

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

王诗耕, 刘义军, 童小雨, 范勇, 李贝贝, 王旭, 崔景景, 陈安良

【摘要】 目的:探讨不同等级深度学习图像重建(DLIR)算法对肾上腺肿瘤的检出、组学特征可重复性和组学模型鉴别肿瘤类型效能的影响。方法:回顾性收集 41 例肾上腺功能性腺瘤(FAA)和 46 例肾上腺转移瘤(AM)患者的临床和影像资料。CT 增强扫描完成后,对静脉期的原始数据采用 4 种强度等级(DL1、DL2、DL3、DL4)的 DLIR 算法进行重建。首先采用主、客观指标比较 4 种等级间图像质量的差异;然后使用 Research Portal V1.1 科研平台对各组重建图像上肾上腺肿瘤进行分割并提取 450 个影像组学特征,包括原始图像特征 90 个和拉普拉斯(LoG)滤波后的高阶特征(高斯核:0.5、1.0、1.5、2.0)360 个。采用一致性相关系数(CCC)评估采用不同图像重建等级测量的 FAA 和 AM 组学特征的可重复性。最后,在各组重建图像中采用逐步特征选择策略,筛选出最优特征集并构建鉴别 FAA 和 AM 的组学模型。利用五折交叉验证法验证 4 个组学模型的鉴别效能,利用分层交叉验证法测评 4 个模型的泛化能力。结果:DL2 和 DL3 在肾显示上腺肿瘤的清晰度方面最优,得分为 4(4,5),优于 DL1 相应得分 4(3,5)和 DL4 相应得分 4(3,4),且差异具有统计学意义($F=139.045, P<0.05$)。随着 DLIR 降噪等级的提升,原始特征 CCC 值 >0.85 的个数逐渐减少,DL4 中 FAA 和 AM 特征可重复的比例仅占 39.3%(21/90)和 50.9%(29/90)。组学特征经过 LoG 滤波(高斯核 2.0)处理后,CCC 值 >0.85 的个数增加,DL4 中 FAA 和 AM 特征可重复的比例占 91.1%(82/90)和 93.3%(84/90)。4 个组学模型在测试集中的曲线下面积(AUC)和符合率均 >0.75 ,DeLong 检验显示 AUC 的差异无统计学意义($Z=0.177\sim 1.284, P=0.199\sim 0.859$)。但分层交叉验证显示,DL4 重建图像的泛化能力最弱,AUC 和符合率均 <0.75 。结论:高降噪等级的 DLIR 算法会降低对肾上腺肿瘤显示的清晰度以及组学模型的泛化性。虽然 LoG 滤波器(高斯核:2.0)有助于提升组学特征测量的可重复性,但仍建议在肾上腺影像诊断和组学模型训练时,使用中低降噪等级的 DLIR 图像。

【关键词】 体层摄影术;X 线计算机;影像组学;深度学习;重建算法;肾上腺肿瘤;可重复性

【中图分类号】 R814.42;R736.6 **【文献标识码】** A **【文章编号】** 1000-0313(2024)08-1081-08

DOI:10.13609/j.cnki.1000-0313.2024.08.015

开放科学(资源服务)标识码(OSID):



The effect of deep-learning image reconstruction algorithm on the visualization and classification of adrenal tumors WANG Shi-geng, LIU Yi-jun, TONG Xiao-yu, et al. Department of Radiology, the First Affiliated Hospital of Dalian Medical University, Liaoning 116011, China

【Abstract】 Objective: To observe the impact of different levels of deep-learning image reconstruction (DLIR) algorithm on the conspicuity of adrenal tumors, the reproducibility of radiomics features, and the performance of radiomics models in differentiating tumor types. **Methods:** The clinical and CT data of forty-one patients with functioning adrenocortical adenoma (FAA) and forty-six patients with adrenal metastases (AM) were collected retrospectively. After completion of CT enhanced scans, the raw data at venous phase was reconstructed using four levels of DLIR (DL1, DL2, DL3 and DL4) respectively. The image quality of different reconstruction levels was compared first using subjective and objective evaluation. Adrenal tumor segmentation and 450 radiomics features extraction in each group of reconstructed images were established by the Research Portal V1.1. Radiomics features including 90 features from original image features as well as 360 features from higher-order features after Laplacian of Gaussian (LoG) filtering (with Gaussian kernels of 0.5, 1.0, 1.5 and 2.0, respec-

作者单位:116011 辽宁大连,大连医科大学附属第一医院放射科(王诗耕、刘义军、童小雨、范勇、李贝贝、王旭、陈安良); 100094 北京,联影智能医疗科技(北京)有限公司(崔景景)

