态可持续健康发展，加速CRC CT影像人工智能相关产业的研发与应用。

### 1 CRC CT 影像数据库描述

依据《卫生信息数据元数据规范》中对数据库的定义，本文将CRC CT影像数据库定义为以CRC对比增强CT扫描影像为主题，可以被标注并被计算机识别和处理的CT影像数据集。CRC CT影像数据库构建包括数据库软硬件构建、数据采集和储存、数据库质量控制、数据库运行与维护、数据库质量评价等5个方面。明确数据库入库标准、规范采集流程和标准化化管理对CRC CT影像数据库的建设具有重要意义。

### 2 数据采集和储存

数据采集和储存是数据库构建的核心环节，高质量、高标准的数据才能为临床、教学和科研提供更加科学、有效的依据。为保证数据的完整性、规范性和可审计性，建议在数据采集和储存中采用以下流程（图1）。数据采集前，应取得数据来源单位的伦理委员会批准，并采用详实具体的方案保护患者隐私。根据数据库的建设目的，提前制订清晰的纳入、排除标准和完整的采集参数，有效地提高数据库的质量、可信度和利用率。采集过程中，应符合临床研究规范，患者的知情权、同意书和相关补偿应满足法律法规要求。采集完成后，应对采集数据进行整理和标注，并录入数据库储存。具体而言，CRC CT影像数据库数据采集应包括临床诊疗数据、病理基因数据、原始影像数据和影像标注数据。

**2.1 临床诊疗数据** 临床诊疗数据采集应符合相关伦理和法律法规，采用规范的方式从医院信息系统及影像归档和通信系统中获取。临床诊疗数据应包括患者基本人口学信息（年龄、性别、民族、职业、体重指数、饮食习惯等）、流行病学信息（疾病史、家族史）、血清肿瘤标志物信息（癌胚抗原、甲胎蛋白、CA19-9和

CA125等）、有无其他器官转移（肝脏、肺等）、转移类型（同时型或异时型）、临床TNM分期和内镜检查信息等（进镜深度、肿瘤形态、大小、距肛缘位置、局部浸润范围以及术前有无肠腔狭窄、肠梗阻或穿孔等）。CRC患者预后状况与其采取的治疗方法密切相关，应详细记录患者诊疗策略，如新辅助治疗（化疗、放疗、靶向和免疫治疗等）、手术（内镜下切除或根治性切除术等）、术后辅助治疗方案以及治疗周期和是否更换治疗方案等信息。若患者合并其他器官转移，也应记录针对转移病灶的治疗策略和周期。随访是评估患者疗效和预后的主要依据，建议制订完善的定期随访计划，详细记录患者1年、3年、5年无进展生存期和总生存期。

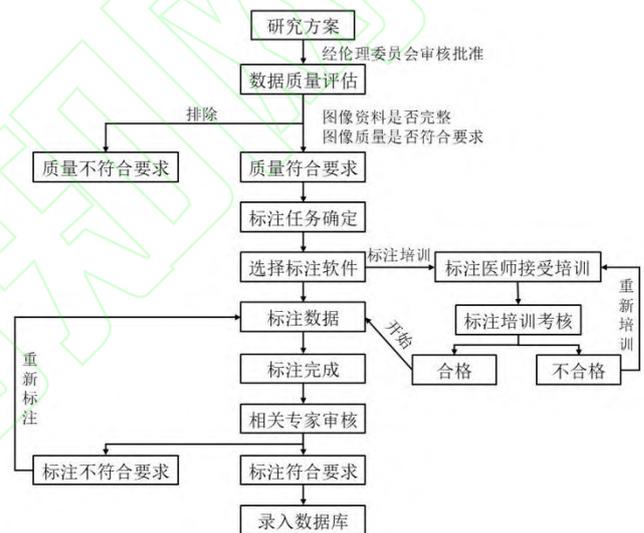


图1 CRC CT影像数据库构建流程

**2.2 病理基因数据** 为更好地开展影像与病理的对照分析，提高人工智能模型的精准性和可靠性，建议采集并录入CRC患者的手术病理结果，如肿瘤病理分型（腺癌、鳞癌或黏液腺癌等）、分化程度（高分化、中分化、低分化或未分化）、浸润深度（pT分期）、邻近淋巴结转移（pN分期）、肿瘤出芽、脉管神经侵犯、近端和远端切缘状况（R0切除）、系膜/环周切缘状况、基因状态（KRAS、NRAS、BRAF、HER-2、MMR蛋白、微卫星不稳定性等）及Ki-67指数等。若手术病理信息中具有其他转移病灶的病理结果，应一并详细记录。采用新辅助治疗的患者，也应记录病理学专家评估的肿瘤退缩分级。

