

· 人工智能影像学专题 ·

深度学习图像重建算法在低管电压双期 CT 小肠造影中的应用研究

汪子雄, 唐友发, 罗彦, 胡学梅, 李震, 胡道予

【摘要】 目的:探讨深度学习图像重建(DLIR)算法在低管电压双期 CT 小肠造影(CTE)图像重建中的价值。方法:回顾性选取 40 例接受 100 kVp 双期 CTE 扫描的研究对象。原始数据使用滤波反投影(FBP)、新一代自适应统计迭代重建(ASiR-V, 30%和 60%)以及 DLIR(L、M、H)算法重建。对比分析不同重建算法在定量[CT 值、标准差(SD)、信噪比(SNR)、对比噪声比(CNR)]和定性(图像主观评分)质量方面的表现。同时,以 FBP 算法的 CNR 为基础,评估其他五种不同算法的相对 CNR 增幅与 BMI 之间的相关性。结果:在图像质量定量分析中,六种重建算法在四个解剖部位的两个期相中 CT 值差异均无统计学意义(P 均 >0.05),DLIR-H 算法优于 FBP 及 ASiR-V,达到最低的 SD、最高的 SNR 及 CNR(P 均 <0.05)。图像定性质量评估结果显示,DLIR-H 及 DLIR-M 算法优于其它算法($P < 0.05$)。DLIR-M 和 DLIR-H 的相对 CNR 增幅与 BMI 呈显著正相关(动脉期和静脉期相关系数 r 分别为 0.319、0.387 和 0.314、0.333, P 值均 <0.05)。结论:DLIR 算法提供了一种优化低剂量 CT 图像质量的有效工具,显著改善了图像质量,降低了噪声,特别适用于 BMI 较高的患者,具有广泛的临床应用潜力。

【关键词】 深度学习;图像重建;CT 小肠造影;低剂量;体层摄影术,X 线计算机

【中图分类号】 R574.5;R814.42 **【文献标志码】** A **【文章编号】** 1000-0313(2025)09-1075-07

DOI:10.13609/j.cnki.1000-0313.2025.09.001

开放科学(资源服务)标识码(OSID):



A study on the application of deep learning image reconstruction algorithm in low-kVp dual-phase CT enterography WANG Zi-xiong, TANG You-fa, LUO Yan, et al. Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China

【Abstract】 Objective: To investigate the performance of deep learning image reconstruction (DLIR) in enhancing image quality for low-kVp dual-phase CT enterography (CTE). **Methods:** This retrospective analysis included 40 individuals who underwent dual-phase CTE at 100kVp. Image datasets were reconstructed using filtered back projection (FBP), adaptive statistical iterative reconstruction-Veo (ASiR-V) at 30% and 60% blending, and DLIR at low, medium, and high levels (DLIR-L, DLIR-M, DLIR-H). Performance of different reconstruction algorithms were compared and analyzed in terms of objective measurements including CT attenuation, image noise (standard deviation, SD), signal-to-noise ratio (SNR), and contrast-to-noise ratio (CNR) and subjective image quality scoring. Meanwhile, the relationship between relative CNR enhancement over FBP and BMI was also examined among other five different algorithms. **Results:** In the quantitative analysis of image quality, there was no statistically significant difference in CT values among the six reconstruction algorithms in the dual-phase of the four anatomical sites (all $P > 0.05$). DLIR-H demonstrated significantly lower SD and higher SNR and CNR than FBP and ASiR-V (all $P < 0.05$). Subjective image quality scores were significantly higher for DLIR-H and DLIR-M ($P < 0.05$). Moreover, the relative CNR improvements achieved with DLIR-M and DLIR-H were positively correlated with BMI in both the arterial and venous phases ($r = 0.319, 0.387$ and $0.314, 0.333$, respectively; all $P < 0.05$). **Conclusion:** DLIR substantially enhances image quality in low-voltage dual-phase CTE by minimizing noise. It exhibits particular value in patients with elevated BMI, supporting its integration into clinical imaging protocols.

作者单位:430030 武汉,华中科技大学同济医学院附属同济医院放射科

作者简介:汪子雄(1997-),男,云南文山人,硕士研究生,住院医师,主要从事腹部影像诊断研究工作。

通讯作者:胡学梅, E-mail: mayjuly3720@163.com

基金项目:国家重点研发计划《基于国产 CT 的超低剂量儿科应用解决方案研究》(项目编号:2024YFC2419300)

【Key words】 Deep learning; Image reconstruction; CT enterography; Low dose; Tomography, X-ray computed

双期 CT 小肠造影(computed tomography enterography, CTE)因其非侵入性、快速性和高准确性优势,在炎症性肠病(inflammatory bowel disease, IBD)及肠道肿瘤的筛查和管理中发挥着重要作用^[1,2]。低管电压(kVp)扫描已被广泛应用,能增强图像对比度并减少对比剂使用量,同时降低辐射剂量,有助于减少潜在危害^[3]。然而,低管电压技术的应用仍面临背景噪声增加和图像质量下降的挑战,尤其在高体质量指数(body mass index, BMI)患者中。因此,迫切需要引入更先进的重建技术来优化图像质量^[4]。传统的滤波反投影(filtered back projection, FBP)技术在低剂量 CT 中受到限制,而自适应统计迭代重建(adaptive statistical iterative reconstruction, ASiR)和新一代自适应统计迭代重建-Veo(adaptive statistical iterative reconstruction-Veo, ASiR-V)已被证实在降噪和图像质量方面优于 FBP^[5,6]。然而,ASiR-V 在低对比病变的空间分辨率和高迭代强度下的“塑料感”外观限制了其进一步降低剂量的潜力^[7]。因此,深度学习图像重建(deep learning image reconstruction, DLIR)作为一种新兴算法,基于大规模数据训练,能够有效降低图像噪声并保留信号,从而提升低剂量 CT 图像质量^[8]。已有研究表明,DLIR 在腹部实体器官(如肝脏)中具有显著价值,具体表现为在降低辐射剂量的同时改善图像质量、降低噪声且保持噪声纹理,并保留肝转移瘤的良好检测性^[9-13]。然而,目前鲜有关于 DLIR 在空腔器官(如小肠与大肠)中的研究。本研究旨在评估 DLIR 算法在低管电压双期小肠 CT 造影中的效果。通过与 FBP 和 ASiR-V 算法的对比分析,探讨不同重建算法对 CT 图像定量与定性质量的影响。此外,本研究还深入探讨了 DLIR 算法在不同肠段及不同 BMI 人群中的表现,以评估其临床应用潜力。

材料与方法

1. 研究对象

纳入 2024 年 3 月至 2024 年 5 月应用 GE Revolution 256 排 CT 扫描仪行 100 kVp 双期 CTE 的患者。病例排除标准:①年龄 < 18 岁(2 例);②图像运动伪影严重(3 例);③相关临床数据缺乏(2 例);④存在明确的肠道(包括肠系膜及系膜血管)病变(肠道肿瘤 12 例,炎症 3 例)。最终 40 例患者被纳入本研究,患者纳入排除流程图见图 1。本回顾性研究经华中科技大学同济医学院附属同济医院伦理委员会批准。

2. CT 扫描与图像后处理

所有研究对象均采用 256 排 CT 扫描仪(Revolution, GE Health Care, USA)进行检查。检查前禁食 4 h,并于检查前 40 min 内饮水 1300 mL,于检查前立即再饮用 200 mL 以扩张胃肠道。扫描范围从膈顶至坐骨结节。采用智能增强追踪技术,在膈顶水平降主动脉放置感兴趣区(region of interest, ROI),当 CT 值达到 180 HU 后延迟 8 s 启动动脉期扫描,静脉期在动脉期开始 18 s 后启动。使用双筒高压注射器经静脉留置针注射对比剂(优维显, Ultravist 350 mg I/mL, 拜耳公司)。对比剂总碘量按研究对象体表面积(body surface area, BSA)调整($14.2 \text{ g I/m}^2 \text{ BSA}$), BSA 计算公式为 $\text{BSA} = 0.1173 \times \text{W}^{0.6466}$ (W 为体重,单位 kg)计算。对比剂注射时间固定为 20 s,随后以相同流率注射 20 mL 生理盐水。扫描参数:探测器配置 $256 \times 0.625 \text{ mm}$,光束准直 80 mm,螺距 0.992:1,旋转时间 0.6 s,管电压 100 kVp,自动管电流调制模式,扫描厚度 5 mm。

每例患者的原始数据分别采用 FBP、ASiR-V 和 DLIR 算法重建。其中 ASiR-V 使用两个权重(30%、60%),DLIR 使用三种强度(DLIR-L、DLIR-M、DLIR-H)。重建层厚为 1.25 mm。

3. 图像定量质量评估

图像定量质量评估由 1 位具有 5 年影像诊断经验的阅片者在 GE HealthCare 专用工作站(AW 4.7 版本)上完成。首先在 ASiR-V60% 图像上手动绘制圆形 ROI,选定部位包括空肠、回肠、右半结肠和左半结肠的肠壁,以及竖脊肌和皮下脂肪组织。随后将 ROI 自动复制至其他重建图像,确保不同重建方法下 ROI 位置一致。肠道选取适度扩张肠段,肠壁 ROI 直径为

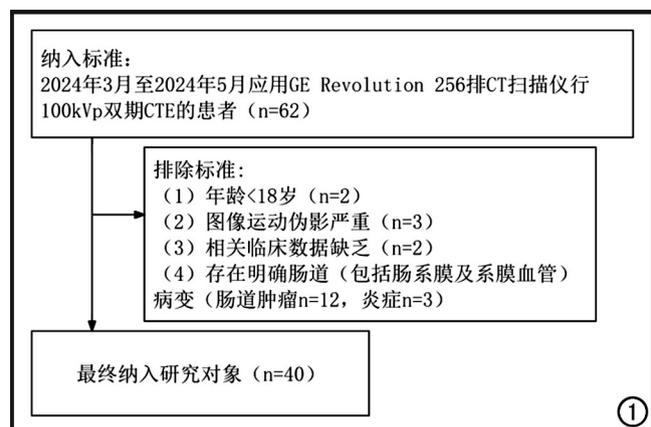


图 1 患者纳入排除流程图。

2 mm,避开肠腔内容物、肠系膜脂肪及血管;竖脊肌与皮下脂肪的 ROI 直径为 5 mm,位置为第 3 腰椎水平。所有 ROI 均由 1 位具有 15 年腹部影像诊断经验的高年资放射科医师复核。

记录每个 ROI 的 CT 值平均值及标准差(standard deviation, SD)。信噪比(signal-to-noise ratio, SNR)计算公式为: $SNR_{\text{目标}} = \text{平均值}_{\text{目标}} / \text{标准差}_{\text{目标}}$;对比噪声比(contrast-to-noise ratio, CNR)计算公式为: $CNR_{\text{目标}} = (\text{平均值}_{\text{目标}} - \text{平均值}_{\text{皮下脂肪}}) / \text{标准差}_{\text{皮下脂肪}}$ 。计算 ASiR-V 和 DLIR 相对于 FBP 的相对 CNR 增幅:相对 CNR 增幅目标 = $(CNR_{\text{目标}} - CNR_{\text{FBP}}) / CNR_{\text{FBP}}$ 。最后计算各重建方法在空肠、回肠、右半及左半结肠的平均相对 CNR 增幅,以分析其与 BMI 的相关性。

4. 图像定性质量评估

图像定性质量由 3 位腹部影像阅片者独立评估,阅片者 1(研究生)、阅片者 2(主治医师)、阅片者 3(主治医师)的影像诊断经验分别为 2 年、8 年、10 年。所有图像随机排序后分发给阅片者,动脉期与门静脉期图像同时评分。阅片者对图像重建方式及临床信息均不知晓。

图像总体质量(包括噪声、清晰度、肠壁与肠腔及肠系膜脂肪之间的对比)按 5 分制进行评分:1 分,噪声重,对比与清晰度极差,图像质量极差;2 分,噪声明显,对比与清晰度差,图像质量差;3 分,噪声中等,对比与清晰度尚可,图像质量中等;4 分,噪声不显著,对比与清晰度良好,图像质量好;5 分,噪声轻微,对比与清晰度优秀,图像质量优秀。

5. 统计学分析

所有数据均采用 Python(版本 3.9)进行统计学分析。首先通过 Shapiro-Wilk 检验判断数据正态性,采用 Levene 检验评估数据方差齐性。多组间连续变量比较时,根据正态性与方差齐性结果,选择单因素方差分析(ANOVA)或 Kruskal-Wallis 检验进行组间差异分析。两组间连续变量比较时,根据正态性与方差齐性结果,选择独立样本 *t* 检验、Welch 校正 *t* 检验或非

参数检验中的 Mann-Whitney *U* 检验。分类变量间的差异通过卡方检验进行评估。相对 CNR 增幅与 BMI 之间的相关性采用 Pearson 相关系数进行分析。检验水准设定为 $\alpha = 0.05$,以 $P < 0.05$ 为差异具有统计学意义。

结果

1. 研究对象特征

本研究共纳入 40 名研究对象,其中男 21 例,女 19 例,平均年龄(52.5 ± 12.7)岁(范围 18~73 岁),平均 BMI 为(22.48 ± 2.85) kg/m^2 (范围 15.56~27.68 kg/m^2)。所有纳入研究对象的 CTE 均经 2 位高年资放射科医师阅片并结合临床资料(包括病史、实验室、内镜及病理检查结果,既往及随访的影像资料)排除了明确的肠道(包括肠系膜及系膜血管)病变。

2. 辐射剂量与对比剂剂量

动脉期与静脉期图像的容积 CT 剂量指数(volume CT dose index, CTDIvol)分别为(7.83 ± 0.81) mGy 与(7.83 ± 0.81) mGy,剂量长度乘积(dose length product, DLP)分别为(414.33 ± 56.66) mGy·cm 与(414.32 ± 56.63) mGy·cm,有效剂量(effective dose, ED)分别为(6.22 ± 0.85) mSv 与(6.21 ± 0.85) mSv。平均对比剂使用量为(68 ± 7) mL,注射流率为(3.4 ± 0.3) mL/s。

3. 图像质量定量和定性分析

在图像质量定量分析中,六种重建算法在四个解剖部位的两个期相中 CT 值差异均无统计学意义(表 1)。在 SD 方面,FBP 最高,其次为 ASiR-V30%、DLIR-L、ASiR-V60%与 DLIR-M 接近,最低为 DLIR-H(表 2,图 2a、b)。CNR 方面,最低为 FBP,ASiR-V30%、DLIR-L、ASiR-V60%、DLIR-M 与 DLIR-H 依次递增(表 3,图 2c、d)。SNR 方面,除 ASiR-V60%与 DLIR-M 间差异无统计学意义外,其它趋势与 CNR 一致(表 4,图 2e、f)。

在图像质量定性分析中,尽管 3 位阅片者的具体评估细节存在差异,但各算法间质量差异的趋势保持

表 1 六种重建算法在双期 CTE 中小肠结肠 CT 值结果 (HU)

部位	FBP	ASiR-V30%	ASiR-V60%	DLIR-L	DLIR-M	DLIR-H	统计量	P 值
动脉期								
空肠	84.10±19.51	84.01±19.30	83.81±19.17	84.09±18.99	83.95±18.86	83.80±18.69	$\chi^2 = 0.033$	1.000
回肠	58.64±12.51	58.32±12.53	57.78±12.53	58.67±12.19	58.56±12.30	58.31±12.27	$F = 0.029$	1.000
右半结肠	49.88±14.14	49.63±13.81	49.36±13.42	49.96±13.28	49.79±13.08	49.51±12.95	$\chi^2 = 0.111$	1.000
左半结肠	51.67±13.23	51.22±12.96	50.78±12.69	51.42±12.55	51.19±12.31	50.84±12.13	$F = 0.028$	1.000
静脉期								
空肠	95.97±11.39	95.85±11.27	95.65±11.23	95.82±11.17	95.81±11.09	95.63±11.07	$\chi^2 = 0.047$	1.000
回肠	82.17±15.07	81.73±15.04	81.30±14.95	81.90±14.82	81.57±14.77	81.28±14.66	$\chi^2 = 0.115$	1.000
右半结肠	65.90±12.82	65.15±12.48	64.40±12.22	65.34±11.39	64.81±11.05	64.13±10.73	$F = 0.120$	1.000
左半结肠	68.97±15.95	68.69±15.28	68.40±14.72	69.34±14.05	69.21±13.55	69.04±13.06	$F = 0.023$	1.000