作者简介:王诗耕(1997-),男,河北张家口人,硕士研究生,主要从事医学影像技术工作。

通讯作者:刘义军, E-mail: yijunliu1965@126.com

tively). The concordance correlation coefficient (CCC) was used to evaluate the reproducibility of the radiomics features of FAA and AM after the DLIR level was increased. Finally, a step-wise feature selection strategy was employed in each group of reconstructed images to select the optimal feature set and construct a radiomics model to differentiate FAAs from AMs. The discrimination performance of the four models was verified using five-fold cross-validation, while their generalizability was assessed using stratified cross-validation. **Results:** Images of DL2 and DL3 were optimal in displaying adrenal lesions with a score of 4 (4,5), which were better than those of DL1 with a corresponding score of 4 (3,5) and DL4 with a corresponding score of 4 (3,4), with statistical difference ($F=139.045, P<0.05$). With the increase in DLIR level, the number of original features with a CCC value >0.85 gradually decreased, and the reproducibility proportion of FAA and AM features in DL4 was only 39.3% (21/90) and 50.9% (29/90), respectively. After the radiomics features were processed with the LoG filter (Gaussian kernel: 2.0), the number of features with CCC values >0.85 increased, with reproducibility proportions in DL4 being 91.1% (82/90) for FAA and 93.3% (84/90) for AM. The area under the curve (AUC) and accuracy of the four radiomics models were all >0.75 in the test sets, with no statistically significant differences according to DeLong test ($Z=0.177\sim 1.284, P=0.199\sim 0.859$). However, stratified cross-validation showed that the generalization ability of images reconstructed with DL4 was the weakest, with both AUC values and accuracy <0.75 . **Conclusion:** High-level noise reduction in DLIR images reduces the conspicuity of adrenal tumors as well as the generalizability of the radiomics models. Although the LoG filter with Gaussian kernel of 2.0 can help for enhancing the reproducibility of radiomics features, it is still recommended to use medium- to low-noise-level DLIR images for adrenal imaging diagnosis and radiomics model training.

【Key words】 Tomography, X-ray computed; Radiomics; Deep learning; Reconstruction algorithm; Adrenal gland neoplasm; Reproducibility

在 CT 扫描偶然发现的肿瘤中以肾上腺肿瘤最为常见,准确判定其良恶性、功能性与非功能性对患者的诊疗和预后至关重要^[1-2]。目前已有多项研究表明,基于 CT 影像组学特征构建的分类模型在鉴别肾上腺肿瘤类型方面具有较高的临床价值^[3-4]。但确保此类模型普适性的关键在于组学特征的可重复性,所以了解 CT 扫描参数对提取的组学特征的影响是当前业内共同关注的问题^[5-6]。随着影像技术的发展,新一代深度学习图像重建(deep-learning image reconstruction, DLIR)算法可在不改变图像纹理的情况下抑制图像噪声^[7],其降噪等级越高降噪效果就越强,而且中高降噪等级的 DLIR 正逐步在人体实质器官成像中得到应用和推广^[8]。但目前鲜有文献报道 DLIR 等级对小病灶及其影像组学特征的影响,所以本研究旨在其它成像参数不变的情况下,观察不同 DLIR 降噪等级是否会影响肾上腺肿瘤的检出、组学特征的可重复性以及模型的效能。

材料与方法

1. 研究对象

回顾性搜集 2022 年 1 月—2023 年 2 月在本院行肾上腺 CT 增强扫描的 90 例患者的病例资料。纳入

标准:①临床资料完整且 CT 图像原始数据(raw data)完整;②经临床和术后病理证实为单发功能性肾上腺腺瘤(functioning adrenocortical adenoma, FAA),或经术后病理或 3 个月以上 CT 影像随访证实为单发肾上腺转移瘤(adrenal metastases, AM);③无碘对比剂过敏、甲状腺亢进和肾功能不全等 CT 增强检查禁忌证。排除标准:①图像质量不佳,有明显呼吸运动伪影($n=2$);②胸腰段椎体内有金属植入物($n=1$)。最终共纳入 87 例患者,男 44 例,女 43 例。

本研究经本院伦理委员会批准(PJ-KS-KY-2023-349),免除患者知情同意书。

2. CT 检查方法

使用联影 uCT760 128 层螺旋 CT 机。患者取仰卧位、足先进、双上肢上举,扫描范围自膈肌上缘至髂前上棘水平。扫描参数:120 kVp,自动管电流,探测器宽度 40 mm,扫描层厚 5 mm,0.5 s/r,螺距 0.9875,显示野(display field of view, DFOV) 280 mm \times 280 mm,矩阵 512 \times 512。使用 Ulrich 双通道高压注射器,以 4.0 mL/s 的流率经右肘正中静脉注入非离子型对比剂碘佛醇(350 mg I/mL),碘剂量 500 mg I/kg,对比剂注射完毕后以相同流率注射生理盐水 30 mL。动脉期扫描采用自动阈值触发(Smart Prep)技术,监

测层面选取膈下腹主动脉段,阈值为 180 HU,动脉期扫描结束后延迟 30 s 行静脉期扫描。扫描完成后,对静脉期原始数据采用 4 种降噪等级(DL1、DL2、DL3、DL4)的 DLIR 算法分别进行图像重建,4 次重建时其它的重建参数一致:重建层厚 1 mm,层间距 1 mm,标准算法(B_SOFT_B)。