**2.3 原始影像数据** 医学数字成像和通信（digital imaging and communications in medicine, DICOM）标

准是医学影像储存、传输和显示的常用格式，CRC CT 数据库中的原始影像数据均应采用DICOM格式录入。依据每个病例采用的治疗方式，依次采集治疗前（基线）和治疗后不同时间点的腹部薄层图像（平扫及增强三期），图像视野应覆盖膈肌至耻骨联合下缘，包含CRC原位肿瘤、邻近区域及远处器官等完整信息。扫描体位均宜采取仰卧位且双臂举过头顶。记录采集机器的厂家和型号、扫描参数、造影剂种类及注射速度等。依据国家医疗器械行业标准《医疗器械唯一标识基本要求：YY/T 1630—2018》<sup>[11]</sup>，原始影像数据采集完成后，应将每例患者的数据样本或每个影像序列存放的文件夹设置唯一的编码，以便后续记录及溯源核对信息。建议按照CRC原发灶部位或有无其他部位转移进行分类储存，并补充条件筛选功能，以便对数据进行重新分组与归类。

2.4 影像标注数据 标准化高质量的影像标注数据是保证人工智能模型训练准确性的基础，能有效促进人工智能产品的开发和应用<sup>[12]</sup>。而获取可靠的影像标注数据很大程度上依赖制订和落实完善的标注流程和规范。CRC CT影像数据库的标注流程建议见图1，标注对象应包含原位肿瘤和转移病灶，标注内容应包含影像征象及影像图像的标注。标注软件可选择ITK-SNAP（<https://www.itksnap.org>）和3D Slicer（<https://www.slicer.org>）等开源软件。标注过程中应安排相关医学专家具体负责和持续参与数据的标注过程，严格把控数据标注的质量。

为提高影像征象标注的效率和质量，避免遗漏、缺失和错误，推荐使用结构化报告进行标注，模板见表1。对于合并肝转移或肺转移的病例，可使用肝转移瘤及肺转移瘤标注结构化报告，模板见表2、3。

表 1 CRC CT 数据库原位肿瘤标注结构化报告

影像编码：	性别：	年龄：	检查时间：
检查项目：		临床诊断：	
肿瘤位置	<input type="checkbox"/> 左半结肠 <input type="radio"/> 横结肠（左） <input type="radio"/> 乙状结肠 <input type="radio"/> 降结肠 <input type="radio"/> 结肠脾曲	<input type="checkbox"/> 右半结肠 <input type="radio"/> 横结肠（右） <input type="radio"/> 盲肠 <input type="radio"/> 升结肠 <input type="radio"/> 结肠肝曲	
肿瘤形态	<input type="checkbox"/> 肿块型	<input type="checkbox"/> 浸润型	<input type="checkbox"/> 溃疡型
肿瘤大小	长径____mm 上下径____mm	短径____mm 厚度____mm	
肿瘤密度（CT 值）	平扫____Hu 静脉期____Hu	动脉期____Hu 延迟期____Hu	
肿瘤 T 分期	<input type="checkbox"/> 侵犯至黏膜下层（T1） <input type="checkbox"/> 肿瘤侵犯固有肌层，但未穿透固有肌层（T2） <input type="checkbox"/> 肿瘤突破固有肌层（T3）<5 mm <input type="checkbox"/> 肿瘤突破固有肌层（T3）≥3 mm <input type="checkbox"/> 肿瘤侵犯超出腹膜覆盖的表面（T4a） <input type="checkbox"/> 侵犯邻近器官（T4b）		
淋巴结	区域可疑阳性淋巴结数目____个，最大短径____mm 腹膜后可疑阳性淋巴结数目____个，最大短径____mm 异质性强化淋巴结数目____个，最大短径____mm		
远处转移	<input type="checkbox"/> 肝脏转移 其他转移部位_____	<input type="checkbox"/> 肺部转移	<input type="checkbox"/> 腹膜种植转移
肿瘤环周状况	累及环周范围	<input type="checkbox"/> >1/2	<input type="checkbox"/> >1/2
	肿瘤外壁	<input type="checkbox"/> 光滑	<input type="checkbox"/> 毛糙
	肿瘤与肠壁间强化	<input type="checkbox"/> 有	<input type="checkbox"/> 无
	小动脉征	<input type="checkbox"/> 有	<input type="checkbox"/> 无
	壁外血管侵犯	<input type="checkbox"/> 有	<input type="checkbox"/> 无

续表 1

影像编码: 检查项目:	性别:	年龄: 临床诊断:	检查时间:
	脂肪间隙结节	<input type="checkbox"/> 有	<input type="checkbox"/> 无
	脂肪间隙状况	<input type="checkbox"/> 清晰	<input type="checkbox"/> 浑浊
	其他_____		
其他异常征象	肿瘤穿孔	<input type="checkbox"/> 有	<input type="checkbox"/> 无
	肠梗阻	<input type="checkbox"/> 有	<input type="checkbox"/> 无