表 2 六种重建算法在双期 CTE 中小肠结肠 SD 结果 (HU)

部位	FBP	ASiR-V30%	ASiR-V60%	DLIR-L	DLIR-M	DLIR-H	统计量	P 值
动脉期								
空肠	26.58±4.24	20.08±3.21	13.57±2.37	16.98±2.78	13.50±2.20	9.77±1.64	$\chi^2=188.187$	<0.001*
回肠	24.70±4.26	18.87±3.45	13.03±2.93	16.02±3.11	12.88±2.65	9.53±2.34	$\chi^2=173.594$	<0.001*
右半结肠	23.68±3.91	18.26±3.11	12.89±2.54	15.66±3.11	12.89±2.82	10.01±2.59	$\chi^2=162.661$	<0.001*
左半结肠	24.34±4.22	18.67±3.39	13.01±2.68	16.64±3.19	13.63±2.76	10.46±2.43	$\chi^2=157.213$	<0.001*
皮下脂肪	18.91±2.81	14.50±2.21	10.14±1.63	10.86±1.66	8.48±1.33	6.01±1.07	$\chi^2=198.700$	<0.001*
静脉期								
空肠	25.05±5.12	18.69±3.87	12.24±2.70	16.12±3.30	12.71±2.58	9.18±1.75	$\chi^2=171.224$	<0.001*
回肠	24.12±5.09	18.07±3.78	11.94±2.56	15.78±3.66	12.62±2.87	9.20±2.14	$\chi^2=156.970$	<0.001*
右半结肠	22.98±5.04	17.36±3.95	11.74±3.02	15.32±3.67	12.51±3.10	9.59±2.64	$\chi^2=143.090$	<0.001*
左半结肠	22.45±5.48	17.31±4.42	12.20±3.57	14.92±3.78	12.24±3.16	9.32±2.60	$\chi^2=126.210$	<0.001*
皮下脂肪	19.10±2.78	14.72±2.23	10.39±1.73	11.07±1.94	8.59±1.63	6.07±1.39	$\chi^2=196.813$	<0.001*

注: * 差异有统计学意义

表 3 六种重建算法在双期 CTE 中小肠结肠 CNR 结果

部位	FBP	ASiR-V30%	ASiR-V60%	DLIR-L	DLIR-M	DLIR-H	统计量	P 值
动脉期								
空肠	10.72±2.38	13.98±3.11	20.01±4.45	18.76±4.38	24.01±5.54	34.07±8.15	$\chi^2=173.280$	<0.001*
回肠	9.37±2.33	12.20±3.04	17.42±4.19	16.36±4.00	20.91±4.85	29.57±6.56	$\chi^2=180.240$	<0.001*
右半结肠	8.88±2.23	11.57±2.86	16.52±3.84	15.49±3.68	19.80±4.44	27.99±5.96	$\chi^2=193.310$	<0.001*
左半结肠	8.98±2.18	11.69±2.80	16.68±3.78	15.65±3.65	19.98±4.38	28.25±6.00	$\chi^2=182.801$	<0.001*
静脉期								
空肠	11.21±2.17	14.55±2.85	20.68±4.31	19.50±4.28	25.21±5.60	36.00±8.65	$\chi^2=180.142$	<0.001*
回肠	10.47±2.14	13.56±2.79	19.24±4.11	18.17±4.03	23.44±5.19	33.42±7.76	$\chi^2=182.044$	<0.001*
右半结肠	9.57±1.65	12.37±2.20	17.53±3.37	16.58±3.26	21.37±4.27	30.42±6.62	$\chi^2=185.728$	<0.001*
左半结肠	9.74±1.89	12.63±2.47	17.93±3.70	16.97±3.60	21.92±4.69	31.32±7.30	$\chi^2=182.029$	<0.001*

注: * 差异有统计学意义

表 4 六种重建算法在双期 CTE 中小肠结肠 SNR 结果

部位	FBP	ASiR-V30%	ASiR-V60%	DLIR-L	DLIR-M	DLIR-H	统计量	P 值
动脉期								
空肠	3.23±0.83	4.26±1.08	6.31±1.58	5.06±1.30	6.34±1.57	8.74±2.11	$\chi^2=149.299$	<0.001*
回肠	2.45±0.74	3.20±0.96	4.63±1.40	3.81±1.15	4.74±1.40	6.43±1.87	$\chi^2=116.119$	<0.001*
右半结肠	2.17±0.70	2.79±0.87	3.94±1.16	3.30±1.01	4.03±1.24	5.23±1.70	$\chi^2=100.873$	<0.001*
左半结肠	2.18±0.61	2.82±0.79	4.04±1.13	3.17±0.81	3.87±0.99	5.06±1.38	$\chi^2=110.942$	<0.001*
静脉期								
空肠	3.99±0.92	5.34±1.25	8.18±1.96	6.19±1.44	7.83±1.76	10.76±2.32	$\chi^2=162.016$	<0.001*
回肠	3.64±1.47	4.83±1.99	7.30±3.01	5.63±2.57	6.99±3.05	9.56±4.02	$\chi^2=118.845$	<0.001*
右半结肠	2.97±0.75	3.89±0.98	5.74±1.52	4.47±1.25	5.45±1.54	7.09±2.01	$\chi^2=123.558$	<0.001*
左半结肠	3.26±1.11	4.22±1.43	6.04±2.14	4.93±1.53	6.02±1.88	7.94±2.56	$\chi^2=106.287$	<0.001*

注: * 差异有统计学意义

一致。统计学分析结果表明, ASiR-V60% 与 DLIR-L 之间以及 DLIR-M 与 DLIR-H 之间差异无统计学意义 ($P>0.05$), 而其余组别之间差异均有统计学意义 ($P<0.05$)。因此图像质量评分排序为: FBP<ASiR-V30%<ASiR-V60%=DLIR-L<DLIR-M=DLIR-H (图 3)。

4. 相对 CNR 增幅与 BMI 的相关性

在动脉期和静脉期中, DLIR-M 和 DLIR-H 算法的相对 CNR 增幅与 BMI 呈显著正相关 (DLIR-M 和 DLIR-H 的相关系数分别为 0.319、0.387 和 0.314、0.333, P 均 <0.05), 而 ASiR-V 算法的相关性均不显著 ($r=0.078\sim 0.173$, P 均 >0.05 , 图 4)。

讨论

本研究结果表明, DLIR-H 算法在 100 kVp 双期 CTE 图像重建中的表现优于其它几种算法。图像质量定量评估中 DLIR-H 达到最低的 SD、最高的 SNR 及 CNR (P 均 <0.05)。DLIR-M 和 DLIR-H 的相对 CNR 增幅与 BMI 呈显著正相关 (动脉期和静脉期相关系数 r 分别为 0.319、0.387 和 0.314、0.333, P 值均 <0.05)。图像定性质量评估结果显示, DLIR-H 及 DLIR-M 算法优于其它算法 ($P<0.05$)。

在本研究中, DLIR-H 相比于 FBP 及 ASiR-V 在双期 CTE 中显著提升图像质量, 与以往在肝脏及胰腺研究中的结论一致^[14,15]。在多个维度上, 包括 SNR、

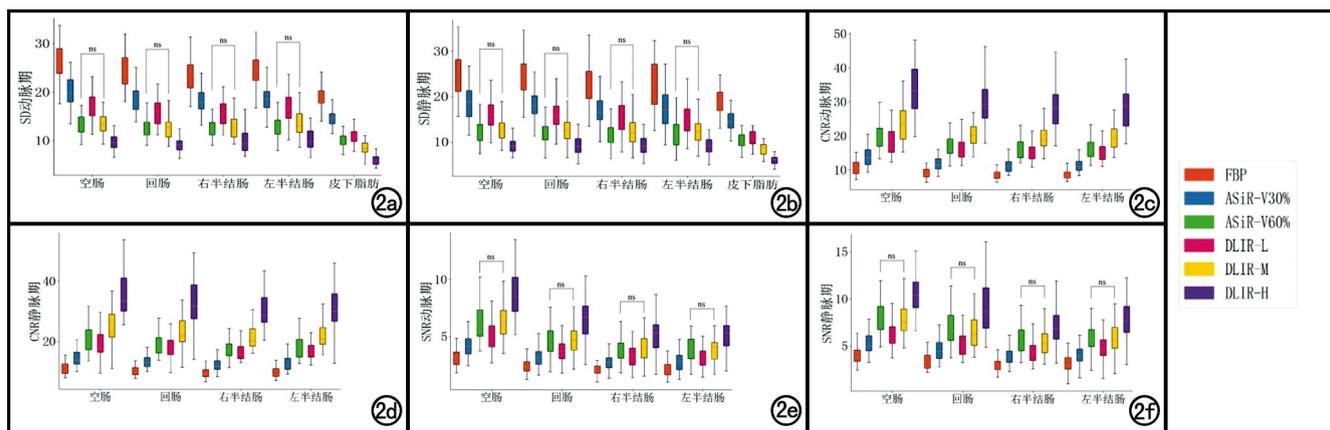


图2 图像定量质量评估。图中标记线及“ns”表示标记的算法组间差异无统计学意义($P > 0.05$),其余各组间差异均有统计学意义($P < 0.05$)。a) 动脉期 SD; b) 静脉期 SD; c) 动脉期 CNR; d) 静脉期 CNR; e) 动脉期 SNR; f) 静脉期 SNR。

CNR 以及图像的主观评分,均观察到显著的性能提升。特别是 DLIR-H 和 DLIR-M 两种算法,其 CNR 值均显著超越了 ASiR-V60% 算法,这表明 DLIR 算法在降低噪声的同时,有效提升了图像的对比度。DLIR-H 算法不仅在定量指标上实现了显著的性能提升,而且在图像质量的主观评价中,也被认为优于其他重建算法。这一现象可能归因于 DLIR 算法在减少噪声的同时,保持了图像的高质量,这对于复杂结构的肠道成像尤为重要。DLIR 算法为临床提供了一种有效的工具,有助于提高对低对比度病变的诊断准确性,尤其是在图像噪声较高的情况下,DLIR 算法的优势尤

为显著。

本研究引入“相对 CNR 增幅”作为评价 ASiR-V 与 DLIR 在不同 BMI 下提升图像质量效率的指标。结果显示,ASiR-V 在不同 BMI 研究对象中的表现差异无统计学意义,而 DLIR-M 及 DLIR-H 算法的图像质量提升与 BMI 呈正相关。在高 BMI 研究对象中,ASiR-V 受限于原始噪声过大导致图像质量下降,难以实现低 kV 扫描,需使用较高电压,从而增加辐射剂量。而 DLIR 从高质量训练数据中学习识别与消除噪声,因此其性能主要依赖于训练数据质量而非输入数据质量。对于原始噪声大的高 BMI 研究对象,DLIR

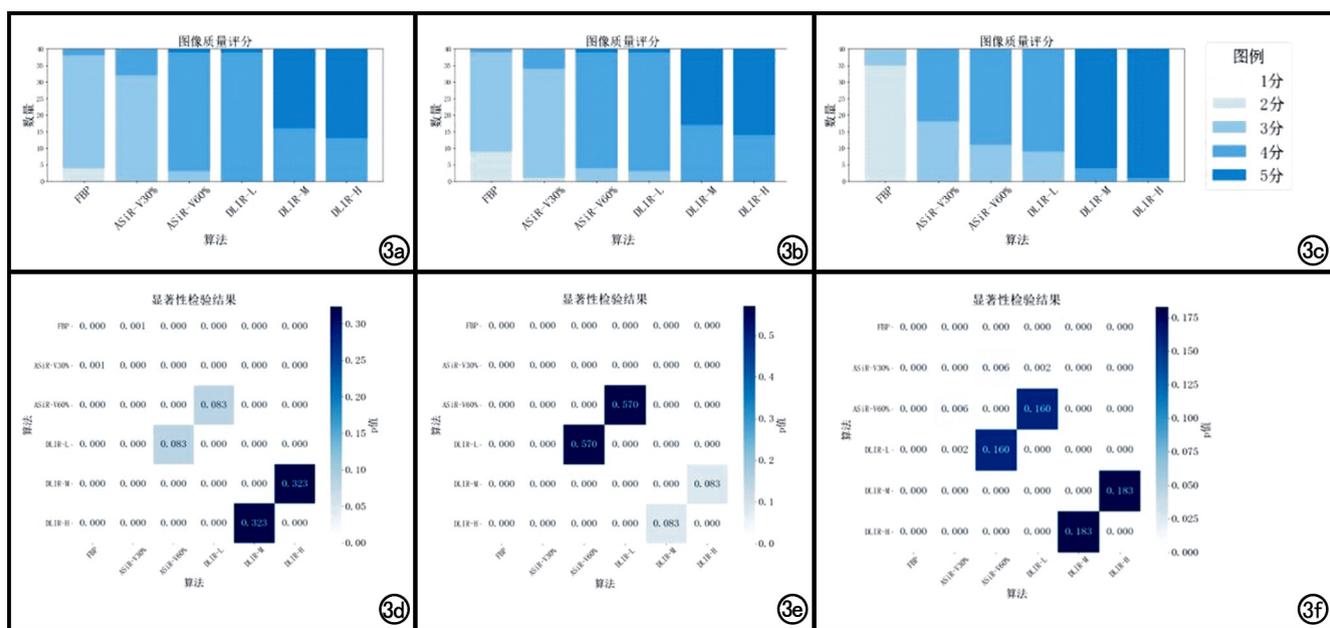


图3 图像定性质量评估。在3位阅片者的评分中,除 ASiR-V60% 与 DLIR-L 以及 DLIR-M 与 DLIR-H 两组外,其余组别之间差异均有统计学意义($P < 0.05$)。a) 阅片者 1 的图像质量柱状图; b) 阅片者 2 的图像质量柱状图; c) 阅片者 3 的图像质量柱状图; d) 阅片者 1 的显著性检验结果; e) 阅片者 2 的显著性检验结果; f) 阅片者 3 的显著性检验结果。

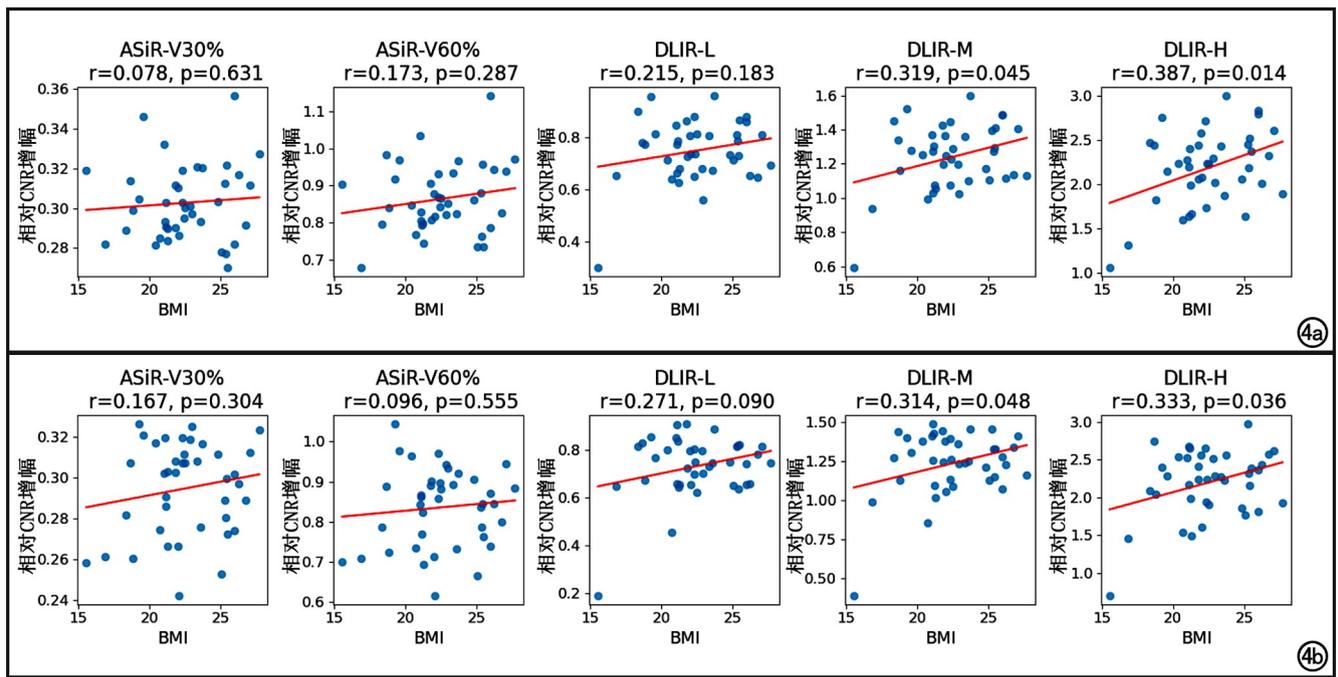


图4 相对 CNR 增幅与 BMI 的 Pearson 相关性分析结果。展示了 BMI 与各指标 (ASiR-V30%、ASiR-V60%、DLIR-L、DLIR-M、DLIR-H) 之间的关系。每个子图绘制了 BMI 与对应指标的散点图,红线为回归直线,反映二者间的线性趋势。标题中标注了 Pearson 相关系数(r)和对应的 P 值,以评估各指标与 BMI 之间的统计关联性。a) 动脉期; b) 静脉期。