3. 图像质量评估

客观评价:在各组重建图像上测量无病变侧肾上腺和腹壁皮下脂肪组织的 CT 值,获得均值和 SD 值。所有数据测量 3 次取平均值。以脂肪组织的 SD 值作为背景噪声,计算肾上腺的 SNR 和 CNR:

$$\text{SNR} = \frac{\text{CT 值}_{\text{肾上腺}}}{\text{SD 值}_{\text{脂肪}}} \quad (1)$$

$$\text{CNR} = \frac{\text{CT 值}_{\text{肾上腺}} - \text{CT 值}_{\text{脂肪}}}{\text{SD 值}_{\text{脂肪}}} \quad (2)$$

主观评价:由 2 位具有 5 年以上工作经验的影像科医师采用盲法对图像噪声和肾上腺肿瘤显示清晰度采用 5 分制进行评分,2 项指标的评分均 ≥ 3 分为图像质量符合临床要求,评分标准见表 1^[9-10]。

表 1 主观评分标准

评分标准	肾上腺显示情况	图像噪声
5	肾上腺边缘锐利,组织结构显示清晰	无明显图像噪声,整体图像质量优
4	肾上腺边缘较锐利,组织结构显示较清晰	图像噪声轻微,整体图像质量好
3	肾上腺边缘和组织结构显示较模糊	图像噪声适中,整体图像质量一般
2	肾上腺边缘和组织结构显示模糊	图像噪声明显,整体图像质量较差
1	肾上腺边缘和组织结构显示不清晰	图像噪声严重,整体图像质量差

4. 病灶分割和特征提取

使用 Research Portal V1.1 (Shanghai United Imaging Intelligence, Co., Ltd.) 科研平台进行影像组学分析。

对肾上腺肿瘤进行手动分割。在所有包含肿瘤层面上沿病灶轮廓逐层勾画 ROI,并生成三维容积感兴趣区(volume of interest, VOI)。首先,在 DL1 重建图像中,由一位研究生和一位具有 10 年工作经验的副主任医师各随机选取 30 例患者的图像,独立进行 ROI 的勾画,以进行观察者间的一致性检验;随后,对其余 27 例患者的图像,由研究生进行 ROI 的勾画和肿瘤分割,并由副主任医师进行核准。在 DL1 重建图像勾画完成后,将 ROI 复制到其它组别重建图像中。

使用平台内嵌入的 Pyradiomics 工具包对各组重建图像上病灶 VOI 进行特征提取,共提取了 450 个组学特征,包括原始特征 90 个(一阶特征 18 个、纹理特征 72 个)和经过拉普拉斯滤波变换(Laplacian of the Gaussian, LoG)后的高阶统计特征 360 个(高斯核为

0.5、1.0、1.5、2.0)。纹理特征中包括灰度共生矩阵(gray level co-occurrence matrix, GLCM)特征 21 个、灰度游程矩阵(gray level run length matrix, GLRLM)特征 16 个、灰度区域矩阵(gray level size zone matrix, GLSZM)特征 16 个、灰度相关矩阵(gray level difference matrix, GLDM)特征 14 个、邻域灰度差分矩阵(neighborhood gray-tone difference matrix, NGTDM)特征 5 个。在分析过程中 4 组重建图像上使用的 ROI 相同,这可以排除肿瘤分割勾画的影响,故本研究中未对形状特征进行分析。

5. 特征筛选和模型建立

基于 Research Portal V1.1 科研平台,利用 Z 分数归一化对不同特征的尺度进行统一,使特征值均分布在 0~1 之间。在 4 组重建图像中以 FAA 和 AM 为标签,参考特征逐步筛选策略,依次采用 3 种方法筛选每组重建图像的最优特征集:递归特征消除(recursive feature elimination, RFE)、最小冗余最大相关(minimum-Redundancy Maximum-Relevancy, mRMR)和最小绝对缩减与选择算子(the least absolute shrinkage and selection operator, LASSO)^[11]。然后,应用随机森林分类器,构建鉴别 FAA 和 AM 的组学模型,并采用五折交叉验证法评估 4 个模型的预测效能。

本研究中进一步采用分层交叉验证法来评估 4 个模型的泛化能力。具体方法:在其中一组图像(如 DL1)上训练模型后,将其应用于由其它 3 组图像(如 DL2、DL3 和 DL4)构成的独立验证集,此验证过程依次轮换进行。绘制各模型对应的 ROC 曲线,并计算曲线下面积(AUC)、敏感度、特异度和符合率。

6. 统计学分析

采用 Medcalc 19.2、SPSS 25.0 和 Python 3.9.0 软件对数据进行统计学分析。符合正态分布的计量资料以均数±标准差表示。非正态分布的计量资料和计数资料以中位数(上、下四分位数)[M(P25, P75)]表示。各组 CT 值、SD 值、SNR 和 CNR 采用 ANOVA 单因素方差分析比较差异。使用 Kappa 检验比较 2 位观察者对 4 组重建图像主观评分的一致性,取 2 名观察者评分标准差较小的进行统计分析。主观评分的组间比较采用 Friedman 检验。计算一致性相关系数(concordance correlation coefficient, CCC)来评估不同等级重建图像上提取组学特征的可重复性以及观察者间的一致性,当 CCC ≥ 0.85 认为可重复性好。进行特征可重复性评估时,以系统默认的重建等级 DL1 作为参考,与其它 3 组数据进行比较。使用独立样本 t 检验或 χ^2 检验比较一般资料的组间差异或 FAA 和 AM 组学特征的 CCC ≥ 0.85 的比例差异。采用 ROC 曲线评估影像组学模型鉴别肾上腺功能性腺瘤与转移