注: CRC: 结直肠癌; 肿瘤密度 (CT 值) 应对肿瘤实性部分测量至少 3 次以上, 取平均值作为最终测量结果

表 2 CRC CT 数据库肝转移瘤标注结构化报告

影像编码:	性别:	年龄:	检查时间:
1.脂肪肝	<input type="checkbox"/> 有	<input type="checkbox"/> 无	
2.肝转移瘤数目	<input type="checkbox"/> 1~3 个	<input type="checkbox"/> 4~7 个	<input type="checkbox"/> 8 个及以上
3.肝转移瘤分布			
尾状叶	<input type="checkbox"/> S1		
肝左叶	<input type="checkbox"/> S2 <input type="checkbox"/> S3 <input type="checkbox"/> S4		
肝右叶	<input type="checkbox"/> S5 <input type="checkbox"/> S6 <input type="checkbox"/> S7 <input type="checkbox"/> S8		
4.肝转移瘤描述	病灶位于_____段, 最大径_____mm		
	病灶形态	<input type="checkbox"/> 圆形/卵圆形	<input type="checkbox"/> 分叶形
	病灶边界	<input type="checkbox"/> 清晰	<input type="checkbox"/> 不清晰
	病灶质地	<input type="checkbox"/> 均匀	<input type="checkbox"/> 混杂
	病灶密度	平扫_____Hu	动脉期_____Hu
		静脉期_____Hu	延迟期_____Hu
	强化程度	<input type="checkbox"/> 强化 <input type="checkbox"/> 轻度	<input type="checkbox"/> 中度 <input type="checkbox"/> 明显
5.与重要血管的关系			
门静脉右支 主干	<input type="checkbox"/> 未见显示	<input type="checkbox"/> 推移移位	<input type="checkbox"/> 紧邻 <input type="checkbox"/> 分界清楚
门静脉右支 分支	<input type="checkbox"/> 未见显示	<input type="checkbox"/> 推移移位	<input type="checkbox"/> 紧邻 <input type="checkbox"/> 分界清楚
门静脉左支 主干	<input type="checkbox"/> 未见显示	<input type="checkbox"/> 推移移位	<input type="checkbox"/> 紧邻 <input type="checkbox"/> 分界清楚
门静脉左支 分支	<input type="checkbox"/> 未见显示	<input type="checkbox"/> 推移移位	<input type="checkbox"/> 紧邻 <input type="checkbox"/> 分界清楚
肝右静脉 主干	<input type="checkbox"/> 未见显示	<input type="checkbox"/> 推移移位	<input type="checkbox"/> 紧邻 <input type="checkbox"/> 分界清楚
肝中静脉 主干	<input type="checkbox"/> 未见显示	<input type="checkbox"/> 推移移位	<input type="checkbox"/> 紧邻 <input type="checkbox"/> 分界清楚
肝左静脉 主干	<input type="checkbox"/> 未见显示	<input type="checkbox"/> 推移移位	<input type="checkbox"/> 紧邻 <input type="checkbox"/> 分界清楚
下腔静脉 主干	<input type="checkbox"/> 未见显示	<input type="checkbox"/> 推移移位	<input type="checkbox"/> 紧邻 <input type="checkbox"/> 分界清楚
6.肝门区淋巴结	<input type="checkbox"/> 有	<input type="checkbox"/> 无	
	最大淋巴结大小_____mm×_____mm		
7.不确定转移灶	<input type="checkbox"/> 有	<input type="checkbox"/> 无	
8.不确定转移灶位置分布			
尾状叶	<input type="checkbox"/> S1		
肝左叶	<input type="checkbox"/> S2 <input type="checkbox"/> S3 <input type="checkbox"/> S4		
肝右叶	<input type="checkbox"/> S5 <input type="checkbox"/> S6 <input type="checkbox"/> S7 <input type="checkbox"/> S8		
9.其他			

注: CRC: 结直肠癌; 对于CT显示最大直径<10 mm的病灶, 除具有典型转移瘤表现时纳入转移灶, 其他情况建议纳入不确定转移灶, 进一步行肝脏增强MRI进行判断

表3 CRC CT数据库肺转移瘤标注结构化报告

影像编码:	性别:	年龄:	检查时间:
1.肺转移瘤数目	<input type="checkbox"/> 1~3 个	<input type="checkbox"/> 4~7 个	<input type="checkbox"/> 8 个及以上
2.肺转移瘤分布			
右肺	<input type="checkbox"/> 上叶	<input type="checkbox"/> 中叶	<input type="checkbox"/> 下叶
	<input type="checkbox"/> 尖段	<input type="checkbox"/> 外段	<input type="checkbox"/> 背段
	<input type="checkbox"/> 后段	<input type="checkbox"/> 内段	<input type="checkbox"/> 内基底段
	<input type="checkbox"/> 前段		<input type="checkbox"/> 前基底段
			<input type="checkbox"/> 外基底段
			<input type="checkbox"/> 后基底段
左肺	<input type="checkbox"/> 上叶	<input type="checkbox"/> 下叶	
	<input type="checkbox"/> 尖后段	<input type="checkbox"/> 背段	
	<input type="checkbox"/> 前段	<input type="checkbox"/> 内基底段	
	<input type="checkbox"/> 上舌段	<input type="checkbox"/> 前基底段	
	<input type="checkbox"/> 下舌段	<input type="checkbox"/> 外基底段	
		<input type="checkbox"/> 后基底段	
3.肺转移瘤描述			
	位于____叶____段, 病灶直径____mm		
	病灶形态	<input type="checkbox"/> 圆形/卵圆形	<input type="checkbox"/> 分叶形
	病灶边界	<input type="checkbox"/> 清晰	<input type="checkbox"/> 模糊
	病灶质地	<input type="checkbox"/> 均匀	<input type="checkbox"/> 混杂
	病灶密度	平扫____Hu	动脉期____Hu
		静脉期____Hu	延迟期____Hu
	强化程度	<input type="checkbox"/> 不强化	<input type="checkbox"/> 轻度 <input type="checkbox"/> 中度 <input type="checkbox"/> 明显
4.淋巴结			
	肺门区	<input type="checkbox"/> 有	<input type="checkbox"/> 无
	纵隔区	<input type="checkbox"/> 有	<input type="checkbox"/> 无
	最大淋巴结大小____mm×____mm		
5.其他			