在 CNR 改善上表现更为显著,使得低 kV 扫描在高 BMI 研究对象中亦能获得优质图像。

本研究中所有研究对象均使用 100 kVp 扫描,辐射剂量较低,动静脉期 CTDI_{vol} 分别为 (7.83 ± 0.81) mGy 和 (7.83 ± 0.81) mGy,仅为美国放射学会推荐腹部增强 CT 诊断参考水平 (19 mGy) 的 41.2%^[16],该剂量水平与已有研究中基于体型设定 kVp 的低剂量腹部 CT 及 CTE 相当^[17]。

临床实践中,低 kVp 可降低辐射剂量与对比剂使用量,然而为确保穿透能力,腹部常规仍使用 120 kVp,尤其是在肥胖患者中^[18]。本研究结果显示,对于 BMI 介于 15.56~27.68 kg/m² 的研究对象,即使使用 100 kVp,图像质量与诊断效能依然优良。提示在 DLIR 算法加持下,无需因 BMI 调整 kVp 设定,可维持低 kVp 所带来的对比增强优势。而在 BMI 偏低患者中,仍建议个体化 kVp 方案,以进一步降低辐射剂量与对比剂使用量。在急诊环境中,患者 BMI 难以即时获知,DLIR 的应用有助于提升工作流程效率并保障影像质量与安全性。

本研究排除了存在明确肠道(包括肠系膜及系膜血管)病变的人群,仅纳入经临床及影像诊断为“肠道未见明显异常”的“正常”人群作为研究对象,是考虑到炎症、肿瘤或其他病理状态可能会影响图像质量,导致噪声、对比度等因素的变化,从而不能纯粹地评估重建

算法的性能。同时考虑到肠道病变的存在可能会影响肠壁 ROI 的勾画。本研究通过选取经临床及影像诊断为“肠道正常”的人群作为研究对象,排除了疾病因素的干扰,能够更准确地评估各重建算法在低 kVp 双期 CTE 中的表现。后续研究将纳入肠道病变的患者,进一步探讨在疾病状态下 DLIR 在改善 CTE 图像质量以及对肠道病变诊断中的价值,如病变检出率,病变状态下肠壁黏膜、浆肌层等肠壁微结构细节的评估情况,以及对肠道病变的诊断符合率及把握度等。

值得注意的是,尽管 DLIR 具备诸多优点,但其模型复杂性与可解释性仍是限制因素。虽然本研究中未观察到,但 DLIR 仍有潜在“幻觉”(hallucination)与“反幻觉”(inverse hallucination)风险,可能引发误诊^[19]。因此,DLIR 未来的发展仍依赖大量高质量训练数据^[20]。

综上所述,在适用于广泛 BMI 人群的 100 kVp 双期 CTE 中,DLIR-H 算法显著改善了图像质量并降低了噪声,这展现出该算法具有良好性能并且仍有发展空间及潜力,值得进一步深入研究与验证。

参考文献:

- [1] Meng J, Luo Z, Chen Z, et al. Intestinal fibrosis classification in patients with Crohn's disease using CT enterography-based deep learning: comparisons with radiomics and radiologists[J]. Eur Radiol, 2022, 32(12): 8692-8705.
- [2] Bruining DH, Zimmermann EM, Loftus EV, et al. Consensus rec-

- ommendations for evaluation, interpretation, and utilization of computed tomography and magnetic resonance enterography in patients with small bowel Crohn's disease[J]. *Radiology*, 2018, 286(3):776-799.
- [3] Lee Y, Hwang J, Ryu H, et al. Image quality and diagnostic accuracy of reduced-dose computed tomography enterography with model-based iterative reconstruction in pediatric Crohn's disease patients[J]. *Sci Rep*, 2022, 12(1):2147.
- [4] Kapoor A, Mahajan G, Kapoor A. Evaluation of low-kVp low-volume iodinated contrast protocol for coronary CT angiography using retrospective ECG gating[J]. *Indian J Radiol Imaging*, 2022, 31(4):910-916.
- [5] Chen LH, Jin C, Li JY, et al. Image quality comparison of two adaptive statistical iterative reconstruction (ASiR, ASiR-V) algorithms and filtered back projection in routine liver CT[J]. *Br J Radiol*, 2018, 91(1088):20170655.
- [6] 付芮, 张成亮, 沈云, 等. 深度学习图像重建算法对图像质量的影响: 体模研究[J]. *放射学实践*, 2023, 38(12):1617-1622.
- [7] Deng D, Dong H. Advantages of contrast-enhanced CT combined with DCE-MRI in identifying malignant parotid tumor[J]. *Am J Transl Res*, 2022, 14(12):9047-9056.
- [8] Zhong J, Xia Y, Chen Y, et al. Deep learning image reconstruction algorithm reduces image noise while alters radiomics features in dual-energy CT in comparison with conventional iterative reconstruction algorithms: a phantom study[J]. *Eur Radiol*, 2022, 33(2):812-824.
- [9] Park J, Shin J, Min IK, et al. Image quality and lesion detectability of lower-dose abdominopelvic CT obtained using deep learning image reconstruction[J]. *Korean J Radiol*, 2022, 23(4):402-412.
- [10] Racine D, Becce F, Viry A, et al. Task-based characterization of a deep learning image reconstruction and comparison with filtered back-projection and a partial model-based iterative reconstruction in abdominal CT: A phantom study[J]. *Phys Med*, 2020, 76:28-37.
- [11] Lin X, Gao Y, Zhu C, et al. Improved overall image quality in low-dose dual-energy computed tomography enterography using deep-learning image reconstruction[J]. *Abdom Radiol (NY)*, 2024, 49(9):2979-2987.
- [12] Lee JM, Kang HJ, Park SJ, et al. Image quality improvement of low-dose abdominal CT using deep learning image reconstruction compared with the second generation iterative reconstruction[J]. *Curr Med Imaging*, 2023, 20:e250523217310.
- [13] 王诗耕, 刘义军, 童小雨, 等. 深度学习重建算法对肾上腺肿瘤的检出及鉴别效能的影响[J]. *放射学实践*, 2024, 39(8):1081-1088.
- [14] Wang H, Yue S, Liu N, et al. Deep learning reconstruction vs. standard reconstruction for abdominal CT: the influence of BMI[J]. *Eur Radiol*, 2023, 34(3):1614-1623.
- [15] Jensen C, Liu X, Tamm E, et al. Image quality assessment of abdominal CT by use of new deep learning image reconstruction: initial experience[J]. *Am J Roentgenol*, 2020, 215(1):1-8.
- [16] Jensen C, Gupta S, Saleh MM, et al. Reduced-dose deep learning reconstruction for abdominal CT of liver metastases[J]. *Radiology*, 2022, 303(1):211838.
- [17] Feng C, Zhu D, Zou X, et al. The combination of a reduction in contrast agent dose with low tube voltage and an adaptive statistical iterative reconstruction algorithm in CT enterography[J]. *Medicine (Baltimore)*, 2018, 97(12):e0151.
- [18] Levinson MI, Hoerner M, Mustafa A, et al. Abdominal CT imaging applications of low kilovoltage peak techniques[J]. *J Am Coll Radiol*, 2019, 16(4):482-484.
- [19] Wang S, Su Z, Ying L, et al. Accelerating magnetic resonance imaging via deep learning[J]. *Proc IEEE Int Symp Biomed Imaging*, 2016, 2016:514-517.
- [20] Koetzier L, Mastrodicasa D, Szczykutowicz T, et al. Deep learning image reconstruction for CT: technical principles and clinical prospects[J]. *Radiology*, 2023, 306(3):e221257.

(收稿日期:2025-06-09 修回日期:2025-09-02)

人工智能大语言模型在医学影像学中的应用进展

许莹, 赵心明

【摘要】 本文系统性综述了大语言模型(LLMs)在医学影像学领域的最新研究进展与应用前景。首先介绍了 LLMs 的基本概念、技术原理及发展历程,然后从影像诊断、疾病临床分期、报告生成、科研辅助、影像医学考试与教育、健康宣教与患者沟通等六个方面详细阐述了 LLMs 的应用现状。研究表明,LLMs 在影像诊断、结构化报告生成、临床决策支持、科研工作开展、影像医学教育、医患沟通等方面展现出显著优势,但在直接影像诊断方面仍存在局限性。最后,本文探讨了当前面临的挑战并进行展望,为 LLMs 在医学影像学领域的进一步研究和应用提供参考。

【关键词】 大语言模型;医学影像学;人工智能;应用进展

【中图分类号】 TP18;R81 **【文献标识码】** A **【文章编号】** 1000-0313(2025)09-1082-06

DOI:10.13609/j.cnki.1000-0313.2025.09.002

开放科学(资源服务)标识码(OSID):



近年来,人工智能(artificial intelligence, AI)技术的迅猛发展正在深刻改变医学影像学的实践模式和研究范式。作为 AI 领域的重要突破,大语言模型(large language models, LLMs)凭借其强大的自然语言理解与生成能力,在医疗领域展现出巨大潜力。医学影像学作为临床诊疗的核心环节,涉及海量结构化与非结构化数据的整合与分析,而 LLMs 的引入有望优化临床诊疗流程并提升科研效率。目前,LLMs 在医学影像学中的应用主要集中在以下几个方向:①影像诊断及鉴别诊断;②疾病临床分期及分类;③影像报告生成、结构化及质量控制;④科研数据提取及分析;⑤影像学考试及医学生教育;⑥健康宣教与医患沟通。本文系统综述了 LLMs 在医学影像学领域的最新研究进展,从技术原理、应用场景及现存挑战三个维度展开探讨,旨在为未来研究提供理论参考和实践指导。

大语言模型概述

LLMs 作为当前人工智能领域最具突破性的技术之一,凭借其卓越的跨领域文本生成与理解能力,正在深刻改变多个学科领域。这类基于深度学习技术的自然语言处理模型,通过在海量文本数据上进行预训练,捕捉语言的统计规律和深层语义表征,从而具备文本生成、智能问答、多语言翻译、内容摘要以及逻辑推理等多样化能力^[1]。其核心技术架构源于 2017 年谷歌研究院提出的 Transformer 模型,该架构采用创新的

自注意力机制,有效解决了传统循环神经网络在处理长距离依赖关系时的局限性,为后续大规模预训练模型的发展奠定了理论基础^[2,3]。从技术演进的角度来看,LLMs 的发展经历了三个重要阶段:早期的统计语言模型阶段、中期的神经网络模型阶段,以及当前基于 Transformer 的大规模预训练模型阶段^[4]。2018 年,OpenAI 发布的生成式预训练转换器(generative pre-trained transformer, GPT)-1 首次采用单向 Transformer 解码器结构,开创了生成式预训练的新模式;2020 年问世的 GPT-3 凭借 1750 亿参数和强大的少样本学习能力,将 LLMs 推向了通用人工智能的新高度;而随后 ChatGPT(基于 GPT-3.5/4 架构)的广泛应用,则标志着这项技术开始真正走向产业化落地^[2]。

值得注意的是,2023 年 3 月发布的 GPT-4 最初仅支持纯文本处理,这在很大程度上限制了其在医疗领域的应用潜力。为此,OpenAI 于同年 10 月推出了多模态版本 GPT-4V(vision),该版本能够同时处理视觉和文本输入,大大拓展了模型在医学影像分析中的应用场景^[5]。多模态大语言模型(multimodal large language models, MLLMs)的核心优势在于其能够模拟人类,通过整合多种感官信息(文本、图像、音频、视频等)来理解和处理复杂问题^[6]。本文主要简述通用 LLMs 2023 年之后的发展,虽然本文主要聚焦于单模态 LLMs 在医学文本处理中的应用,但也会适当探讨 MLLMs 在影像领域的创新实践。

大语言模型在医学影像学中的应用进展

1. 影像诊断及鉴别诊断

近年来,越来越多的研究开始探索 LLMs 在各种急危重症和肿瘤性病变诊断中的临床应用价值。以腹

作者单位:100021 北京,国家癌症中心 国家肿瘤临床医学研究中心 中国医学科学院北京协和医学院肿瘤医院 影像诊断科

作者简介:许莹(1996—),女,河南郑州人,博士,住院医师,主要从事消化系统恶性肿瘤的影像与转化研究工作。

通讯作者:赵心明, E-mail: xinmingzh@sina.com

基金项目:中央高水平医院临床科研业务费资助(80102022505)

部创伤为例,肝、脾、肾是最常见的损伤器官,约占全部创伤病例的 80%^[7]。在急诊临床实践中,医生通常需要综合患者的病史、体格检查和影像学检查结果才能做出准确诊断^[8]。如何充分利用现有的医疗信息辅助临床决策,成为提升急诊救治效率的关键。Wang 等^[9]开发了一套基于平扫 CT 图像和非结构化文本数据的腹部创伤多模态诊断系统(screening for multi-organ assessment in rapid trauma, SMART); 研究结果显示,基于非结构化文本数据的 SMART_GPT 模块灵敏度为 81.3%,受试者工作特征曲线下面积(area under curve, AUC)为 0.88; 而基于 CT 图像的 SMART_Image 模块则表现出更高的灵敏度(87.5%),但 AUC 略低(0.81)。在实际急诊场景模拟中,该系统将患者平均等待时间缩短了 64.24%,显著提高了救治效率。在急性肝衰竭(acute liver failure, ALF)诊断方面,一项比较研究评估了五种主流 LLMs(ChatGPT 3.5/4、BARD、CLAUDE 和 COPILOT)的临床表现,并将这些结果与使用检索增强生成(retrieval augmented generation, RAG)技术的加强版 ChatGPT4 进行了比较。结果显示增强版 ChatGPT4 在诊断准确性(3.67 分)、回答清晰度(4.04 分)和临床相关性(4.01 分)三个维度均显著优于其他模型^[10]。Horiuchi 等^[11]针对骨骼肌肉系统影像开展了一项有趣的研究,比较了 GPT-4、GPT-4V 和不同年资影像科医生的诊断水平。研究发现,纯文本版本的 GPT-4 诊断准确率达到 43%(46/106),显著高于视觉版本 GPT-4V 的 8%(9/106),其表现与影像科住院医师相当。类似地,Mitsuyama 等^[12]在脑肿瘤诊断研究中发现,GPT-4 的诊断能力与神经影像亚专业医师相当。此外,有研究证实 GPT-4 根据影像特征提供鉴别诊断的性能与专家小组的一致性为 68.8%,且 GPT-4 提出的鉴别诊断有 93.8%被认为是可接受的替代方案^[13]。