瘤的效能,对各模型 AUC 的比较采用 DeLong 检验。以 $P < 0.05$ 为差异有统计学意义。

结果

1. 基本临床资料的比较

87 例中原发性醛固酮增多症 27 例,皮质醇增多症 14 例,共 41 个 FAA;46 个 AM,原发恶性肿瘤包括肺癌($n=26$)、肝癌($n=13$)和乳腺癌($n=7$)。两组患者基本临床资料的比较见表 2。两组间各项临床资料的差异均无统计学意义($P > 0.05$)。

2. 图像质量

对不同 DL 等级重建图像的客观评价(参数测量)和主观评分结果见表 3 和图 1~2。

客观评价:除肾上腺的 CT 值之外,不同 DLIR 等级之间其它指标测量值的差异均有统计学意义($P < 0.05$);随着 DL 降噪等级升高,肾上腺的 SD 值逐渐降低,SNR 和 CNR 逐渐升高。

主观评分:图像噪声和肿瘤显示清晰度的评分在不同 DLIR 等级之间的差异均有统计学意义($P < 0.05$)。虽然随着 DL 降噪等级的升高,图像噪声呈逐步下降的趋势(得分逐渐升高),但在肾上腺肿瘤显示清晰度方面,DL2 组和 DL3 组中的得分相对较高,

表 2 两组患者的基本临床资料的比较

指标	功能性腺瘤 ($n=41$)	转移瘤 ($n=46$)	统计量	P 值
性别比(男/女)	19/22	25/21	0.556 ^a	0.456
年龄/岁	62.7±10.8	65.6±10.3	-1.293 ^b	0.200
身高/cm	166.8±8.81	168.5±7.47	-0.996 ^b	0.322
体重/kg	69.66±14.23	68.76±10.17	0.341 ^b	0.734
BMI/kg/m ²	24.92±3.95	24.16±3.02	1.010 ^b	0.315
病变侧(R/L)	18/23	24/22	0.594 ^a	0.441

注:R=右侧;L=左侧。^a为 χ^2 值;^b为 t 值。



图 1 患者男,64 岁,左侧 AM(箭)。a)DL1 重建图像,图像噪声适中,呈现出较多细小的亮点或暗点,评分为 3 分,肿瘤边界显示较模糊,评分为 3 分;b)DL2 重建图像,图像噪声轻微,细小的亮点或暗点较少,评分为 4 分,肿瘤边界较清晰,评分为 4 分;c)DL3 重建等级图像,图像噪声轻微,细小的亮点或暗点较少,评分为 4 分,肿瘤边界显示较清晰,评分为 4 分;d)DL4 重建图像,图像上无明显噪声,细小的亮点或暗点明显消失,评分为 5 分,肿瘤边界显示较模糊,评分为 3 分。图 2 患者女,43 岁,左侧 FAA(箭)。a)DL1 重建图像,图像噪声适中,呈现出较多细小的亮点或暗点,评分为 3 分,肿瘤边界显示较模糊,评分为 3 分;b)DL2 重建图像,图像噪声轻微,细小的亮点或暗点较少,评分为 4 分,肿瘤边界显示较清晰,评分为 4 分;c)DL3 重建图像,噪声轻微,细小的亮点或暗点较少,评分为 4 分,肿瘤边界较清晰,评分为 4 分;d)DL4 重建图像,无明显噪声,细小的亮点或暗点明显消失,评分为 5 分,肿瘤边界显示较模糊,评分为 3 分。

表 3 不同 DL 等级重建图像的主观评价结果

指标	DL1	DL2	DL3	DL4	统计量	P 值
客观指标						
CT 值	83.65±14.94	85.41±14.93	85.25±15.06	85.16±15.09	0.260 ^a	0.854
SD 值	20.21±3.27	17.86±3.06	16.42±2.95	11.81±2.29	128.277 ^a	<0.001
SNR	4.25±1.07	4.91±1.22	5.35±1.39	7.46±1.93	80.692 ^a	<0.001
CNR	9.69±2.43	11.64±2.64	12.76±2.96	16.74±3.62	88.709 ^a	<0.001
主观指标						
图像噪声	4(3,4)	4(3,5)	4(4,5)	4(4,5)	88.345 ^b	<0.001
肿瘤显示清晰度	4(3,5)	4(4,5)	4(4,5)	4(3,4)	139.045 ^b	<0.001

注:^a为 F 值;^b为 χ^2 值。

DL4 组最差。

3. 可重复性分析

对两位观察者独立勾画 ROI 所提取的组学特征进行一致性检验,结果显示所提取特征的 CCC 均 > 0.85,平均 0.90 ± 0.04 ,表明所有组学特征在不同观察者间的一致性均较好。