注: CRC: 结直肠癌; 若患者肺转移瘤未经手术病理确诊时, 应选择影像学上具有典型转移表现的病灶进行标注

原位肿瘤和肝转移病灶推荐在静脉期进行标注, 肺转移病灶推荐在肺窗进行标注。根据数据库的建设目的, 影像图像标注可使用精细标注沿病灶边缘勾画感兴趣区 (ROI) 或感兴趣容积, 即2D或3D, 勾画方式可采用手动或半自动分割工具辅助。为确保图像标注的准确性和稳定性, CRC CT影像数据库储存数据推荐均使用人工手动3D勾画法, 以适用于监督类影像组学特征提取和多中心人工智能研究开展。标注结果的命名和保存格式应统一, 推荐采用病例影像编码命名和Nifty格式保存。

2.4.1 原位肿瘤标注 由于结直肠为管状柔性中空器官, 独特的解剖特点给CRC图像标注带来了一定挑战。原位肿瘤的图像标注需要结合不同图像和层面确定肿瘤部位、形态和边界, 推荐使用以下3种方式进行

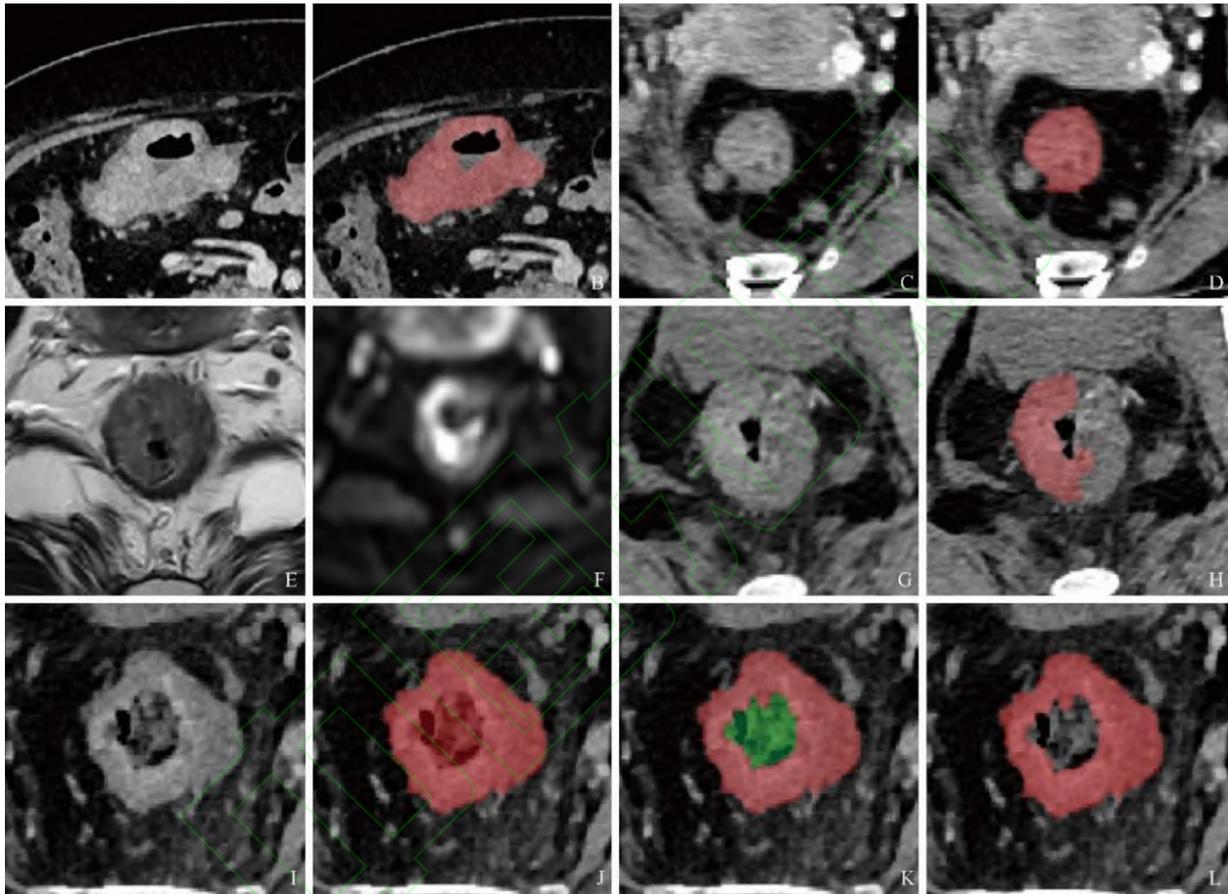
标注, ①直接勾画: 依靠完备的解剖知识, 识别肿瘤累及范围, 直接沿肿瘤边缘勾画 (图2A~D)。②参考对照勾画: 由于CT对肿瘤软组织的分辨率有限, 可参照同时段MRI检查T2WI、扩散加权成像序列进行精细勾画病灶 (图2E~H)。③减影勾画: 沿肿瘤外缘和内缘分别勾画ROI, 然后将外缘和内缘ROI相减, 获得减影ROI (图2I~L)。详细标注方法和原则可参考《结直肠癌CT和MRI标注专家共识 (2020)》<sup>[5]</sup>。