虽然 LLMs 基于文本输入的诊断准确性较高,但其直接基于图像进行影像诊断的准确性尚需提高。Brin 等^[14]评估了 GPT-4V 对 230 张急诊影像(包括超声、CT 和 X 线)的分析能力。虽然模型在识别成像方式(100%准确率)和解剖定位(87.1%准确率)方面表现优异,但其整体诊断准确率仅为 35.2%,其中 CT、X 线和超声的诊断准确率分别为 36.4%、66.7% 和 9.1%。Zhang 等^[15]也发现,GPT-4 在肝脏疾病 CT 诊断中的异常检出准确率为 54%,但疾病诊断准确率仅为 23%,远低于专业 AI 影像诊断系统。在一项早期研究中,一个同时使用临床参数和 X 线胸片的基于 Transformer 的模型比仅使用临床参数或影像数据的模型表现更好^[16]。这些研究结果提示,当前 LLMs 尚不能作为独立的影像诊断工具,但其与临床参数结合

使用时可能展现出更好的应用前景。

2. 疾病临床分期及分类

肝癌临床分期是评估疾病严重程度和制定治疗方案的重要依据。目前国内主要采用中国肝癌分期方案,临床医生通常需要根据影像学报告内容进行人工评估。这一过程不仅耗时耗力,还存在主观判断差异。Li 等^[17]开发了一种基于 LLMs 的知识感知注意力网络,能够从自由文本的影像学报告中自动提取关键信息并完成肝癌临床分期。该模型在测试中展现出优异的性能,准确率达到 90.3%,宏观 F1 得分和召回率均为 90.0%,为临床医生提供了可靠的决策支持。在肝脏影像报告和数据系统(liver imaging reporting & data system, LI-RADS)分类方面,有研究团队基于 GPT-4 架构开发了专用模型 LiverAI^[18]。该模型能够自动分析 MRI 报告中的自由文本描述,并给出符合 LI-RADS 标准的分类结果,其表现与影像科医生评估结果无统计学差异。实际应用数据显示,该模型可将数据管理效率提高 60%,使影像科医生工作量减少 45%。Gu 等^[19]的独立验证研究也证实,GPT-4 在 LI-RADS 特征提取和分类任务中具有较强能力,准确率达到 85%。

然而,LLMs 在肿瘤分期方面的表现存在较大差异。一项针对日本胰腺癌患者的研究评估了 GPT-4 从日文影像报告中识别 TNM 分期的能力^[20]。结果显示,模型对 T、N、M 分期的准确率分别为 73%、91%和 93%,Kappa 系数分别为 0.45、0.79 和 0.83,表明其对 T 分期的判断与专家共识的一致性仅为中等水平。后续研究采用 Claude 3.5 Sonnet 模型对相同数据集进行评估^[21],发现其对 T 分期的准确率提升至 80%,Kappa 系数提高至 0.65,显示出更好的性能,但对于 N、M 分期差异无统计学意义。Wu 等^[22]评估三种主流 LLMs(Claude-2、GPT-3.5 和 GPT-4)对 LI-RADS 2018 版、肺部 CT 筛查报告和数据系统(lung CT screening reporting & data system, Lung-RADS) 2022 版、卵巢及附件报告和数据系统(ovarian-adnexal reporting & data system, O-RADS)磁共振成像进行 RADS 类别的判定能力;结果显示,所有模型在 LI-RADS 分类任务中的表现均优于其它系统,其中 Claude-2 在小样本提示下的最佳准确率为 57%。以上研究提示虽然目前 LLMs 在疾病临床分期、分类中取得一定成果,但距离在临床实例中进行应用实践尚有一定距离。

3. 影像报告

LLMs 在影像报告领域的应用主要集中在三个方面:自动报告生成、自由文本结构化转换和报告质量控制。

在报告生成方面, Huang 等^[23] 比较了 LLMs 生成报告与影像科医生报告的质量差异。研究随机抽取 500 例急诊 X 线检查图像, 采用李克特量表评估显示, LLMs 生成的报告在临床准确性和文本质量方面与影像科医生报告相当, 且优于远程影像报告。然而, 另一项针对 50 份 X 线胸片报告的研究发现^[24], LLMs (GPT-4) 在根据影像描述生成诊断结论时, 其连贯性、全面性和事实一致性仍显著低于影像科医生, 且存在更高的医疗风险。

结构化报告因其标准化、易解读和便于数据挖掘等优势, 正逐渐成为影像报告的发展趋势^[25]。然而, 实施结构化报告具有挑战性, 尤其是在将大量现有的自由文本报告转换为结构化格式时。Adams 等^[26] 创新性使用 GPT-4 将自由文本报告自动转换为结构化格式, 该系统能够智能选择最适合的报告模板, 单份报告处理成本仅约 0.1 美元。Hasani 等^[27] 的对比研究显示, GPT-4 生成的结构化报告在多个质量维度上优于传统影像科医生报告, 特别是在清晰度、易理解性和结构规范性等方面差异具有统计学意义。在胰腺癌 CT 评估方面, 相关研究证实 GPT-4 不仅能生成高质量的结构化报告, 还能准确评估肿瘤可切除性, 其生成报告被外科医生认为更准确且节省时间^[28]。此外, LLMs 在核医学和介入影像学报告生成中也展现出应用潜力^[29,30]。最新发布的 GPT-4o(omni) 模型还具备多语言语音转文本能力^[31], 为影像报告工作流程的进一步优化提供了新可能。基于上述研究, GPT-4 可以成为生成标准化影像学报告的可靠工具, 有望提高影像科工作效率、临床沟通效率并简化数据提取和分析流程。

在报告质量控制方面, Gertz 等^[32] 研究发现 GPT-4 检测报告错误的速度与影像科医生相当, 但处理速度更快、成本更低。Schmidt 等^[33] 通过比较五种 LLMs (GPT-3.5-turbo、GPT-4、text-davinci-003、Llama-v2-70B-chat 和 Bard) 对于影像报告的错误检测性能, 结果显示 GPT-4 以 76.9% 的准确率表现最佳。

4. 科学研究

虽然根据国际医学期刊编辑委员会的标准, LLMs 不符合学术作者资格^[34], 但其作为科研辅助工具的价值已得到广泛认可。在文献综述方面, LLMs 能够显著提升文献检索和初步归纳的效率, 但其生成的综述内容仍需专业人员严格把关^[6]。在数据标注和数据库构建方面, LLMs 展现出独特优势。当今的影像学数据贯穿于整个医疗的全过程, 数据规模庞大、更新频繁且类型多样。系统构建一个庞大而复杂的图相关联数据库往往耗时耗力, 常需要医师逐个浏览和手动标注。LLM 通过标注报告的训练, 有望通过影像学

诊断的疾病谱对报告进行归纳整理, 关联图文数据, 从而以点及面构建不同的专病、共病及随访数据库。这可以帮助影像科医师高效地管理影像数据, 进行数据分析及征象总结, 更好地利用影像数据进行研究和临床应用。Mohamad 等^[35] 使用 LLMs 对踝关节 X 线报告进行自动标注, 生成的标签用于训练卷积神经网络 (convolutional neural networks, CNN) 识别踝关节骨折, 最终模型准确率达到 89.5%, 与人工标注训练的效果相当。Li 等^[36] 仅用 8 小时就完成了 333 例肺癌患者临床信息的提取, 且数据质量与人工提取相当。类似地, 本地化部署的 LLMs 在识别失代偿性肝硬化特征和头痛症状方面也表现出高灵敏度和特异度^[37,38]。这些研究体现了 LLMs 在加速提取真实世界数据以用于观察性研究方面具有巨大潜力。在数据分析方面, OpenAI 开发的高级数据分析工具 (原称为代码解释器) 作为 GPT-4 插件 (OpenAI. <https://chat.openai.com/>), 能够根据自然语言指令完成数据清洗、基础分析和可视化等任务, 大幅提升了研究效率。但需要注意的是, 高级数据分析功能目前仍处于测试阶段, 输出结果需要专业人员验证。

5. 影像学考试及医学生教育

在影像学专业考核方面, 相关研究表明 GPT-4 能够通过基于文本的影像科委员会考试, 且其推理能力显著优于前代 GPT-3.5^[39,40]。同时, GPT-4 可以用来开发影像学委员会格式的选择题和解析, 以丰富教学资源库^[41]。在医学教育领域, LLMs 为教学改革提供了新思路。通过个性化学习路径设计、互动式案例模拟和临床技能培训等方式, LLMs 有望显著提升医学教育质量^[42]。但需要强调的是, 考试通过率并不能完全代表 LLMs 在实际临床环境中的可靠性和安全性。

6. 健康宣教及患者沟通

医学影像报告通常包含大量专业术语, 容易造成患者理解困难, 特别是当报告提示异常发现时, 可能引发不必要的焦虑^[43]。相关研究表明, LLMs 不仅能够准确回答患者关于各类疾病的疑问, 在沟通同理心方面甚至优于部分医生^[44]。在肺癌预防、筛查和影像学报告相关咨询中, ChatGPT-3.5 的回答准确率 (70.8%) 显著高于 Google Bard^[45]。Doshi 等^[46] 的研究表明 LLMs 可有效简化影像报告, 提高报告的可读性。尽管 LLMs 可以提高患者对影像学报告的理解度, 但其准确性仍有待提高^[47]。正如 Doshi 等^[46] 所提出的, 将 LLMs 生成的简化报告与影像科医生出具的原始报告一起整合到医疗记录中, 可能会减轻患者的焦虑和困惑。LLMs 在患者沟通方面成果显著。相关研究表明 GPT-4 对急性肝衰竭相关问题的回答准确性、清晰度和相关性方面得分均较高^[10]。在肝硬化和

肝细胞癌治疗方面, ChatGPT 虽然能够正确回答诊断、治疗和预防医学等领域的问题, 但回答深度和实用性仍有不足^[48]。在乳腺癌筛查宣教方面, ChatGPT 也展现出较大的应用潜力^[49]。

总结与展望

尽管 LLMs 在医学影像学领域展现出广阔的应用前景, 但仍面临若干重要挑战^[6,50], 主要包括如下:

①数据隐私与安全: 医疗数据的敏感性和隐私保护要求使得模型训练和应用面临严格的合规审查; ②模型幻觉问题: LLMs 可能生成看似合理但实际错误的医学信息, 这种“幻觉”现象在临床应用中可能造成严重后果, 需要开发更可靠的验证机制; ③领域适应性局限: 通用 LLMs 在专业医学知识理解和应用方面仍存在不足, 需要针对医疗场景进行专门的优化和微调; ④评估标准缺失: 目前缺乏统一的评估框架和标准来衡量 LLMs 在医疗场景中的性能和安全性, 亟需建立专业的评测体系; ⑤伦理与责任问题: 当 LLMs 辅助或参与医疗决策时, 如何界定医疗责任成为亟待解决的法律和伦理问题; ⑥资源密集: 需要大量资源包括计算资源、数据资源(高质量数据集)和基础设施建设。

不同 LLMs 的技术架构差异显著影响其在医学影像任务中的性能。在模型架构方面, 基于单向 Transformer 解码器的模型(如 GPT 系列)擅长序列生成, 适用于报告自动生成; 而混合专家模型(mixture of experts, MoE)通过激活少量专家网络处理输入, 在提升模型容量的同时保持计算效率, 更适用于多任务医疗场景。此外, 多模态大模型(如 GPT-4V)通过整合视觉编码器处理影像信息, 但当前视觉-语言对齐质量仍待提升。在训练与医学适配方面, RAG 技术通过整合外部医学知识库, 有效减少模型幻觉。任务性能方面, 模型上下文窗口长度直接影响其处理长序列临床报告的能力, 窗口扩展有助于整合分散的临床信息。未来需开发更统一的评估框架, 以量化不同架构模型在具体医学影像任务中的效能与可靠性。

MLLMs 在医学影像学中具有显著潜力。通过融合影像、文本、时间序列等多源信息, MLLMs 有望实现更全面的临床辅助决策, 例如在影像中直接识别病变并关联临床指南生成诊断建议, 或集成多期相、多模态影像实现动态病情评估。此外, MLLMs 也可用于构建更自然的医患交互界面, 如基于影像自动生成患者可读的健康说明。然而, 实现这一愿景仍需解决多模态对齐、标注数据稀缺、计算复杂度高和跨中心泛化等问题。国内外正在加强相关研究合作, 未来有望出现更安全、高效、透明的多模态医疗 AI 系统。

值得注意的是, 当前国内外研究在 LLMs 应用于

医学影像领域呈现出不同的发展特点。在国际上, 美国、欧洲等多以大型科技公司(如 Google、OpenAI)与顶尖学术机构合作推进, 研究方向偏重于多模态大模型(如 GPT-4V 等)在影像-报告关联分析、跨模态检索与诊断推理中的探索。相比之下, 国内研究更注重结合临床实际需求, 华为、阿里巴巴、腾讯等企业联合多家医院开展了许多针对特定疾病(如肝癌、肺结节)的本地化模型研发, 清华大学、北京大学等机构也在推动医学 LLMs 的研究(如华佗 GPT、MedicalGPT), 但在多模态融合与国际化影响力方面仍有提升空间。

大语言模型正在深刻改变医学影像学的研究和实践模式。当前 LLMs 在影像诊断、临床分期、报告生成、科研辅助、影像医学考试与教育、健康宣教与患者沟通等方面已展现出实用价值, 但在直接影像诊断等核心任务上仍存在明显局限。随着技术的不断进步和临床验证的深入, LLMs 有望成为影像科医生的智能助手, 共同提升医疗质量和效率。然而, 实现这一愿景需要技术人员、临床医生、管理者和政策制定者的共同努力, 以解决技术、伦理和管理等方面的挑战。

参考文献:

- [1] Nakao T, Miki S, Nakamura Y, et al. Capability of GPT-4V (ision) in the Japanese national medical licensing examination: evaluation study[J]. *JMIR Med Educ*, 2024, 10: e54393.
- [2] Kim S, Lee CK, Kim SS. Large language models: a guide for radiologists[J]. *Korean J Radiol*, 2024, 25(2): 126-133.
- [3] Xu P, Zhu X, Clifton DA. Multimodal learning with transformers: a survey[J]. *IEEE Trans Pattern Anal Mach Intell*, 2023, 45(10): 12113-12132.
- [4] 刘泓泽, 王耀国, 唐圣晟, 等. 医学大语言模型的应用现状与发展趋势研究[J]. *中国数字医学*, 2024, 19(8): 1-7, 13.
- [5] Han T, Adams LC, Bressen KK, et al. Comparative analysis of multimodal large language model performance on clinical vignette questions[J]. *JAMA*, 2024, 331(15): 1320-1321.
- [6] Bhayana R. Chatbots and large language models in radiology: a practical primer for clinical and research applications[J]. *Radiology*, 2024, 310(1): e232756.
- [7] Arıkanoglu Z, Turkoglu A, Taskesen F, et al. Factors affecting morbidity and mortality in hollow visceral injuries following blunt abdominal trauma[J]. *Clin Ter*, 2014, 165(1): 23-26.
- [8] O'Neill SB, Hamid S, Nicolaou S, et al. Changes in approach to solid organ injury: what the radiologist needs to know[J]. *Can Assoc Radiol J*, 2020, 71(3): 352-361.
- [9] Wang Y, Zhang J, Li M, et al. SMART: development and application of a multimodal multi-organ trauma screening model for abdominal injuries in emergency settings[J]. *Acad Radiol*, 2025, 32(5): 2655-2666.
- [10] Malik S, Frey LJ, Gutman J, et al. Evaluating artificial intelligence-driven responses to acute liver failure queries: a comparative analysis across accuracy, clarity, and relevance[J]. *Am J Gastroenterol*, 2024, Dec 17. DOI: 10.14309/ajg.

000000000003255.