对 DLIR 1 级分别与 DLIR 2~3 级提取的组学特征进行 CCC 分析并绘制热图(图 3)。总体而言,DLIR 等级越高对 FAA 和 AM 组学特征的可重复性影响就越大;但部分一阶特征(P10、P90 等),GLCM 特征中的集群突出(cluster prominence)、集群阴影(cluster shade),GLRLM 特征中的灰度不均匀性(gray level non-uniformity),和 NGTDM 特征中的繁忙度(busyness)、强度(strength)等特征均未受到 DLIR 等级的影响,表现出很好的可重复性(CCC 均 > 0.85)。当原始图像特征经高斯核为 2.0 的 LoG 预处理后,可重复性好的组学特征的比例均超过 90%,对于纹理特征中

灰度(Gray Level)类特征,其可重复性均得到了提升(图 4)。

DLIR 等级为 2 时:FAA 组和 AM 组中,基于原始图像提取的组学特征中可重复性好的比例分别为 97.8%(88/90)和 90.0%(81/90),两组间的差异有统计学意义($\chi^2 = 4.744, P < 0.05$),FAA 组中可重复好的特征占比较大。经 LoG(高斯核 2.0)预处理后,FAA 组和 AM 组中所有特征的可重复性均较好。

DLIR 等级为 3 时:FAA 组和 AM 组的原始图像特征中可重复好的比例分别为 85.6%(77/90)和 73.3%(66/90),两组间的差异有统计学意义($\chi^2 = 4.116, P < 0.05$),FAA 组中可重复好的特征占比较大。经 LoG(高斯核 2.0)预处理后,FAA 组和 AM 组中可重复性好的特征占比分别为 97.8%(88/90)和 98.9%(89/90)。

DLIR 等级为 4 时:FAA 组和 AM 组中,基于原始图像提取的组学特征中可重复性好的比例分别为

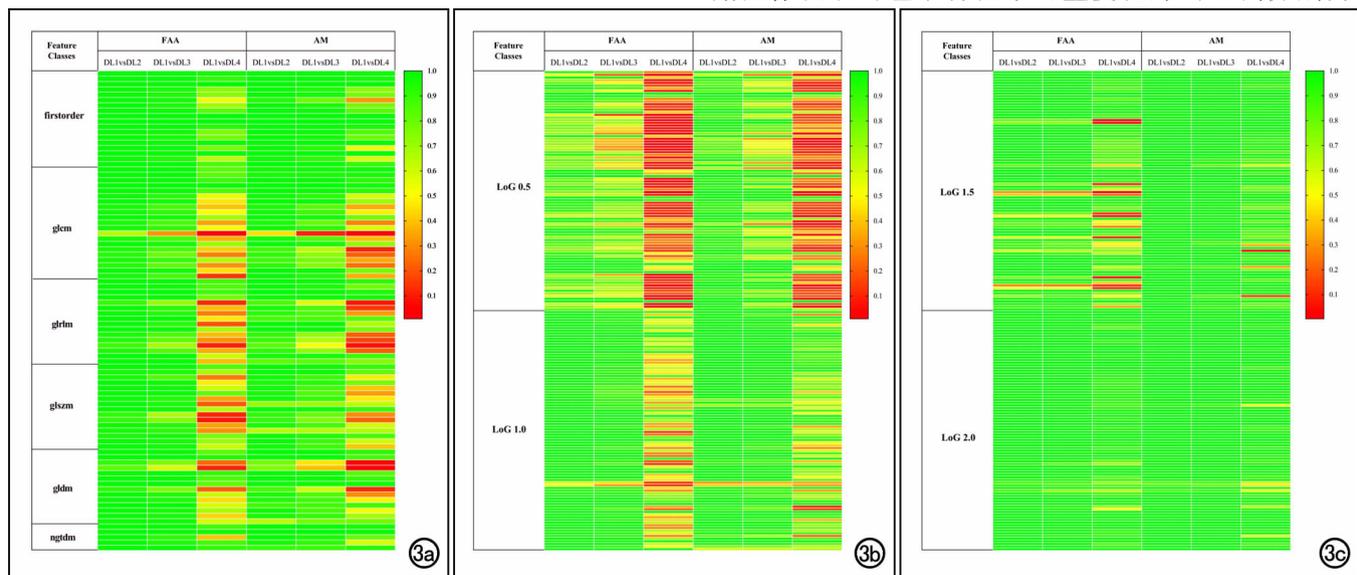


图 3 FAA 组和 AM 组中 DL1 分别与 DL2、DL3 和 DL4 比较的组学特征 CCC 值热图。a)FAA 组和 AM 组中,随着 DLIR 等级的升高各种类别的原始组学特征中 CCC > 0.85 的特征数(绿色单元格)逐渐减少;b)经 LoG(高斯核 0.5 和 1.0)预处理后的组学特征中,随着 DLIR 等级升高 CCC > 0.85 的特征数(绿色单元格)逐渐减少;c)经 LoG(高斯核 1.5 和 2.0)处理后的特征中,随着 DLIR 等级升高 CCC > 0.85 的特征数(绿色单元格)逐渐减少。