2.4.2 转移病灶标注 适用于转移性病灶的鉴别、淋巴结分期、全身系统性治疗疗效、局部治疗疗效评估等研究数据搭建。①器官转移灶的标注: 主要集中于肝转移、肺转移病灶的精细勾画, 标注方法及流程与肝肿瘤、肺肿瘤的标注一致 (图3)。②转移性淋巴结标注: 由病理证实的区域内淋巴结转移,

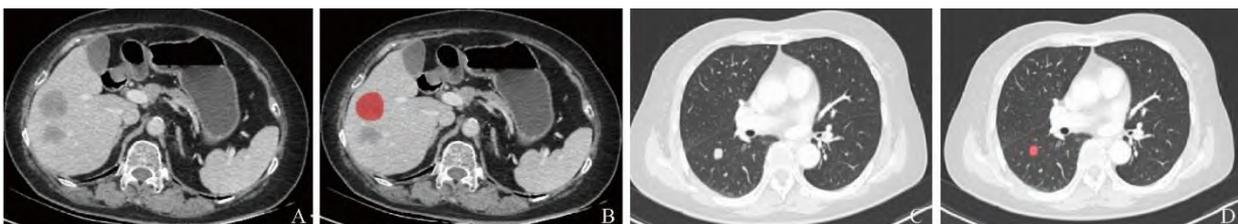
并精细勾画范围内明确与病理对应的转移性淋巴结。转移性淋巴结的识别和勾画存在难度，主要原因为影像-病理的一一对应关系，可参考CT图像中的三维立体定位，以更好地显示和影像-病理对应<sup>[13-14]</sup>。

为保障原位肿瘤及转移病灶标注的一致性，通常建议采用预训练的方式明确病灶的标注范围，由副主任以上医师评估小样本数据预训练的勾画效果。此外，每例数据的勾画至少需要标注2次，并随机抽取小样

本(≥30例)数据进行交互式勾画，评估组内和组间勾画的一致性。采用组内相关系数(ICC)或Dice系数评估一致性的差异，ICC或Dice系数≥0.80认为标注病灶的ROI一致性好，并取两者标注的并集作为标注结果入库储存。当ICC或Dice系数<0.80时，由副主任以上医师对ROI的标注进行判断和修改，并由其确定最终入库存储的结果<sup>[15]</sup>。



**图2** CRC原位肿瘤CT勾画方法。直接勾画法(A~D)：A、B为浸润型结肠癌，A为CT静脉期图像，示结肠肠壁弥漫性增厚，肠腔局部狭窄，沿病变边缘手动勾画ROI；B为标注后图像。C、D为直肠癌，C为CT静脉期图像，示直肠壁明显增厚，呈轻度强化，邻近脂肪间隙见多个小结节影；D为标注后图像。参考对照勾画法(E~H)：E~G分别为T2WI、DWI及CT静脉期图像，示直肠肠壁弥漫性增厚，环周范围<1/2，边缘光整；H为标注后图像。减影勾画法(I~L)：I为CT静脉期图像，示浸润型直肠癌，肠壁弥漫性增厚，周围脂肪间隙模糊；J示沿病灶外缘勾画ROI，K示沿病灶内缘勾画ROI，L为减影后图像



**图3** CRC转移病灶CT勾画。A、B为肝转移病灶，A为CT静脉期图像，示肝转移灶呈圆形或类圆形，边界清晰，增强后较肝实质强化低，勾画时需避开周围血管和其他病灶；B为标注后图像。C、D为肺转移病灶，C为CT平扫肺窗图像，示肺转移灶位于右肺下叶背段，边界清晰，勾画时需避开周围血管；D为标注后图像

### 3 数据库质量控制

3.1 数据采集的质量要求 CRC CT影像数据库的标准数据来源需要考虑多样性和代表性,在数据采集阶段,应考虑选择不同省、市、地区中至少10个以上具有代表性的临床机构。在患者层面,应考虑患者的人口学数据以及危险因素的均衡性,确保采集到的数据在临床意义上具有代表性。CT设备方面主要考虑设备的生产厂家、型号、同一设备的不同成像参数等。

设备厂家应包括但不限于中国联影、中国东软、美国GE、德国Siemens、荷兰Philips、日本佳能等主流设备供应商。在造影剂方面,应记录碘离子造影剂的生产厂家、种类和浓度,同时确保造影剂、高压注射器及针头等在有效期内。在扫描操作方面,应当符合相关腹部影像学检查规范与专家共识,工作人员应经过充分的培训,且为患者进行适当的宣讲工作,确保采集图像的质量。具体细节见表4。