- [11] Horiuchi D, Tatekawa H, Oura T, et al. ChatGPT's diagnostic performance based on textual vs. visual information compared to radiologists' diagnostic performance in musculoskeletal radiology [J]. *Eur Radiol*, 2025, 35(1): 506-516.
- [12] Mitsuyama Y, Tatekawa H, Takita H, et al. Comparative analysis of GPT-4-based ChatGPT's diagnostic performance with radiologists using real-world radiology reports of brain tumors [J]. *Eur Radiol*, 2025, 35(4): 1938-1947.
- [13] Kottlors J, Bratke G, Rauen P, et al. Feasibility of differential diagnosis based on imaging patterns using a large language model [J]. *Radiology*, 2023, 308(1): e231167.
- [14] Brin D, Sorin V, Barash Y, et al. Assessing GPT-4 multimodal performance in radiological image analysis [J]. *Eur Radiol*, 2025, 35(4): 1959-1965.
- [15] Zhang Y, Wu L, Wang Y, et al. Unexpectedly low accuracy of GPT-4 in identifying common liver diseases from CT scan images [J]. *Dig Liver Dis*, 2024, 56(4): 718-720.
- [16] Khader F, Müller-Franzes G, WANG T, et al. Multimodal deep learning for integrating chest radiographs and clinical parameters; a case for transformers [J]. *Radiology*, 2023, 309(1): e230806.
- [17] Li Y, Zheng X, Li J, et al. LKAN; LLM-Based knowledge-aware attention Network for clinical staging of liver cancer [J]. *IEEE J Biomed Health Inform*, 2025, 29(4): 3007-3020.
- [18] Matute-González M, Darnell A, Comas-Cufi M, et al. Utilizing a domain-specific large language model for LI-RADS v2018 categorization of free-text MRI reports; a feasibility study [J]. *Insights Imaging*, 2024, 15(1): 280.
- [19] Gu K, Lee JH, Shin J, et al. Using GPT-4 for LI-RADS feature extraction and categorization with multilingual free-text reports [J]. *Liver Int*, 2024, 44(7): 1578-1587.
- [20] Suzuki K, Yamada H, Yamazaki H, et al. Preliminary assessment of TNM classification performance for pancreatic cancer in Japanese radiology reports using GPT-4 [J]. *Jpn J Radiol*, 2025, 43(1): 51-55.
- [21] Suzuki K, Claude 3.5 sonnet indicated improved TNM classification on radiology report of pancreatic cancer [J]. *Jpn J Radiol*, 2025, 43(1): 56-57.
- [22] Wu Q, Wu Q, Li H, et al. Evaluating large language models for automated reporting and data systems categorization; cross-sectional study [J]. *JMIR Med Inform*, 2024, 12: e55799.
- [23] Huang J, Neill L, Wittbrodt M, et al. Generative artificial intelligence for chest radiograph interpretation in the emergency department [J]. *JAMA Netw Open*, 2023, 6(10): e2336100.
- [24] Sun Z, Ong H, Kennedy P, et al. Evaluating GPT4 on impressions generation in radiology reports [J]. *Radiology*, 2023, 307(5): e231259.
- [25] Larson DB, Towbin AJ, Pryor RM, et al. Improving consistency in radiology reporting through the use of department-wide standardized structured reporting [J]. *Radiology*, 2013, 267(1): 240-250.
- [26] Adams LC, Truhn D, Busch F, et al. Leveraging GPT-4 for post hoc transformation of free-text radiology reports into structured reporting; a multilingual feasibility study [J]. *Radiology*, 2023, 307(4): e230725.
- [27] Hasani AM, Singh S, Zahergivar A, et al. Evaluating the performance of Generative Pre-trained Transformer-4 (GPT-4) in standardizing radiology reports [J]. *Eur Radiol*, 2024, 34(6): 3566-3574.
- [28] Bhayana R, Nanda B, Dehkharghanian T, et al. Large language models for automated synoptic reports and resectability categorization in pancreatic cancer [J]. *Radiology*, 2024, 311(3): e233117.
- [29] Huemann Z, Lee C, Hu J, et al. Domain-adapted large language models for classifying nuclear medicine reports [J]. *Radiol Artif Intell*, 2023, 5(6): e220281.
- [30] Can E, Uller W, Vogt K, et al. Large language models for simplified interventional radiology reports; a comparative analysis [J]. *Acad Radiol*, 2025, 32(2): 888-898.
- [31] Busch F, Prucker P, Komenda A, et al. Multilingual feasibility of GPT-4o for automated Voice-to-Text CT and MRI report transcription [J]. *Eur J Radiol*, 2025, 182: 111827.
- [32] Gertz RJ, Dratsch T, Bunck AC, et al. Potential of GPT-4 for detecting errors in radiology reports; implications for reporting accuracy [J]. *Radiology*, 2024, 311(1): e232714.
- [33] Schmidt RA, Seah JCY, Cao K, et al. Generative large language models for detection of speech recognition errors in radiology reports [J]. *Radiol Artif Intell*, 2024, 6(2): e230205.
- [34] Flanagan A, Bibbins-domingo K, Berkwits M, et al. Nonhuman "Authors" and implications for the integrity of scientific publication and medical knowledge [J]. *JAMA*, 2023, 329(8): 637-639.
- [35] Mohamad FA, Donle L, Dorfner F, et al. Open-source large language models can generate labels from radiology reports for training convolutional neural Networks [J]. *Acad Radiol*, 2025, 32(5): 2402-2410.
- [36] Li Y, Law J, Le LW, et al. Assessing the feasibility and external validity of natural language processing-extracted data for advanced lung cancer patients [J]. *Lung Cancer*, 2025, 199: 108080.
- [37] Wiest IC, Ferber D, Zhu J, et al. Privacy-preserving large language models for structured medical information retrieval [J]. *NPJ Digit Med*, 2024, 7(1): 257.
- [38] Le Guellec B, Lefèvre A, Geay C, et al. Performance of an open-source large language model in extracting information from free-text radiology reports [J]. *Radiol Artif Intell*, 2024, 6(4): e230364.
- [39] Bhayana R, Bleakney RR, Krishna S. GPT-4 in radiology: improvements in advanced reasoning [J]. *Radiology*, 2023, 307(5): e230987.
- [40] Bhayana R, Krishna S, Bleakney RR. Performance of ChatGPT on a radiology board-style examination: insights into current strengths and limitations [J]. *Radiology*, 2023, 307(5): e230582.
- [41] Mistry NP, Saeed H, Rafique S, et al. Large language models as tools to generate radiology board-style multiple-choice questions [J]. *Acad Radiol*, 2024, 31(9): 3872-3878.
- [42] Ballard DH, Antigua-Made A, Barre E, et al. Impact of ChatGPT and large language models on radiology education; association of academic radiology-radiology research alliance task force white paper [J]. *Acad Radiol*, 2025, 32(5): 3039-3049.
- [43] Anyidoho PA, Verschraegen CF, Markham MJ, et al. Impact of

- the immediate release of clinical information rules on health care delivery to patients with cancer[J].JCO Oncol Pract, 2023, 19(5):e706-e713.
- [44] Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum[J]. JAMA Intern Med, 2023, 183(6):589-596.
- [45] Rahsepar AA, Tavakoli N, Kim GHJ, et al. How AI responds to common lung cancer questions: ChatGPT vs Google bard[J]. Radiology, 2023, 307(5):e230922.
- [46] Doshi R, Amin KS, Khosla P, et al. Quantitative evaluation of large language models to streamline radiology report impressions: a multimodal retrospective analysis[J]. Radiology, 2024, 310(3):e231593.
- [47] Amin KS, Davis MA, Doshi R, et al. Accuracy of ChatGPT, Google bard, and microsoft bing for simplifying radiology reports[J]. Radiology, 2023, 309(2):e232561.
- [48] Yeo YH, Samaan JS, Ng WH, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma[J]. Clin Mol Hepatol, 2023, 29(3):721-732.
- [49] Haver HL, Ambinder EB, Bahl M, et al. Appropriateness of breast cancer prevention and screening recommendations provided by ChatGPT[J]. Radiology, 2023, 307(4):e230424.
- [50] Thirunavukarasu AJ, Ting DSJ, Elangovan K, et al. Large language models in medicine[J]. Nat Med, 2023, 29(8):1930-1940.

(收稿日期:2025-07-03 修回日期:2025-08-28)

欢迎订阅 2026 年《放射学实践》

《放射学实践》是由国家教育部主管,华中科技大学同济医学院主办,与德国合办的全国性影像学学术期刊,创刊至今已 41 周年。遵照同行评议、价值导向、等效应用原则,国内各大学会、协会、组织机构通过科技工作者推荐、专家评议、结果公示等规定程序,《放射学实践》杂志入选中国科协发布 10 大领域《我国高质量科技期刊分级目录》业内认可的较高水平期刊。《放射学实践》杂志再次入选 2023 年版北京大学和北京高校图书馆期刊工作研究会共同主持的国家社会科学基金项目“学术期刊评价及文献计量学研究”研究成果——《中国核心期刊要目总览》。

本刊坚持服务广大医学影像医务人员的办刊方向,关注国内外影像医学的新进展、新动态,全面介绍 X 线、CT、磁共振、介入放射及放射治疗、超声诊断、核医学、影像技术学等医学影像方面的新知识、新成果,受到广大影像医师的普遍喜爱。

本刊为中国科学引文数据库(CSCD)核心库来源期刊、《中文核心期刊要目总览》收录期刊、中国科技核心期刊、中国高质量科技期刊分级目录临床医学领域医学影像学期刊,并在中国学术期刊分区中位列 Q1 区。

主要栏目:论著、继续教育园地、专家荐稿、研究生展版、图文讲座、本刊特稿、实验研究、传染病影像学、影像技术学、外刊摘要、学术动态、请您诊断、病例报道、知名产品介绍、信息窗等。

本刊为月刊,每册 25 元,全年定价 300 元。

国内统一刊号:ISSN 1000-0313/CN 42-1208/R 邮政代号:38-122

电话:(027)69378385

E-mail:fsxsjzz@163.com 网址:http://www.fsxsj.net

编辑部地址:430199 武汉市蔡甸区中法新城同济医院专家社区别墅 C 栋

· 人工智能影像学专题 ·

全流程影像人工智能创新研究与智慧流程管理标准解读

白义钧, 陆加明, 陈夫涛, 胡晴, 王正阁, 张鑫, 张冰

【摘要】 医学影像在临床诊疗中的重要性日益提升, 人工智能(AI)技术正逐步应用于影像科全流程管理。本文对南京鼓楼医院牵头制定的两项行业标准进行解读, 系统阐述了 AI 在预约调度、图像采集、质量控制、辅助诊断、报告生成及设备管理等环节的应用进展。同时, 探讨了大语言模型在影像诊断与教学中的应用实践。最后, 针对当前存在的数据标准不统一、模型泛化能力不足、可解释性欠缺及伦理问题等挑战, 提出构建标准化、可控性强且可持续发展的影像 AI 生态体系的发展方向。

【关键词】 人工智能; 医学影像; 全流程管理; 大语言模型

【中图分类号】 TP18; R81 **【文献标识码】** A **【文章编号】** 1000-0313(2025)09-1088-05

DOI: 10.13609/j.cnki.1000-0313.2025.09.003

开放科学(资源服务)标识码(OSID):



人工智能(artificial intelligence, AI)快速发展, 对医疗健康领域产生了深远影响。在医疗人工智能众多应用方向中, 医学影像占据了 20% 的应用场景, 是目前应用最广泛、价值最突出的方向之一。国家卫生健康委员会等多个部门联合发布了《关于印发卫生健康行业人工智能应用场景参考指引》的通知, 要求积极推进卫生健康行业“人工智能+”应用创新发展。据统计, 在本文提出的 83 个应用场景中, 医学影像结合人工智能衍生出医学影像智能辅助诊断、医学影像数据智能辅助质控、临床专病智能辅助决策、基层全科医生智能辅助决策、医学影像智能辅助治疗等应用场景。在此背景下, 医学影像人工智能既符合国家的发展战略, 也满足临床医疗的实际需求。

医学影像在临床诊断中的重要性日益凸显。在医疗诊断相关数据中, 医学影像已成为支撑临床决策的核心信息来源之一, 约占 70%。尤其在大型三甲等医院, CT 与 MRI 等影像技术作为主要的诊断工具, 其门诊使用率可达 50% 以上; 而在非三甲医院中, 该类影像检查的使用率通常低于 25%。这一显著差异体现了高等级医院在影像资源配置与临床依赖度方面的集中趋势。随着影像数据规模的持续增长, 传统人工分析手段已难以满足高效、精准诊断的需求。因此, 借助人工智能与大模型等先进技术, 对影像数据进行自动化处理与智能辅助诊断, 成为提升诊疗效率与质量

的关键路径。

基于目前国家发展战略及临床实际需求, 南京鼓楼医院牵头发布了《江苏省“影像平台”质控标准》及《区域远程医学影像中心图像及诊断报告质量控制》两个行业标准, 本文将对这两个标准进行详细解读, 结合人工智能及其最新成果“大模型”在临床工作全流程中的多个应用环节进行综述。

医学影像全流程管理概念提出

《江苏省“影像平台”质控标准》(项目号: 苏医管 2022(9))中, 首次提出了“影像平台流程管理”这一创新概念。医学影像科全流程管理的应用场景主要包含预约登记、报到候诊、护理服务、图像采集、信息传输与存储、对比剂及其反应的预防与处理、图像后处理与排版、辅助诊断及结构化报告、图像与报告的浏览打印与处理、设备管理等共 10 个一级流程(图 1)。首先, 预约服务是指患者提前安排影像检查的过程, 主要用于确认患者信息和预约检查时间。接下来, 分诊服务会根据患者的症状和需求, 将其合理分流至相应的检查项目和影像科室。然后, 护理服务将为患者提供扫描前的准备工作与协助, 包括帮助患者正确定位和缓解紧张情绪。在图像采集阶段, 影像设备将进行实际检查, 获取所需的影像数据。随后, 采集的图像将通过信息传输与存储系统进行数字化传输和存储, 以便后续访问和诊断。若使用对比剂, 在对比剂不良事件处理环节会及时应对可能发生的不良反应, 确保患者安全。图像后处理与排版环节会对图像进行优化调整, 确保图像清晰度满足诊断需求。接着, 医生将通过诊断服务对图像进行分析并作出医学判断。图像与报告服务将生成诊断报告, 并与患者和医生共享。最后, 设备管理环节将确保影像设备的正常运行, 包括维护和维修

作者单位: 210008 南京, 南京大学医学院附属鼓楼医院医学影像科(白义钧、陆加明、陈夫涛、胡晴、王正阁、张鑫、张冰); 210033 南京, 南京大学社会学院心理系(白义钧); 210008 南京, 南京大学医学影像与人工智能研究所(张冰); 210033 南京, 南京大学脑科学研究院(张冰)

作者简介: 白义钧(1996-), 女, 黑龙江绥化市人, 博士研究生, 主要从事医学影像人工智能研究工作。

通讯作者: 张冰, E-mail: zhangbing_nanjing@nju.edu.cn

基金项目: 国家自然科学基金(82330059, 82271965); 南京市卫生科技发展专项一般性课题(YKK22083, YKK23103); 南京鼓楼医院临床研究专项资金项目(2024-LCYJ-PY-33); 江苏省卫健委医学科研面上项目(M2024057)

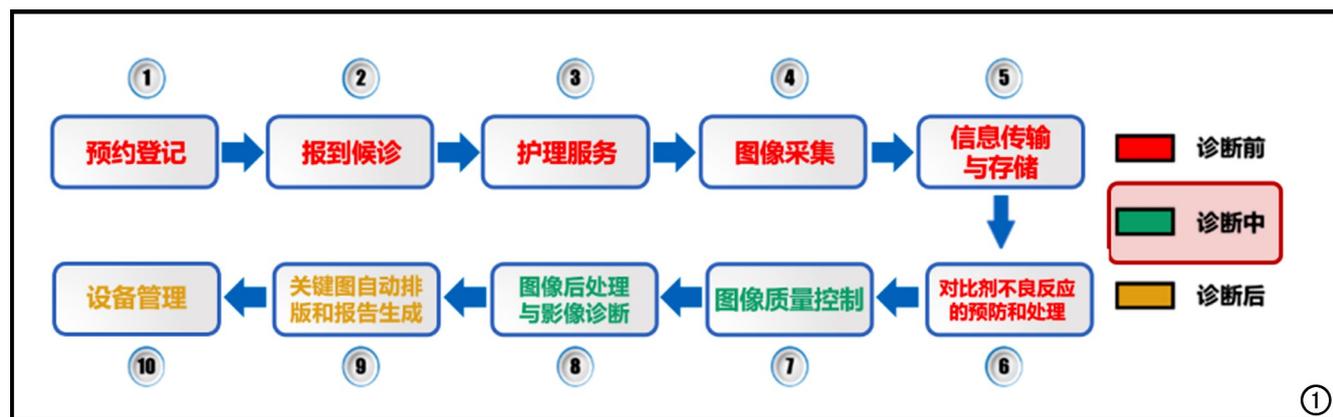


图 1 医学影像科全流程管理的一级流程。

工作,使设备始终保持最佳状态。整个流程通过精细化管理,确保影像检查过程高效、安全且准确。

基于患者就诊流程,医学影像诊断流程可分为诊断前、诊断中及诊断后三个阶段,涵盖诊断过程的核心环节。将人工智能技术融入医学影像科全流程管理的各种应用场景,有望实现对诊断各阶段流程的智能化重构与优化,从而全面提升影像诊断效率与质量。