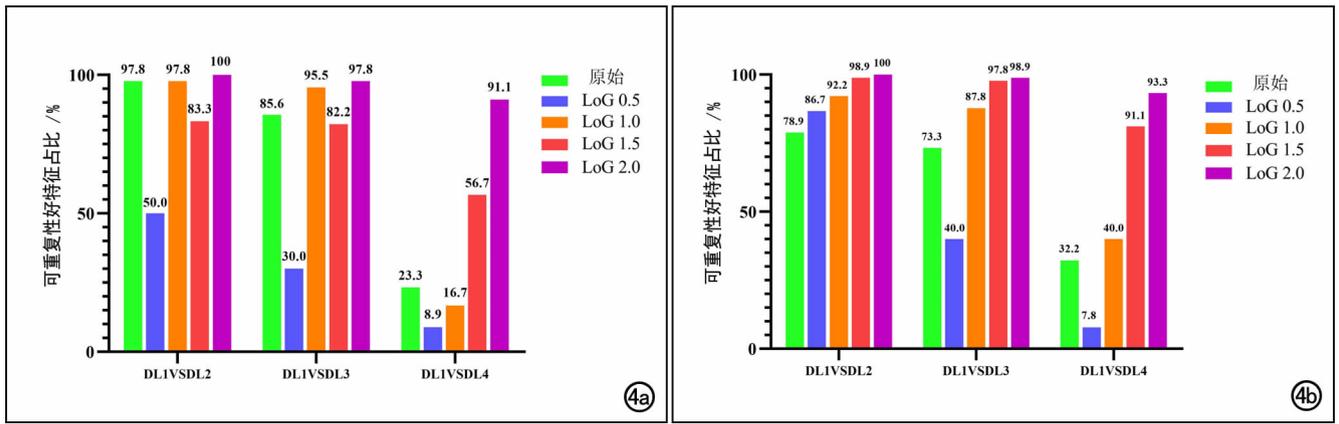


图 4 经 LoG 处理前后组学特征的可重复性变化。a) FAA 组中可重复性好($CCC > 0.85$)的特征占比,随着 DLIR 等级的升高而逐渐降低,但随着 LoG 高斯核的增加而有所提高;b) AM 组中可重复性好的特征占比,随着 DLIR 等级的升高而逐渐降低,但随着 LoG 高斯核的增加而有所提高。

39.3%(21/90)和 50.9%(29/90),两组间差异无统计学意义($\chi^2 = 1.772, P = 0.183$)。经 LoG(高斯核 2.0)预处理后,FAA 组和 AM 组中可重复性好的特征占比分别为 91.1%(82/90)和 93.3%(84/90)。

4. 组学模型效能的评估

DLIR 1~4 级四组重建图像分别筛选出 7、6、6 和 7 个最优组学特征,其中有 3 个经 LoG(高斯核 2.0)预

处理后的组学特征参与了模型 DL1、DL2 和 DL3 的构建,详见表 3。在测试集中,4 组重建图像构建的组学模型的 AUC 分别为 0.835、0.829、0.815 和 0.788。进一步进行 DeLong 检验,结果见表 4,分析结果显示各模型 AUC 的差异均无统计学意义($Z = 0.177 \sim 1.284, P = 0.199 \sim 0.859$)。

表 4 1~4 级 DLIR 重建图像构建的模型中组学特征集

模型的特征集	特征权重*
DL1	
Log2.0-ngtdm-Strengtha	0.210
Log0.5-glrlm-LongRunEmphasis	0.156
Log1.0-glcm-ClusterProminence	0.150
Log2.0-glszm-GrayLevelNonUniformityNormalizeda	0.149
Log2.0-firstorder-Kurtosisa	0.117
Log1.0-glszm-LargeAreaLowGrayLevelZoneEmphasis	0.068
Original-glszm-LargeAreaHighGrayLevelEmphasis	0.017
DL2	
Log0.5-firstorder-Uniformity	0.231
Log2.0-glszm-SizeZoneNonUniformityNormalized	0.188
Log2.0-glszm-GrayLevelNonUniformityNormalizeda	0.182
Log2.0-ngtdm-Strengtha	0.100
Log4.0-glszm-SmallAreaLowGrayLevelZoneEmphasis	0.075
Log2.0-firstorder-Kurtosisa	0.032
DL3	
Log1.0-glcm-ClusterProminence	0.175
Log2.0-glszm-GrayLevelUniformityNormalizeda	0.158
Log2.0-ngtdm-Strengtha	0.147
Log2.0-glszm-SizeZoneNonuniformityNormalized	0.146
Log2.0-firstorder-Kurtosisa	0.077
Log1.0-glszm-SmallAreaEmphasis	0.087
DL4	
Log1.0-glcm-ClusterProminence	0.274
Log2.0-glszm-GrayLevelVariance	0.190
Log2.0-glszm-LowGrayLevelZoneEmphasis	0.095
Log2.0-glszm-LargeAreaLowGrayLevelZoneEmphasis	0.078
Log2.0-glszm-SmallAreaLowGrayLevelZoneEmphasis	0.074
Log0.5-glszm-SmallAreaEmphasis	0.076
Log2.0-glszm-SmallAreaEmphasis	0.062