表4 CRC CT数据采集的设备与参数

项目	内容
设备厂家	中国联影、中国东软、美国GE、德国Siemens、荷兰Philips、日本佳能等主流设备
探测器	16~320排
检查前准备	检查前1周禁服含金属的药物或行消化道钡剂造影,24 h内清空肠道;当日禁食4 h以上,检查前15~20 min口服温水500~1 000 ml
体位	仰卧位,足先进,两臂上举,身体置于检查床正中间,水平线对准人体腋中线
扫描参数	管电压100~120 kV,有效管电流200~300 mAs(或自动毫安技术),螺距0.984~1.375,转速0.6~0.8 s/周,视野300~350 mm×300~350 mm
注射参数	均采用静脉内团注对比剂,对比剂含碘浓度270~370 mg/ml,流速2.5~3.5 ml/s,用量80~100 ml
扫描期相	采用三期扫描,动脉期延迟25~30 s,门静脉期延迟50~60 s,实质期延迟120~180 s
层厚与层间距	层厚≤5 mm,层间距≤5 mm
重建方式	标准算法或软组织算法
窗宽与窗位	窗宽200~250 Hu,窗位30~50 Hu

注: CRC: 结直肠癌

3.2 数据管理的质量要求 ①数据脱敏。数据库建设主体责任方应将DICOM数据头文件中所有涉及患者隐私相关的信息统一进行删除处理,如医院名称、医院地址、患者姓名、出生年月、家庭地址、图像扫描信息等。记录数据脱敏的具体过程,以确保数据脱敏的可追溯性和透明度。②数据清洗。数据库建设主体责任方应定期进行数据清洗,并记录清洗操作日志,检查录入数据是否符合标准要求,如非CRC病例、不符合DICOM格式、CT图像不完整或具有严重伪影的数据应统一删除,确保数据内容、格式、标识、图像正确和完整。③数据查重。数据清洗完毕后,建议对数据进行查重,删除重复录入的数据,确保录入数据的唯一性,避免与外部公开数据集重合。

类储存,二者建立一一对应的映射关系。当数据发生更新时,映射关系也同步更新。数据库的访问和使用需要经过管理人员授权,防止发生数据泄露、篡改和丢失等,具体权限分配方案由数据库建设方自行定义。数据库的客户端可采用网络浏览器(如Web页面)实现数据查询,达到网上资源共享。同时实现数据库数据查询、随访、统计分析操作快捷、方便,后续扩展功能强大。在维护方面,数据库责任方应对数据库进行定期检测和更新,并设置数据库版本控制功能,当数据库更新至新版本出现功能不稳定或故障时,可以通过版本控制回溯至上一稳定版本,确保数据库稳定运行。

### 4 数据库运行与维护

数据库建设主体责任方需要对数据库的运行和维护负责,将数据同时存储在本地和云端服务器,防止数据库损坏及数据丢失,确保储存数据的安全性和保密性。数据库中的原始数据和处理后的数据均应归

### 5 数据库质量评价

为保证数据库的质量,数据库建设主体责任方应符合国家医疗器械行业标准——《人工智能医疗器械质量要求和评价第2部分:数据集通用要求》<sup>[16]</sup>对数据库进行全面系统的质量评价。依据数据库的建设目的和用途,应设置具体、详细的评价内容和评价流程。

评估过程中应形成研究资料进行留存,作为后续管理、运行和使用的客观依据<sup>[12]</sup>,以下是该过程的一些关键点:①记录评价流程和结果;②撰写风险分析报告;③提出改进措施计划;④变更管理过程记录;⑤留存审核和批准记录;⑥持续监控记录。评价内容和评估流程记录应贯穿数据库构建、维护和使用全流程中。具体而言,CRC CT影像数据库质量评价内容应包含数据质量特性、数据库风险控制文档和说明文档3个方面。

数据质量特性评价可以通过抽样检查、操作检查和过程验证等方式进行,评价指标包括数据库的完整性、准确性、一致性、唯一性、可靠性、时效性、可访问性、依从性、保密性、可追溯性、可理解性、可移植性、可恢复性和代表性等。

数据库风险管理文档的检查,是对数据库的整体数据质量的风险进行评估,用于控制数据库的分布偏倚、统计偏倚,评估内容主要为,①风险评估:识别和分析数据库可能面临的风险,如数据泄露、数据丢失、不当访问等。应定期对数据库进行维护,及时发现数据库的漏洞,并避免造成数据不正当使用。②风险控制策略:制订相应的风险控制措施,包括预防性措施和应对性措施。③监控和审计:描述如何监控数据库的风险和控制措施的有效性,以及定期审计的流程。④应急计划:提供针对潜在风险的应急响应计划,以减少风险事件的影响。数据库构建主体责任方可以运用专家评议,组织第三方医学专家对数据库进行质量评审,参照行业标准《YY/T 0316—2016医疗器械风险管理对医疗器械的应用》中采用的方法<sup>[12]</sup>,采用设计问卷或评分表等形式,定期开展风险评估和管理。