1. AI 赋能“诊断前”阶段

AI 技术正逐步实现从单点应用向全流程型集成系统的转变。该流程包括预约登记、报到候诊、护理服务、图像采集、信息传输与存储、对比剂不良反应的预防和处理等 6 个一级流程。AI 赋能的前置流程系统在实际部署中可显著提升预约效率、缩短平均排队时间并优化就诊流程,为推动影像科室全流程智能化奠定基础^[1]。

在预约登记、报道候诊与护理服务环节,依托自然语言处理(natural language processing, NLP)、知识图谱、智能调度与推荐算法,医学影像科室可构建智能化预约系统,对患者主诉进行语义分析并精准匹配相应的影像检查类型和时段,从而提高检查预约效率,缓解资源配置不合理问题^[2,3]。

在图像采集环节,AI 算法支持的自适应成像协议已逐步取代传统固定参数模式,为个性化精准影像提供更高灵活性与系统智能性^[4,5]。相关研究表明,强化学习与贝叶斯优化逐步被引入 MRI 与 CT 的扫描协议自适应配置中,实现对扫描电压、层厚、曝光时间等参数的动态最优调控^[6],对抗生成网络(generative adversarial network, GAN)能在保持结构细节的同时显著提升图像的信噪比与清晰度,已广泛应用于低剂量与 MRI 影像增强^[7]。与此同时,部分医疗机构已部署基于深度视觉与空间定位技术的非接触式体位引导系统,实现患者在扫描前的智能化位置校准与检查准备,从而缩短流程耗时、减少操作误差并提升感染控制

能力^[1,8]。

信息传输与存储是实现模型临床部署的核心环节。PACS-AI 平台通过集成 DICOM 标准,确保影像数据在不同系统间的兼容性,并采用本地存储结合云端推理的混合架构,在保障患者隐私的前提下提升处理效率。系统支持加密传输与访问控制,确保图像数据在存储与传输过程中的安全性与合规性,有效支撑多中心临床应用与模型验证^[9,10]。

2. AI 赋能“诊断中”阶段

该环节包括图像质量控制、影像后处理与影像诊断两个一级流程。

在医学影像人工智能(AI)系统的研发与临床转化过程中,质量控制(quality control, QC)是确保影像及 AI 报告在实际临床中能够持续稳定地发挥作用,减少潜在风险,提高医疗决策的安全性和可靠性的重要保障。在 2025 年发表在 Nature 的文章提出“生成式医学图像理解(GenMI)”的新范式,系统性探讨了利用多模态大模型(如 GPT-4、Gemini)从医学图像生成临床报告,并实现自动化质量控制。文章明确指出,AI 系统不仅可生成放射报告,还可通过引入第二个 AI 模型对其输出进行内容审核、逻辑一致性检验和幻觉识别,构成“AI 质控 AI”的闭环流程。这一模式强调利用知识图谱、病灶检测模型及外部知识增强机制,有效提升报告准确性与可解释性,为建立可信、多功能、持续反馈的医学 AI 系统奠定了基础^[11]。影像质量控制是保障诊断准确性与可靠性的关键环节,传统影像质控方式依赖于人工经验,面临效率低、主观性强、重复劳动等问题。Sanchez 等^[12]基于胎儿 T₂W MRI 影像提取 332 个影像质量控制指标,建立开源机器学习框架 FetMRQC,用于预测专家评分并实现对低质量影像的自动筛选。dMRI 广泛用于生物组织的微观结构和大脑的结构连接性且容易受到多种伪影的影响。Samani 等^[13]开发基于 CNN 和迁移学习的自

自动化质量控制工具“OC-Automator”,旨在预测 dMRI 中的各种伪影,如运动伪影、磁敏感伪影等,准确率高达 98%。基于人工智能的方法通过自动化影像质量评估,显著提高了影像控制的效率和准确性。通过这些创新的 AI 方法,影像质控的效率与准确性得到了显著提升,进而优化了临床影像分析流程。

在 AI 辅助诊断环节, AI 影像人工智能辅助诊断方面取得一系列进展。如通过融合 DCE-MRI 多模态影像组学特征,结合机器学习技术,以实现成人胶质瘤 IDH1 突变状态的高准确率预测^[14];依据 Logistic 回归分析结果,结合常规 MRI 特征与 DCE-MRI 定量及定性参数,建立术前预测 CDKN2A/B 纯合缺失状态的模型。研究发现, Ktrans 在全体患者及 IDH 突变型、IDH 野生型亚组中均为独立预测因子,预测模型 AUC 分别为 0.90、0.95 和 0.84。结果表明, DCE-MRI 具备良好的分子病理预测价值,有望为胶质瘤的精准分型和个体化治疗提供重要影像学支持^[15]。基于 MRI 影像组学与 ResNet 深度学习网络融合的模型,用于非侵入性预测弥漫性胶质瘤的 MGMT 启动子甲基化状态^[16]。此外,采用丘成桐院士提出的计算共形几何方法,结合创新型 SE-ResNet 深度学习方法,进行脑胶质瘤 WHO 分级预测。面对稀缺标注样本,尤其是在罕见病与小数据集场景中,小样本学习已被广泛引入医学影像分类任务中。这些方法通过参数共享、度量学习与模型初始化策略,实现了在仅有极少训练样本的情况下,对病灶类型进行有效识别^[17],在临床应用中,该策略已被证实能显著缓解数据不足对模型性能的限制。Mu 等^[18]研究发现, AI 可用于冠状动脉 CT 血管成像(CTA)影像优化,并自动计算冠状动脉钙化评分(CAC)。其计算结果与传统 Agatston 评分高度一致(Pearson 相关性 $r = 0.96, R^2 = 0.92$),并且在低剂量扫描条件下依然保持较高的影像质量。这表明 AI 影像优化技术可在降低辐射剂量的同时保证诊断价值。综上所述,诊断中阶段的 AI 系统正沿着融合感知、多源建模、跨任务协同与可解释性增强的方向发展,为高精度、低延迟、可信任的医学影像诊断提供新范式。

3. 诊断后流程

该流程主要包括关键图自动排版和报告生成、设备管理两个一级流程。

在图像及报告发布环节,鼓楼医院影像科将深度学习和自适应布局结合开发了智能影像布局系统(intelligent imaging layout system, IILS),适合全自动或者半自动的影像学图像的排版问题。可以将以往影像扫描到排版花费约 2 h 的工作时间,降低到约 1000 s 完成,从而大幅提高影像科工作流程的效率、降低医疗

成本、提高影像质量,进一步帮助影像医生、临床医生和患者之间共同构建高质量规范的诊疗体系^[19]。近年来,基于大模型的自动化报告生成在放射学领域取得了显著进展,旨在缓解全球放射科医生短缺所带来的工作负担,并提升影像报告的生成效率和质量,降低人工书写错误率^[20,21]。多模态视觉-语言模型(VLMs),如 Flamingo-CXR,已被用于自动生成胸部 X 光报告,并在多项评估指标上超越现有方法,部分情况下甚至优于人工报告^[22]。此外,结构化信息提取技术(如 RIEEL)通过指令微调,使大模型能够端到端地提取放射学报告中的标准化信息,从而提高信息一致性并降低数据标注成本^[23]。然而,当前大模型仍然面临临床质量评价体系缺乏、难以适应不同临床场景、结构化信息标准化不足以及对罕见病症适应性较低等挑战。未来研究应致力于完善自动化报告的临床评价体系、提升模型的泛化能力,并探索少样本学习和知识蒸馏等技术,以推动自动化报告生成在临床中的广泛应用。此外,传统医学数据等不容易被患者理解, AI 生成报告的可视化和患者友好性设计,能有效提升患者的理解力。如 Gou 等^[24]探讨了 AI 在诊断后影像报告自动化中的实际部署,通过大语言模型(large language models, LLM)如 GPT-4d、DeepSeek 等,将影像分析结果转化为患者友好型描述,并对诊断报告进行解读,将数据分析和挖掘结果以图形化界面形式呈现。

设备管理:智能设备管理作为诊断后流程中的关键一环,通过物联网与 AI 的融合,实现了对医疗影像设备的实时监控与维护调度。Gou 等^[24]在其综述中指出, AI 系统已广泛应用于影像流程路径的优化与设备资源调配,提升了整体医疗系统的运转效率。从组织层面看, Alami 等^[25]强调了医疗机构在 AI 部署后,需同步升级其数据流管理机制与流程协调策略,以确保诊断后环节的“智能连贯性”,实现 AI 真正的临床价值最大化。

图像及诊断报告质量控制

随着远程医学技术的发展,区域影像中心逐步承担起基层医疗机构影像采集、处理与诊断的集约化服务功能。然而,由于参与机构层级差异、设备性能不一及操作流程不规范,图像质量与诊断报告准确性存在较大波动,影响远程诊疗效果与患者安全。在《区域远程医学影像中心图像及诊断报告质量控制》(TCSBME 038-2021)中,首次提出在影像科质量管理中引入“随机质控”、“主动质控”和“集中质控”三大核心机制,并将其贯穿全流程质控管理当中。在图像伪影质控方面,相关研究设置了阅片质控模块、阅片质控配置模块和输出模块,基于医学图像基础阅片信

息实现对医学图像质量控制,以生成匹配的阅片输出信息,该系统能有效提升医生阅片的工作效率,并为后续阅片输出信息的统计和回溯提供有利条件^[26]。在急诊 CT 伪影校正方面,张冰研究团队基于生成对抗网络智能定位算法,通过解剖结构特征自适应机制,使头部 CT 的灰质信噪比(SNR)从 4.47 提升至 8.92,对比噪声比(CNR)从 0.15 优化至 1.13,降低了漏诊率^[27]。在影像报告质控方面,南京鼓楼医院开发了影像报告实时质控系统,该系统基于人工智能技术可自动识别医学影像报告中的常见错误,包括文本错误(如多字、少字、错别字和单位误用)以及方位、部位、性别、年龄和阴阳性等信息的不一致提示,通过对报告内容进行智能审校,显著降低了人为低级错误的发生率。此举不仅提升了影像报告的质量与准确性,还有效减少了因报告错误引发的潜在医疗纠纷,相关成果已被收录于《人工智能蓝皮书:中国医疗人工智能发展报告(2023)》,具有重要的推广与示范价值。

大模型在医学影像及全流程管理中的应用

全国各大医院已启动 AI 大模型研发与部署,赋能智慧医疗驶入快车道。如北京协和、中山大学附属医院、东南大学附属医院等已建立智慧医疗体系,具备大模型训练、部署及应用能力。如南京鼓楼医院使用 DeepSeek-R1 (670B),研发了多场景医疗智能体,累计服务患者近 20 万人次,实现从智能问诊到病例生成以及检查报告诊断生成的全链条 AI 赋能。在此背景下,人工智能在临床诊断中的应用研究也不断深化。近期,研究者提出了“AI Hospital”多智能体交互诊断框架,进一步推动了大语言模型在真实医疗环境中的评估与落地。该框架构建了由“医生”(由 LLM 扮演)与虚拟患者、检查员、主任医师等 NPC 角色组成的多轮交互系统,模拟实际门诊中的问诊、检查推荐、诊断决策等全过程。AI Hospital 同时构建了多维医学评估体系(multi-view medical evaluation, MVME),从症状收集、检查建议、诊断和治疗等多个维度量化大语言模型表现。研究还引入协作式诊断机制,通过多模型联合问诊和“争议解决”策略提升诊断准确率。实验表明,即便是 GPT-4 等先进模型,其交互式诊断能力仍远低于单轮诊断结果,显示出当前大模型在临床多轮推理、问诊策略等方面仍有显著提升空间^[28]。这一系列成果不仅验证了大模型在医疗领域的应用潜力,也为构建具备实际临床应用价值的 AI 诊断系统提供了重要参考。

为推动医学教育智能化与个性化进程,南京鼓楼医院医学影像科基于大语言模型构建了“鼓医神影学苑”教学系统,深度融合 DeepSeek 模型,探索医学影

像教学场景的智能转型。该系统实现了教学资源的智能生成、内容推送的个性化推荐及影像学辅助学习全流程支持,标志着基于 LLM 的医学影像教育平台正式落地。教学资源模块依托 DeepSeek 的自然语言生成能力,自动构建高质量题库与内容,极大提升了开发效率与更新频次;个性化教学模块则基于学生画像与学习路径,为不同类型学习者匹配差异化教学内容,提升教学精准度;系统还搭载专属智能学辅工具,支持 AI 问答、自动解析、学习难点辅助解释等功能,构建影像学生认知促进与精准反馈闭环;此外,通过“能力成长图谱”,量化学生在影像判读速度、诊断准确率、临床思维完整度等关键指标上的成长情况,构建多维度的动态评价体系。该系统已于 2024 年 12 月申请发明专利,初步形成了医学影像教学智能化、标准化的建设模板。

挑战与未来方向

人工智能在医学影像全流程管理中的应用仍面临多重挑战。数据标准化方面,医疗机构间影像设备、扫描参数及数据格式的差异导致模型跨中心迁移困难^[29]。模型泛化能力方面,基于单一数据集训练的系统在跨中心应用时需额外调优,反映出数据分布差异对性能的影响^[30]。可解释性层面,深度神经网络的“黑箱”特性降低了临床信任度^[31]。伦理合规方面,大规模数据需求与隐私保护存在固有矛盾,且模型偏倚可能加剧医疗公平性问题^[32]。应用场景方面,模型已拓展至放射科、病理科及远程医疗等多科室,提高了服务的可及性与效率。系统落地层面,AI 报告生成系统虽提升了效率,但需解决与现有工作流程整合及计算资源依赖等工程难题。这些限制需要通过数据标准化建设、可解释模型开发、鲁棒性增强算法及隐私保护技术创新等途径系统解决^[33]。

综上所述,医学影像 AI 的发展已跨越技术单点突破阶段,进入“标准牵引生态重构”的新纪元。未来研究需重点突破技术可控性、标准共识性、生态可持续性三大核心命题,最终实现从“赋能工具”到“医疗新基建”的范式跃迁。

参考文献:

- [1] Lv W, Zhu W, Wang M, et al. Application of artificial intelligence in optimizing medical imaging workflows [M]. Berlin: Springer, 2024:111-131.
- [2] Zhou Q, Peng W, Tang D. Automatic recommendation of medical departments to outpatients based on text analyses and medical knowledge graph [J]. J Intell Fuzzy Sys, 2021, 41(2):3289-3299.
- [3] Yu G, Tabatabaei M, Mezei J, et al. Improving chronic disease management for children with knowledge graphs and artificial intelligence [J]. Expert Syst Appl, 2022, 201:117026.