注:*特征权重取绝对值。

表 5 各组模型鉴别 FAA 与 AM 诊断效能

模型类型	AUC		敏感度		特异度		符合率	
	训练集	测试集	训练集	测试集	训练集	测试集	训练集	测试集
DL1	0.903	0.835	0.875	0.847	0.829	0.804	0.853	0.827
DL2	0.867	0.829	0.847	0.826	0.787	0.780	0.819	0.804
DL3	0.875	0.815	0.809	0.783	0.805	0.805	0.807	0.793
DL4	0.832	0.788	0.761	0.739	0.786	0.756	0.773	0.747

各组模型进行分层交叉验证,结果详见表 5。以 DLIR 4 级重建图像作为独立测试集时,模型的 AUC 和符合率均 <0.75 。以这组图像作为训练集构建组学模型,并将该模型应用到其它等级的图像时,AUC 和符合率也均 <0.75 。

表 6 各组模型交叉验证后的诊断效能

交叉测试集	AUC	敏感度	特异度	符合率
DL1 vs. DL2	0.810	0.804	0.731	0.770
DL1 vs. DL3	0.784	0.761	0.780	0.770
DL1 vs. DL4	0.653	0.847	0.439	0.655
DL2 vs. DL1	0.803	0.826	0.829	0.827
DL2 vs. DL3	0.793	0.761	0.756	0.759
DL2 vs. DL4	0.688	0.826	0.537	0.689
DL3 vs. DL1	0.797	0.739	0.780	0.759
DL3 vs. DL2	0.798	0.826	0.732	0.782
DL3 vs. DL4	0.717	0.630	0.756	0.689
DL4 vs. DL1	0.630	0.826	0.439	0.643
DL4 vs. DL2	0.663	0.848	0.489	0.678
DL4 vs. DL3	0.688	0.586	0.756	0.667

讨论

FAA 是肾上腺最常见的原发良性肿瘤,可分泌激素类物质而干扰体内的激素水平,可导致原发性醛固酮增多症和库欣综合征,必须进行手术治疗^[1]。而且,肾上腺内血供丰富,是恶性肿瘤常见的转移部位。早期 AM 通常无特殊症状,CT 影像学表现也常与肾上腺腺瘤重叠,由此可能影响临床判断,使患者失去最佳的治疗机会^[2]。所以近年来国内外学者引入影像组学分析来评估肾上腺肿瘤的性质,以期提高影像诊断水平,指导临床决策^[3-4]。

影像组学作为一种鉴别肿瘤性质的客观分析方法,可通过各种特征化算法高通量地从图像中提取定量特征,能提供更丰富的客观数据,解析影像与生物信息的关联^[6]。Qi 等^[3]回顾性纳入 191 例功能性腺瘤和 98 例非功能性腺瘤患者,发现影像组学模型鉴别 2 类肿瘤的符合率超过了 80%。Winkelmann 等^[4]利用部分一阶特征构建了鉴别肾上腺转移瘤和腺瘤的组学模型,AUC 达到 0.88。Pan 等^[12]在外部验证集中发现,训练完成的组学模型在鉴别肾上腺腺瘤的效能方面相对训练集有所下降,这可能是由于数据集间成像参数的改变影响了组学特征的可重复性,进而降低了模型的诊断效能。所以组学特征的提取在不同成像参数下的可重复性备受关注,选择稳定的特征对模型的

建立和临床应用具有重要意义^[5,13]。目前,大多数学者已针对常规重建参数是否会削弱组学特征的稳定性进行了充分研究^[14-15],但少有学者对新一代图像重建算法 DLIR 进行探讨。而且 DLIR 技术在人体实质器官的影像学检查中已展现出显著的优势,中高等级的 DLIR 算法已得到了推广^[8],但等级提升对小病灶的显示及其组学特征的影响程度尚未可知。所以本文旨在探究不同等级的 DLIR 算法是否能提供足够高的图像质量来观察肾上腺病灶,是否会对 FAA 和 AM 组学特征的可重复性以及组学模型的鉴别能力产生影响。