数据库说明文档是对数据库设计、结构、内容和操作方法的整体系统性描述,目的是帮助使用者及监管机构快速、全面地了解数据库,内容应包含以下4个方面,①目的和范围:明确数据库的目的、预期用途和覆盖范围。②用户指南:提供详细的使用说明,包括如何访问、查询、更新和维护数据库。③操作流程:描述数据库的操作流程,包括数据录入、编辑、审核和导出等步骤。④技术要求:列出使用数据库所需的技术环境和软硬件要求。数据库说明文档是否全面和准确不仅影响使用者的数据库使用率,更影响监管机构对于数据库信任度的评价以及数据库的长久运行。数据库构建主体责任方应制订详细的文档编写标准,设置用户反馈界面,并定期审查和更新说明文档信息。

## 6 小结

医学影像人工智能领域的快速发展为实现精准医疗带来了新机遇。高质量的医学影像标注图像是保障,以深度学习为代表,开发并验证机器学习临床应用模型的核心。单一数据集中心尚无法实现影像大模型的训练和验证,构建CRC CT影像数据库将会有效推动相关研究领域的进一步探索。本共识将会根据影像技术的发展,以及临床诊疗指南定期迭代更新,以期逐步形成CRC CT影像人工智能领域数据库搭建的广泛共识。

**执笔者:**李昇霖(四川省人民医院);袁隆(兰州大学第二医院);卢婷(兰州大学第二医院)

**专家组成员**(按姓氏汉语拼音排序):艾涛(华中科技大学同济医学院附属同济医院);查云飞(武汉大学人民医院);陈峰(海南省人民医院);陈敏(北京医院);崔艳芬(山西省肿瘤医院);程英升(上海交通大学附属第六人民医院);程召平(山东省千佛山医院);程杰军(同济大学附属上海市第一妇婴保健院);董迪(中国科学院自动化研究所);冯湛(浙江大学医学院附属第一医院);范国光(中国医科大学附属第一医院);郭钊(空军军医大学第一附属医院);顾敏(重庆市人民医院);韩丹(昆明医科大学第一附属医院);侯阳(中国医科大学附属盛京医院);洪楠(北京大学人民医院);胡春洪(苏州大学附属第一医院);黄刚(甘肃省人民医院);何晓静(重庆医科大学附属第二医院);居胜红(东南大学附属中大医院);江新青(广州市第一人民医院);姜慧杰(哈尔滨医科大学附属第二医院);金征宇(北京协和医院);焦蕴(东南大学附属中大医院);柯晓艾(兰州大学第二医院);李东(天津医科大学总医院);李绍林(中山大学附属第五医院);梁栋(中国科学院深圳先进技术研究院);梁艺(广州中医药大学第一附属医院);廖伟华(中南大学湘雅医院);刘士远(海军军医大学第二附属医院);刘爱莲(大连医科大学附属第一医院);刘再毅(广东省人民医院);刘凯(海军军医大学第二附属医院);刘文亚(新疆医科大学第一附属医院);卢洁(首都医科大学北京宣武医院);卢再鸣(中国医科大学附属盛京医院);罗良平(暨南大学附属第一医院);罗天友(重庆医科大学附属第一医院);马静(新疆生产建设兵团医院);马林(解放军总医院第一医学中心);彭卫军(复旦大学附属肿瘤医院);蒲红(四川省人民医院);时高峰(河北医科大学第四医院);孙应实(北京大学肿瘤医院);田捷(中国科学院自动化研究所分子影像重点实验室/北京航空航天大学);陶晓峰(上海市第九人民医院);王健(陆军军医大学西南医院);王青(山东大学齐鲁医院);王光彬(山东

省医学影像学研究所);王锡明(山东省立医院);王梅云(河南省人民医院);王希明(苏州大学附属第一医院);王勇(河北医科大学第一医院);汪建华(宁波大学医学院附属医院);鲜军舫(首都医科大学附属北京同仁医院);徐海波(武汉大学中南医院);许茂盛(浙江中医药大学附属第一医院);严福华(上海交通大学医学院附属瑞金医院);杨丽(河北医科大学第四医院);杨华(重庆市中医院);于春水(天津医科大学总医院);于红(上海交通大学附属胸科医院);于兵(中国医科大学附属盛京医院);叶兆祥(天津医科大学肿瘤医院);叶枫(中国医学科学院肿瘤医院)、银武(西藏自治区人民医院);余永强(安徽医科大学第一附属医院);曾献军(南昌大学第一附属医院);曾自三(广西医科大学第一附属医院);张辉(山西医科大学第一医院);张敬(天津医科大学总医院);张明(西安交通大学第一附属医院);张勇(郑州大学第一附属医院);张龙江(东部战区总医院);张敏鸣(浙江大学医学院附属第二医院);张晓琴(内蒙古自治区人民医院);张雪宁(天津医科大学第二医院);张永海(青海省人民医院);张苗(首都医科大学宣武医院);张磊(吉林大学第一医院);张杨(山东大学齐鲁医院);张亚琴(中山大学附属第五医院);张惠茅(吉林大学第一医院);郑传胜(华中科技大学同济医学院附属协和医院);郑敏文(空军军医大学西京医院);郑海荣(中国科学院深圳先进技术研究院);钟红珊(中国医科大学附属第一医院);周俊林(兰州大学第二医院);周石(贵州医科大学附属医院);周福庆(南昌大学第一附属医院);周嶙(重庆医科大学附属第一医院);朱力(宁夏医科大学总医院);朱文珍(华中科技大学同济医学院附属同济医院);邹建勋(浙江省丽水市人民医院);赵国桦(郑州大学第一附属医院)