- [4] Zhang M, Gu S, Shi Y. The use of deep learning methods in low-dose computed tomography image reconstruction; a systematic review[J]. *Complex Intell Sys*, 2022, 8(6): 5545-5561.
- [5] Salaudeen HD, Aleem AO, Solomon EU. Mitigating radiation effects and enhancing patient comfort through image processing and data analysis: utilizing deep learning models and state-of-the-art technology in medical imaging[J]. *Int Res J Mod Eng Technol Sci*, 2024, 6: 2270-2286.
- [6] Tariq K, Munir MA, Aftab HT, et al. Adaptive learning algorithms for low dose optimization in coronary arteries angiography: a comprehensive review[J]. *J Rawalpindi Med Coll*, 2024, 28(2): 336-347.
- [7] Wolterink JM, Leiner T, Viergever MA, et al. Generative adversarial networks for noise reduction in low-dose CT[J]. *IEEE trans Med imaging*, 2017, 36(12): 2536-2545.
- [8] Wang Y, Lu X, Zhang Y, et al. Precise pulmonary scanning and reducing medical radiation exposure by developing a clinically applicable intelligent CT system: toward improving patient care[J]. *EBioMedicine*, 2020, 54: 102724.
- [9] Theriault-Lauzier P, Cobin D, Tastet O, et al. A responsible framework for applying artificial intelligence on medical images and signals at the point of care: the PACS-AI platform[J]. *Can J Cardiol*, 2024, 40(10): 1828-1840.
- [10] 中国急诊医学影像数据库协作组. 急诊医学影像数据库构建及标注专家共识[J]. *中华放射学杂志*, 2024, 58(5): 479-487.
- [11] Rao VM, Hla M, Moor M, et al. Multimodal generative AI for medical image interpretation[J]. *Nature*, 2025, 639(8056): 888-896.
- [12] Sanchez T, esteban O, Gomez Y, et al. FetMRQC: a robust quality control system for multi-centric fetal brain MRI[J]. *Med Image Anal*, 2024, 97: 103282.
- [13] Samani ZR, Alappatt JA, Parker D, et al. QC-automator: deep learning-based automated quality control for diffusion mr images[J]. *Front Neurosci*, 2020, 13: 1456.
- [14] Ye M, Cao Z, Zhu Z, et al. Integrating quantitative DCE-MRI parameters and radiomic features for improved IDH mutation prediction in gliomas[J]. *Front Oncol*, 2025, 15: 1530144.
- [15] Yang H, Zhu Z, Long C, et al. Quantitative and qualitative parameters of DCE-MRI predict CDKN2A/B homozygous deletion in gliomas[J]. *Acad Radiol*, 2024, 31(8): 3355-3365.
- [16] Chen S, Xu Y, Ye M, et al. Predicting MGMT promoter methylation in diffuse gliomas using deep learning with radiomics[J]. *J Clin Med*, 2022, 11(12): 3445.
- [17] Kotia J, Kotwal A, Bharti R, et al. Few shot learning for medical imaging[M]. Berlin: Springer, 2020: 107-132.
- [18] Mu D, Bai J, Chen W, et al. Calcium scoring at coronary CT angiography using deep learning[J]. *Radiology*, 2022, 302(2): 309-316.
- [19] Wang Y, Yan F, Lu X, et al. IILS: intelligent imaging layout system for automatic imaging report standardization and intra-interdisciplinary clinical workflow optimization [J]. *EBioMedicine*, 2019, 44: 162-181.
- [20] Alfarghaly O, Khaled R, Elkorany A, et al. Automated radiology report generation using conditioned transformers[J]. *Inform Med Unlocked*, 2021, 24: 100557.
- [21] Zhang S, Xin X, Wang Y, et al. Automated radiological report generation for chest X-rays with weakly-supervised end-to-end deep learning[J]. *arXiv preprint arXiv, 200610347*, 2020.
- [22] Tanno R, Barrett DG, Sellergren A, et al. Collaboration between clinicians and vision-language models in radiology report generation[J]. *Nat Med*, 2025, 31(2): 599-608.
- [23] Xu Q, Xu X, Zhou C, et al. Towards normalized clinical information extraction in Chinese radiology report with large language models[J]. *Expert Syst Appl*, 2025, 271(2): 126585.
- [24] Gou F, Liu J, Xiao C, et al. Research on artificial-intelligence-assisted medicine: a survey on medical artificial intelligence[J]. *Diagnosics*, 2024, 14(14): 1472.
- [25] Alami H, Lehoux P, Denis JL, et al. Organizational readiness for artificial intelligence in health care: insights for decision-making and practice[J]. *J Health Organ Manag*, 2020, 35(1): 106-114.
- [26] 张冰, 汪洋, 邓芮, 等. 阅片质控装置: 中国, CN210627921U[P]. 2020-05-26.
- [27] Zhang Q, Chen Q, Zhou F, et al. Optimizing the quality of emergency head CT imaging: An automated pipeline for correcting head image position[J]. *Radiography*, 2025, 31(1): 241-246.
- [28] Fan Z, Tang J, Chen W, et al. Ai hospital: benchmarking large language models in a multi-agent medical interaction simulator [J]. *arXiv preprint arXiv, 240209742*, 2024.
- [29] Khalifa M, Albadawy M. AI in diagnostic imaging: revolutionising accuracy and efficiency[J]. *Comput Methods Programs Biomed Update*, 2024, 4: 100146.
- [30] Schuppert C, Rospleszcz S, Hirsch JG, et al. Automated image quality assessment for selecting among multiple magnetic resonance image acquisitions in the German National Cohort study [J]. *Sci Rep*, 2023, 13(1): 22745.
- [31] Chen B, Solebo AL, Taylor P. Automated image quality assessment for anterior segment optical coherence tomograph[C]. *Proceedings of the 2023 Proc IEEE Int Symp Biomed Imaging*, 2023.
- [32] Larson DB, Boland GW. Imaging quality control in the era of artificial intelligence[J]. *J Am Coll Radiol*, 2019, 16(9): 1259-1266.
- [33] 萧毅, 王培军, 刘士远. 中国医学影像人工智能的过去, 现在和未来[J]. *中华放射学杂志*, 2024, 58(11): 1359-1364.

(收稿日期: 2025-04-15 修回日期: 2025-08-04)

人工智能在肿瘤影像学中的应用进展

郑孙易, 刘佳鑫, 崔效楠, 王晶, 叶兆祥

【摘要】 人工智能(AI)正逐步重塑肿瘤影像学的诊疗模式。近年来,随着影像组学、深度学习及大模型等技术的迅速发展,AI已在肿瘤的早期筛查、辅助诊断、疗效评估和预后监测等方面展现出广泛的应用前景。影像组学通过高通量特征提取与建模,深入挖掘CT、MRI及PET等传统影像中潜在的生物学信息,可实现对多种肿瘤的无创诊断、风险分层及预后预测。深度学习依托端到端建模机制,进一步提升了乳腺癌、胰腺癌等病灶的识别率,并在肿瘤免疫治疗反应评估及动态监测中展现出优越性能。大模型则凭借其在复杂文本理解与跨模态数据整合方面的优势,正逐步应用于结构化诊断报告生成及纵向影像随访监测中,助力实现肿瘤影像解读与临床决策的深度融合。然而,AI在肿瘤影像领域的临床转化仍面临数据共享受限、模型可解释性不足及伦理监管缺位等挑战。未来,人工智能将进一步拓展在肿瘤影像教育培训、科研创新及远程诊断等场景中的应用,推动学术资源共享与技术普惠化落地,加速实现肿瘤影像学的智能化、个体化与精准化转型。

【关键词】 人工智能;肿瘤影像;影像组学;深度学习;大模型

【中图分类号】 R73;TP18 **【文献标识码】** A **【文章编号】** 1000-0313(2025)09-1093-05

DOI:10.13609/j.cnki.1000-0313.2025.09.004

开放科学(资源服务)标识码(OSID):



癌症是全球主要的致死原因之一,严重威胁人类生命健康^[1]。医学影像在肿瘤的早期筛查、辅助诊断、疗效评估和预后监测等全流程管理中发挥着至关重要的作用。然而,传统影像学分析方法在实际应用中仍面临诸多挑战。一方面,肿瘤影像的复杂性和异质性对放射科医师的专业判断能力提出了更高要求,尤其是在基层或医疗资源有限地区,诊断结果高度依赖医生的个人经验与技能,难以保证判读的准确性和稳定性;另一方面,基于主观经验的影像判读模式存在较大的个体差异,限制了诊断结果的标准化与可重复性^[2]。此外,随着医学影像数据量的迅速增长,放射科医师的工作负荷不断加重,进而影响其判读的时效性和准确性,导致临床决策延迟或偏差。人工智能(artificial intelligence, AI)作为计算机科学的重要分支,旨在通过模拟人类学习、推理与决策过程,赋予机器自主处理复杂任务的能力。近年来,随着硬件算力的提升和大数据分析技术的发展,AI在医疗领域,特别是在肿瘤

影像学中的应用取得了显著进展^[3-6],其核心技术涵盖影像组学、深度学习和大模型算法等。通过对海量医学影像数据的深度解析,AI在病灶识别、辅助诊断、治疗反应预测和病情动态监测等方面展现出变革性潜力,正在逐步重塑肿瘤影像学的传统诊疗模式。

人工智能在肿瘤影像学中的应用

影像组学一词最早由 Robert Gillies 教授团队于 2010 年提出,并在 2012 年由 Philippe Lambin 教授团队进一步完善^[7,8]。他们指出,医学影像中蕴藏着大量潜在信息,难以通过传统人工阅片方式被完全识别和利用,但可借助高通量计算方法实现系统提取与深入分析^[8]。作为 AI 在医学影像分析中的关键技术,影像组学的基本流程包括图像标准化、感兴趣区分割、高维特征提取与筛选、模型构建与验证等环节。在图像获取阶段,研究人员通常使用 CT、MRI 或 PET 等设备采集高质量图像,并通过标准化处理降低因不同设备参数差异造成的干扰。在图像分割阶段,需对病灶感兴趣区(region of interest, ROI)进行精确勾画,常用方法包括手动、半自动和全自动分割技术。图像分割的准确性将直接影响后续特征提取的质量。基于图像分割结果,研究人员可进一步提取高维定量特征,包括形态学、纹理和统计学特征等。随后,通过机器学习算法进行特征筛选与降维,以去除冗余信息,保留最具诊断价值的特征,最终构建预测模型并结合临床数据开展模型验证与优化。

作者单位: 300060 天津,天津医科大学肿瘤医院放射科,国家恶性肿瘤临床医学研究中心,天津市恶性肿瘤临床医学研究中心,药物成药性评价与系统转化全国重点实验室,天津市消化系统肿瘤重点实验室,天津市肿瘤防治重点实验室(郑孙易、刘佳鑫、崔效楠、叶兆祥);301617 天津,天津中医药大学公共卫生与健康科学学院(王晶)

作者简介: 郑孙易(1992—),男,浙江杭州人,博士研究生,助理研究员,主要从事医学人工智能领域研究工作。

通讯作者: 叶兆祥, E-mail: zye@tmu.edu.cn

基金项目: 国家自然科学基金(编号:82302180、82171932);国家重点研发计划(编号:2021YFC2500400、2021YFC2500402);四大慢病重大专项(编号:2024ZD0520000、2024ZD0520002);天津市医学重点学科(专科)建设项目(编号:TJYXZDXK-010A);天津市教委科研项目计划项目(编号:2024KJ182)

目前,影像组学在肿瘤影像领域的研究多基于高分辨率图像(如 CT、MRI 或 PET),并结合传统机器学习模型完成特定任务的建模分析。该方法通常不需要大规模训练数据,适用于样本量有限的小型队列研究,已被广泛应用于肿瘤的辅助诊断、风险分层、疗效评估和预后预测等多个临床环节,初步显示出良好的应用价值与临床转化潜力。在辅助诊断方面,影像组学可有效提高肿瘤的识别率和区分度,特别是在胰腺癌、肺癌及乳腺癌等早期诊断中发挥了关键作用。Mukherjee 等^[9]从分割后的胰腺 CT 图像中筛选出 34 个影像组学特征,使用支持向量机等常见分类器构建模型,成功实现了胰腺导管腺癌的早期识别(AUC=0.98),其诊断效能明显优于放射科医师。在风险分层方面,影像组学可精准预测肿瘤病理特征与转移风险。Cysouw 团队开展了一项前瞻性队列研究,基于 76 例前列腺癌患者术前 PET-CT 图像构建影像组学模型,用以预测淋巴结受累、远处转移、格里森评分及包膜侵犯等状态,五折交叉验证 AUC 分别为 0.86 ± 0.15 、 0.86 ± 0.14 、 0.81 ± 0.16 和 0.76 ± 0.12 ^[10]。在疗效评估方面,影像组学的表现常优于传统的实体瘤疗效评价标准,特别是在胃癌的新辅助治疗中。Xu 等^[11,12]通过对 292 例局部进展期胃癌患者新辅助治疗后的 CT 增强图像进行分析,采用多种不同特征筛选方法和机器学习分类器的交叉组合,构建的最佳预测模型 AUC 高达 0.92。此外,影像组学还能够有效预测肿瘤预后,为临床个体化治疗提供决策依据。一项来自韩国团队的研究表明,影像组学模型预测结果与直肠癌患者总体生存期和无病生存期均显著相关,在外部测试集中表现优异(AUC=0.80),并能有效区分 IIA 期患者中可能从辅助化疗中获益的亚群^[13]。总体而言,影像组学可在无创条件下深度挖掘肿瘤影像中的潜在生物学信息,拓宽传统影像的临床解释维度,进而服务于精准分型与个体化治疗。然而,基于传统机器学习的影像组学依赖于 ROI 的精准分割,通常需要投入大量人工操作,易受限於医师的个人经验与图像质量,可能导致肿瘤特征稳定性差、模型泛化能力弱、跨中心一致性不足等问题,进而限制其大规模临床转化及推广应用。未来,推动标准化工作流程并加强多中心协作验证,将是提升影像组学在肿瘤影像实践中可用性和可靠性的关键方向。

深度学习(deep learning, DL)作为人工智能技术的重要分支,基于深层神经网络,特别是卷积神经网络(convolutional neural networks, CNNs),能够从原始医学影像中自动学习并提取高阶特征,实现端到端建模,无需依赖手工特征工程。这种方法适用于多模态肿瘤影像数据的处理,如 CT、MRI 和 PET 图像的联

合分析,从而提高肿瘤检测、分类及预后预测的准确性。目前,深度学习已在肿瘤早期筛查、病灶检测与分割、疗效评估与监测及图像增强等方面表现出独特优势。基于美国国家肺癌筛查试验 CT 影像数据开发的深度学习模型,能够有效评估肺结节的恶性风险。该模型在丹麦肺癌筛查研究中表现出优异的验证性能(AUC=0.93),其诊断效能与放射科医师相当,且显著优于传统的 Brock 恶性风险模型。在此基础上,研究人员进一步整合既往 CT 时序信息,构建了新的深度学习模型。该模型在跨中心验证中展现出较高的诊断准确性(AUC=0.97),并显著提升了对动态变化肺结节的鉴别能力。这一进展为在肺癌筛查中实施个体化的随访策略提供了坚实的技术支持和临床依据^[14,15]。此外,深度学习模型还能够自动检测和分割肿瘤病灶,尤其在乳腺癌、肺癌、脑肿瘤等领域,展现出高准确度和敏感度。例如,在乳腺癌筛查中,深度学习模型能够有效识别临床漏诊的早期癌灶。研究表明,该模型可使平均敏感度提高 14%,特异度提高 24%,其诊断性能优于乳腺影像诊断医师。此外,该模型在不同筛查人群、设备及影像模态中均表现出良好的泛化能力与稳定性。尤其值得注意的是,该模型在中国低筛查率人群中表现出优异的判别效能,AUC 达到 0.97。同时,在乳腺 X 线摄影和乳腺断层成像中,该模型也实现了高效且稳健的癌症检测,为乳腺癌的早期干预提供了坚实的技术支撑和临床依据^[16,17]。

在疗效评估方面,深度学习通过对肿瘤细微变化的自动捕捉,可以辅助实现精准化治疗监测。Vanguri 等^[18]构建了一种多模态深度学习模型,该模型整合了 CT 影像、组织病理学图像及基因组数据,用于预测晚期非小细胞肺癌患者对免疫治疗的反应。该模型在验证中表现出良好的预测性能(AUC=0.80),显著优于基于单一模态数据构建的模型。此外,深度学习算法在动态病情监测中的应用也日益成熟,通过多时相影像对比分析,可实时追踪肿瘤体积变化及治疗反应状态,为个体化管理或治疗方案的动态调整提供可靠依据^[19,20]。除了的核心诊断与评估任务方面发挥重要作用外,深度学习在肿瘤图像优化与重建等方面也展现出广阔的应用前景。利用深度学习图像重建技术,肝脏 MRI T₂WI 序列的采集时间可缩短至原来的 1/4,且图像质量优于传统方法^[21]。在低剂量 CT 成像中,基于预训练卷积神经网络的去噪模型可显著提升图像清晰度,有望在保障图像整体质量的同时,进一步降低肿瘤患者的扫描辐射剂量^[22]。总体而言,深度学习凭借其强大的自动学习和复杂模式识别能力,正在加速推动肿瘤影像学由传统人工判读向智能化、自动化分析范式转型。尽管如此,深度学习的可解释性