利益冲突 所有作者声明无利益冲突

### 参考文献

- [1] Eng C, Yoshino T, Ruiz-Garcia E, et al. Colorectal cancer[J]. *Lancet*, 2024, 404(10449): 294-310. DOI: 10.1016/S0140-6736(24)00360-X.
- [2] 郑荣寿, 张思维, 孙可欣, 等. 2016年中国恶性肿瘤流行情况分析[J]. *中华肿瘤杂志*, 2023, 45(3): 212-220. DOI: 10.3760/cma.j.cn112152-20220922-00647.
- [3] Qi J, Li M, Wang L, et al. National and subnational trends in cancer burden in China, 2005-20: an analysis of national mortality surveillance data[J]. *Lancet Public Health*, 2023, 8(12): e943-e955. DOI: 10.1016/S2468-2667(23)00211-6.
- [4] Benson AB, Venook AP, Al-Hawary MM, et al. Rectal cancer, version 2.2022, NCCN clinical practice guidelines in oncology[J]. *J Natl Compr Canc Netw*, 2022, 20(10): 1139-1167. DOI: 10.6004/jnccn.2022.0051.
- [5] 中华医学会放射学分会医学影像大数据与人工智能工作委员会, 中华医学会放射学分会腹部学组, 中华医学会放射学分会磁共振学组. 结直肠癌CT和MRI标注专家共识(2020)[J]. *中华放射学杂志*, 2021, 55(2): 111-116. DOI: 10.3760/cma.j.cn112149-20200706-00894.
- [6] Yao L, Li S, Tao Q, et al. Deep learning for colorectal cancer detection in contrast-enhanced CT without bowel preparation: a retrospective, multicentre study[J]. *EBioMedicine*, 2024, 104: 105183. DOI: 10.1016/j.ebiom.2024.105183.
- [7] Cicalini I, Chiarelli AM, Chiacchiaretta P, et al. Multi-omics staging of locally advanced rectal cancer predicts treatment response: a pilot study[J]. *Radiol Med*, 2024, 129(5): 712-726. DOI: 10.1007/s11547-024-01811-0.
- [8] Yang P, Qiu H, Yang X, et al. SAGL: a self-attention-based graph learning framework for predicting survival of colorectal cancer patients[J]. *Comput Methods Programs Biomed*, 2024, 249: 108159. DOI: 10.1016/j.cmpb.2024.108159.
- [9] Tong T, Li M. Abdominal or pelvic enhanced CT images within 10 days before surgery of 230 patients with stage II colorectal cancer[EB/OL]. <https://www.cancerimagingarchive.net/collection/stageii-colorectal-ct/>. DOI: 10.7937/p5k5-tg43.
- [10] Simpson AL, Peoples J, Creasy JM, et al. Preoperative CT and survival data for patients undergoing resection of colorectal liver metastases[J]. *Sci Data*, 2024, 11(1): 172. DOI: 10.1038/s41597-024-02981-2.
- [11] 国家药品监督管理局. 医疗器械唯一标识基本要求: YY/T 1630—2018[S]. 北京: 中国标准出版社, 2019.
- [12] 中华医学会放射学分会医学影像人工智能工作组. 脑出血CT数据库构建及质量控制专家共识[J]. *中华放射学杂志*, 2024, 58(7): 707-712. DOI: 10.3760/cma.j.cn112149-20231112-00378.
- [13] Eresen A, Li Y, Yang J, et al. Preoperative assessment of lymph node metastasis in colon cancer patients using machine learning: a pilot study[J]. *Cancer Imaging*, 2020, 20(1): 30. DOI: 10.1186/s40644-020-00308-z.
- [14] Li Q, Yang D, Guan Z, et al. Extranodal extension at pretreatment MRI and the prognostic value for patients with rectal cancer[J]. *Radiology*, 2024, 310(3): e232605. DOI: 10.1148/radiol.232605.
- [15] 郝焯, 王浩, 李佳戈. 行业标准《人工智能医疗器械质量要求和评价第3部分: 数据标注通用要求》解析[J]. *协和医学杂志*, 2023, 14(6): 1185-1188. DOI: 10.12290/xhyxzz.2023-0463.
- [16] 王浩, 孟祥峰, 郝焯, 等. 行业标准《人工智能医疗器械质量要求和评价第2部分: 数据集通用要求》解析[J]. *协和医学杂志*, 2023, 14(6): 1180-1184. DOI: 10.12290/xhyxzz.2023-0463.

【收稿日期】2024-11-06 【修回日期】2024-12-07

(本文编辑 张春辉)