差、计算资源需求高等问题仍是其面临的严峻挑战。未来研究应积极引入如 SHAP 值解释法、梯度加权类激活图等可视化工具,增强模型决策过程的透明度和可信度,进一步拓展深度学习在肿瘤影像中的应用深度和广度。

近年来,基于大规模预训练机制的 AI 大模型在肿瘤影像领域逐步显现出应用潜力。大模型通常基于海量文本或图像文本数据对其进行预训练,具备超参数量、深层次语义解析以及跨任务迁移等优势,能够同时支持图像处理与自然语言生成^[23-25]。在文本处理领域,大语言模型(large language models, LLMs)已被广泛应用于肿瘤影像诊断报告的解析与结构化生成,显著提高了临床工作效率,并增强了报告书写的规范性与一致性。已有研究表明,基于 LLMs 自动生成的¹⁸F-FDG PET/CT 乳腺癌结构化随访报告,在主病灶变化与转移灶识别方面的准确率分别达到 89.6%和 96.3%^[26];在肺癌非结构化 CT 报告处理中,GPT-4 在病灶参数提取、转移病灶识别及疾病进展评估等方面整体优于 ChatGPT,准确率超过 95.0%,进一步证实了大模型在肿瘤影像文本挖掘中的卓越性能^[27]。此外,研究人员将 LLMs 与医学知识图谱相融合,构建了一种知识感知注意力网络,用于从非结构化影像报告中自动提取肝癌临床分期信息。该方法在真实临床数据集上的准确率达到 90.3%,显示出良好的应用潜力^[28]。在图像分析方面,视觉大模型展现出较强的跨模态推理与多时序影像理解能力。GPT-4o 能够基于纵向随访图像对肺结节进行动态监测。研究结果显示,随着随访 CT 图像数量的增加,其预测肺结节恶性概率的 AUC 值从 0.72 提升至 0.92。此外,在结节大小变化的识别能力方面,GPT-4o 的表现与放射科医师相当^[29],显示出其在临床管理中提供高效、可靠影像学依据的潜力。总体而言,大模型在自由文本理解、跨模态融合以及自动化报告生成等方面展现出独特优势,为推动肿瘤影像分析的标准化、精准化和智能化发展提供了重要技术支撑。同时,其在促进医工交叉协作、多源数据融合及产业化应用方面也开辟了新的路径,为肿瘤影像领域的未来发展注入了持续创新动能。

人工智能在肿瘤影像中的应用局限性

尽管人工智能在肿瘤影像学中展现出广阔的应用前景,但其临床推广与落地仍面临多重挑战,主要集中在数据层面、技术层面与制度保障层面。首先,肿瘤影像数据的异质性与隐私保护之间的矛盾亟待解决。我国《个人信息保护法》规定医疗数据必须经过严格的脱敏处理,然而当前尚缺乏完善的数据共享机制,很大程度上限制了模型训练样本的规模与多样性。尤其是在

影像组学与深度学习研究中,由于肿瘤类型多样、病灶边界模糊、成像特征复杂,基于小样本的建模容易导致泛化能力下降,从而影响模型在真实临床环境中的可靠性。其次,AI 技术门槛较高且模型可解释性不足,制约了其在临床科室的广泛采纳。深度学习与大模型虽能实现高准确率预测,但其"黑箱"属性令医生难以信任。目前已有梯度加权类激活图等可视化方法用于量化关键特征的贡献度,以增强模型透明度^[30],但在肿瘤影像诊断中尚未形成标准化流程,也缺乏针对不同肿瘤类型的系统性验证。此外,大模型对计算资源的依赖也阻碍了其在基层医院或资源有限地区的部署,成为制约技术公平可及的重要瓶颈。同时,影像标准不统一与数据预处理流程不规范亦对模型稳定性构成影响。由于不同医疗机构间影像设备参数和扫描协议存在较大差异,尤其在多中心肿瘤影像联合分析中,模型易出现性能波动,严重影响其跨机构应用的一致性与可靠性。因此,亟需建立统一的图像采集标准与规范化预处理流程,如 N4 偏置场校正等,以提升模型在肿瘤影像任务中的鲁棒性^[31]。最后,伦理与法律规范尚待完善。随着 AI 系统在肿瘤诊疗流程中的角色日益增强,其参与决策若出现误诊或漏诊时,相关责任归属尚不明确。应尽快构建涵盖数据使用、模型训练、临床部署与责任划分等方面的法律法规体系,为 AI 在肿瘤影像中的安全合规应用提供制度保障。综上所述,推动人工智能在肿瘤影像学中的规范应用,需从数据共享机制、模型可解释性提升、跨中心适应性优化与伦理法律体系建设等多个维度协同推进,以实现从科研探索向临床实践的有效转化。

未来展望

未来,人工智能将在医学教育、科研创新与临床实践等多维度持续拓展应用边界。在教育方面,AI 可通过构建基于大模型驱动的虚拟肿瘤影像案例库,辅助医学生识别不同分型、不同阶段或罕见肿瘤病变的影像特征,提升诊断思维与影像判读能力。目前,上海交通大学医学院已开设"AI+微专业"课程,通过虚拟病例库和智能评估系统,强化学生的影像分析与诊断能力。在科研方面,AI 正加速推动肿瘤影像研究范式由传统经验驱动向数据驱动转变。基于 CT 影像构建的 EGFR 突变预测模型,正在加速影像生物标志物的发现,并推动靶向药的研发进程^[32-33]。此外,先进成像技术(如 5T MRI、能谱 CT)与 AI 联合,将推动肿瘤组织微观结构的解析精度,助力精准影像分型与肿瘤异质性研究^[34]。与此同时,诸如 DeepSeek、文心一言、通义千问等具备中文处理优势的大语言模型,在肿瘤影像领域亦可为临床医生提供高效的文本支持工具,

尤其在科研资料整理、课题申报书辅助撰写等环节展现出良好的应用潜力。在临床实践方面, AI 将有望进一步拓展在肿瘤影像中的角色, 从初筛识别延伸至纵向随访与远程辅助诊断。例如, 通过连续随访 CT 影像构建纵向 AI 分析模型, 可用于预测肿瘤复发风险、疗效波动及治疗调整窗口; 同时, 大模型辅助结构化报告生成与远程质控, 将有望缓解基层医院肿瘤影像诊断人力不足的问题。AI 技术的广泛应用正推动放射科医师从传统的影像解读者向 AI 协同决策者转型, 这一变革要求医师掌握基本的 AI 算法原理, 理解其技术优势和局限性, 具备识别潜在误判风险的能力, 并参与多学科团队合作, 优化 AI 工具的设计, 共同推进轻量化模型架构及硬件加速技术的应用。同时, 医师还应注重伦理审查与临床决策的融合, 建立算法偏见监测机制, 避免因数据偏倚导致的误诊, 并将 AI 输出结果与患者个体特征深度融合, 以保障临床决策的精准性与人性化, 从而全面提升 AI 在临床中的可接受性与信任度。

综上所述, 人工智能正深度融入肿瘤影像学的多个关键环节, 其在早期筛查、精准分型、疗效预测和预后评估等方面的临床应用价值已逐步显现。尽管当前 AI 技术在数据壁垒、模型可解释性和伦理法律法规等方面仍面临诸多挑战, 但其由“辅助工具”向“协同决策伙伴”演进的趋势已不可逆转。未来, 相关部门应进一步强化多学科协作机制, 推动高水平复合型人才培养, 加快技术创新与临床融合步伐, 并通过政策引导和法规完善, 明确 AI 在医疗场景中的责任边界, 助力人工智能在肿瘤影像领域实现高质量、可持续发展。

参考文献:

- [1] Siegel RL, Kratzer TB, Giaquinto AN, et al. Cancer Statistics, 2025 [J]. CA Cancer J Clin, 2025, 75(1): 10-45.
- [2] Abujudeh HH, Boland GW, Kaewlai R, et al. Abdominal and pelvic computed tomography (CT) interpretation: discrepancy rates among experienced radiologists [J]. Eur Radiol, 2010, 20(8): 1952-1957.
- [3] Zheng S, Liu J, Xie J, et al. Differentiating high-grade patterns and predominant subtypes for IASLC grading in invasive pulmonary adenocarcinoma using radiomics and clinical-semantic features [J]. Cancer Imaging, 2025, 25(1): 42.
- [4] Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology [J]. Nat Rev Cancer, 2018, 18(8): 500-510.
- [5] 褚晓彤, 付宇, 郑爽, 等. 人工智能在原发性肝癌影像学的机遇和挑战 [J]. 放射学实践, 2024, 39(9): 1244-1249.
- [6] 王葵, 刘丽娟, 金观桥. 人工智能在低剂量螺旋 CT 早期肺癌筛查中的应用进展 [J]. 放射学实践, 2025, 40(5): 672-676.
- [7] Gillies RJ, Anderson AR, Gatenby RA, et al. The biology underlying molecular imaging in oncology: from genome to anatomy and back again [J]. Clin Radiol, 2010, 65(7): 517-521.
- [8] Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis [J]. Eur J Cancer, 2012, 48(4): 441-446.
- [9] Mukherjee S, Patra A, Khasawneh H, et al. Radiomics-based machine-learning models can detect pancreatic cancer on pre-diagnostic computed tomography scans at a substantial lead time before clinical diagnosis [J]. Gastroenterology, 2022, 163(5): 1435-1446. e3.
- [10] Cysouw MCF, Jansen BHE, Van DE Brug T, et al. Machine learning-based analysis of ¹⁸F-DCFPyL PET radiomics for risk stratification in primary prostate cancer [J]. Eur J Nucl Med Mol Imaging, 2021, 48(2): 340-349.
- [11] Xu Q, Sun Z, Li X, et al. Advanced gastric cancer: CT radiomics prediction and early detection of downstaging with neoadjuvant chemotherapy [J]. Eur Radiol, 2021, 31(11): 8765-8774.
- [12] Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (Version 1.1) [J]. Eur J Cancer, 2009, 45(2): 228-247.
- [13] Woerner J, Nam Y, Jung SH, et al. Leveraging automated machine learning to predict colon cancer prognosis from clinical features and risk groups: a retrospective cohort study [J]. Eur J Surg Oncol, 2025, 51(9): 110194.
- [14] Venkadesh KV, Setio AAA, Schreuder A, et al. Deep learning for malignancy risk estimation of pulmonary nodules detected at low-dose screening CT [J]. Radiology, 2021, 300(2): 438-447.
- [15] Venkadesh KV, Aleef TA, Scholten ET, et al. Prior CT improves deep learning for malignancy risk estimation of screening-detected pulmonary nodules [J]. Radiology, 2023, 308(2): e223308.
- [16] Lotter W, Diab AR, Haslam B, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach [J]. Nat Med, 2021, 27(2): 244-249.
- [17] Camurdan O, Tanyel T, Cerecki EA, et al. Annotation-efficient, patch-based, explainable deep learning using curriculum method for breast cancer detection in screening mammography [J]. Insights Imaging, 2025, 16(1): 60.
- [18] Vanguri RS, Luo J, Aukerman AT, et al. Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer [J]. Nat Cancer, 2022, 3(10): 1151-1164.
- [19] Tao G, Zhu L, Chen Q, et al. Prediction of future imagery of lung nodule as growth modeling with follow-up computed tomography scans using deep learning: a retrospective cohort study [J]. Transl Lung Cancer Res, 2022, 11(2): 250-262.
- [20] Yang N, Yue HL, Zhang BH, et al. Predicting pathological response to neoadjuvant or conversion chemioimmunotherapy in stage IB-III non-small cell lung cancer patients using radiomic features [J]. Thorac Cancer, 2023, 14(28): 2869-2876.
- [21] Shanbhogue K, Tong A, Smerek AP, et al. Accelerated single-shot T₂-weighted fat-suppressed (FS) MRI of the liver with deep learning-based image reconstruction: qualitative and quantitative comparison of image quality with conventional T₂-weighted FS sequence [J]. Eur Radiol, 2021, 31(11): 8447-8457.
- [22] Han M, Shim H, Baek J. Low-dose CT denoising via convolutional neural network with an observer loss function [J]. Med Phys, 2021, 48(10): 5727-5742.

- [23] Gertz RJ, Dratsch T, Bunck AC, et al. Potential of GPT-4 for detecting errors in radiology reports: implications for reporting accuracy[J]. Radiology, 2024, 311(1): e232714.
- [24] Zhou Y, Ong H, Kennedy P, et al. Evaluating GPT-V4 (GPT-4 with vision) on detection of radiologic findings on chest radiographs[J]. Radiology, 2024, 311(2): e233270.
- [25] Sun Z, Ong H, Kennedy P, et al. Evaluating GPT4 on impressions generation in radiology reports [J]. Radiology, 2023, 307(5): e231259.
- [26] Chen K, Xu W, Li X. The potential of gemini and GPTs for structured report generation based on free-text ¹⁸F-FDG PET/CT breast cancer reports[J]. Acad Radiol, 2025, 32(2): 624-633.
- [27] Fink MA, Bischoff A, Fink CA, et al. Potential of ChatGPT and GPT-4 for data mining of free-text CT reports on lung cancer [J]. Radiology, 2023, 308(3): e231362.
- [28] Li Y, Zheng X, Li J, et al. Lkan: LLM-based knowledge-aware attention Network for clinical staging of liver cancer[J]. IEEE J Biomed Health Inform, 2025, 29(4): 3007-3020.
- [29] Mao Y, Xu N, Wu Y, et al. Assessments of lung nodules by an artificial intelligence chatbot using longitudinal CT images[J]. Cell Rep Med, 2025, 6(3): 101988.
- [30] Song D, Yao J, Jiang Y, et al. A new xAI framework with feature explainability for tumors decision-making in ultrasound data: comparing with grad-CAM[J]. Comput Methods Programs Biomed, 2023, 235: 107527.
- [31] 王维. 基于 MRI 的影像组学对椎体转移瘤的原发肿瘤溯源价值的研究[D]. 辽宁: 大连医科大学, 2022.
- [32] Wang S, Yu H, Gan Y, et al. Mining whole-lung information by artificial intelligence for predicting EGFR genotype and targeted therapy response in lung cancer: a multicohort study[J]. Lancet Digit Health, 2022, 4(5): e309-e319.
- [33] Li T, Shetty S, Kamath A, et al. CancerGPT for few shot drug pair synergy prediction using large pretrained language models [J]. NPJ Digit Med, 2024, 7(1): 40.
- [34] Guo Y, Lin L, Zhao S, et al. Myocardial fibrosis assessment at 3T versus 5T myocardial late gadolinium enhancement MRI: early results[J]. Radiology, 2024, 313(2): e233424.

(收稿日期: 2025-03-26 修回日期: 2025-08-05)

《请您诊断》栏目征文启事

《请您诊断》是本刊 2007 年新开辟的栏目, 该栏目以临床上少见或容易误诊的病例为素材, 杂志在刊载答案的同时配发专家点评, 以帮助影像医生更好地理解相关影像知识, 提高诊断水平。栏目开办 14 年来受到广大读者欢迎。《请您诊断》栏目荣获第八届湖北精品医学期刊“特色栏目奖”。

本栏目欢迎广大读者踊跃投稿, 并积极参与《请您诊断》有奖活动, 稿件一经采用稿酬从优。

《请您诊断》来稿格式要求: ①来稿分两部分刊出, 第一部分为病例资料和图片; 第二部分为全文, 即病例完整资料(包括病例资料、影像学表现、图片及详细图片说明、讨论等); ②来稿应提供详细的病例资料, 包括病史、体检资料、影像学检查及实验室检查资料; ③来稿应提供具有典型性、代表性的图片, 包括横向图片(X 线、CT 或 MRI 等不同检查方法得到的影像资料, 或某一检查方法的详细图片, 如 CT 平扫和增强扫描图片)和纵向图片(同一患者在治疗前后的动态影像资料, 最好附上病理图片), 每帧图片均需详细的图片说明, 包括扫描参数、序列、征象等, 病变部位请用箭头标明。

具体格式要求请参见本刊(一个完整病例的第一部分请参见本刊正文首页, 第二部分请参见 2 个月后的杂志最后一页, 如第一部分问题在 1 期杂志正文首页, 第二部分答案则在 3 期杂志正文末页)

栏目主持: 石鹤 联系电话: 027-69378385 15926283035