本研究结果显示,DLIR 等级越高降噪能力就越强,其中 DL2 和 DL3 重建图像能在低噪声下清晰地显示肾上腺肿瘤的边界。但 DL4 重建图像出现了失真化,降低了观察者对瘤内低对比度区域的辨识能力。而且等级越高的 DLIR 图像上提取的组学特征的可重复性就越差,其中 AM 受到的影响较 FAA 更显著。DLIR 是基于深度神经网络完成的图像重建,在高降噪等级下,因训练集中样本数量有限,可能导致病灶辨识度的降低^[16]。具体表现在图像质量方面,就是 DL4 图像虽然噪声大幅度减少,但图像出现特有的失真现象使肾上腺肿瘤边缘清晰度下降;反映在影像组学分析方面,主要就是纹理特征中的灰度类(gray level)特征的可重复性下降。DLIR 算法还具有非线性的特点,这会使与图像空间分辨率相关的组学特征在一定程度上依赖于组织的对比度^[17]。本研究中纳入的 FAA 病灶内大多数含有脂肪成分,与均匀组织之间的对比度稍高于转移瘤,因此相对而言,AM 组学特征中可重复性好的特征占比要低于 FAA,尤其是纹理特征中的反差(inverse difference)特征,而且 DLIR 等级越高,影响程度就越大。当然,DLIR 算法可在不影响解剖和病理结构显示的情况下抑制图像噪声^[7],所以对于一些原始特征而言 IR 等级的变化并没有太大影响。如一阶特征中反映像素强度的 percentile(P10)、P90 等^[4],GLRLM 特征中反映灰度不均匀性(Gray Level Non-Uniformity)以及 NGTDM 特征中反映病灶局部像素灰度与邻近像素间异质性的 strength 等^[18-19]。

本研究中对图像采用高斯核为 2.0 的 LoG 滤波进行预处理后发现,肿瘤组学特征的可重复性得到了

最大幅度的提升。这是因为 LoG 是一种强调灰度变化的滤波增强器,选择较大的高斯核能更好地突出特征尺度,有效地缓解了 DLIR 等级变化对特征的干扰^[20],这对反映肿瘤异质性和灰度变化的特征尤其重要。在训练集中基于 DL4 图像构建组学模型时,特征选择算法自动筛选出高阶特征来构建模型,其中经 LoG(高斯核 2.0)处理后的组学特征占 70%(5/7),且分别使用 DL1、DL2 和 DL3 重建图像进行模型构建时,有 3 个相同并经 LoG 处理后的组学特征均参与了 3 个模型的构建。这从侧面也印证了我们的上述观点,即 DLIR 的降噪等级越高对组学特征的影响就越大;LoG 滤波处理能缓解 DLIR 等级变化对组学特征的影响。

我们还发现,当使用不同 DLIR 等级的图像来训练组学模型时,模型鉴别肾上腺肿瘤类型的能力并不会因 DLIR 等级的变化而受到显著影响。然而,在测评每个模型的泛化能力时,使用降噪等级为 4 的 DLIR 图像作为独立测试集,或者用这类图像训练的组学模型应用于其它等级的图像时,鉴别效能均不理想。因此,我们建议在构建组学模型时尽量避免使用高降噪等级的图像。

本研究的局限性:①因病例收集较困难,本研究中的样本量偏少;②高斯核为 3.0 和 4.0 的 LoG 滤波器对组学特征可重复性的影响程度还有待研究;③本文主要说明 DLIR 等级变化对组学特征及模型效能的影响,所以并未在模型训练中使用多重重建等级图像联合训练的方案。

综上所述,我们在阅读肾上腺 CT 影像和训练组学模型等临床实践中应充分考虑 DLIR 降噪等级的影响。高降噪等级的 DLIR 图像会降低肾上腺肿瘤显示清晰度以及组学模型的泛化性。虽然 LoG(高斯核 2.0)滤波器有助于提升组学特征的可重复性,但我们仍建议在肾上腺 CT 影像诊断和组学模型训练时,使用中低降噪等级的 DLIR 图像。

参考文献:

[1] Alshahrani MA, Bin SM, Alkhunaizan T, et al. Bilateral adrenal abnormalities; imaging review of different entities[J]. *Abdom Radiol (NY)*, 2019, 44(1): 154-179.

[2] Bracci B, de Santis D, Del Gaudio A, et al. Adrenal lesions: a review of imaging[J/OL]. *Diagnostics*, 2022, 12(9): e2171 [Published 2022 Sep 8]. DOI: 10.3390/diagnostics12092171

[3] Qi S, Zuo Y, Chang R, et al. Using CT radiomic features based on machine learning models to subtype adrenal adenoma[J/OL]. *BMC Cancer*, 2023, 23(1): e111 [Published 2023 Jan 31]. DOI: 10.1186/s12885-023-10562-6.

[4] Winkelmann MT, Gassenmaier S, Walter SS, et al. Differentiation of adrenal adenomas from adrenal metastases in single-phased staging dual-energy CT and radiomics[J]. *Diagn Interv Radiol*,

2022, 28(3): 208-216.

- [5] 王旭, 刘义军, 赵明月, 等. 重建矩阵对 CT 影像组学特征稳定性的影响: 体模研究[J]. *中国医学影像技术*, 2022, 38(9): 1412-1416.
- [6] 王旭, 刘磊, 刘义军, 等. 采集模式与能谱成像能对 CT 影像组学特征稳定性的影响: 体模研究[J]. *中国医学影像技术*, 2023, 39(6): 910-914.
- [7] 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.
- [8] 严福华. 深度学习图像重建算法的临床应用和发展前景[J]. *中华放射学杂志*, 2022, 56(11): 1165-1167.
- [9] Bie Y, Yang S, Li X, et al. Impact of deep learning-based image reconstruction on image quality compared with adaptive statistical iterative reconstruction-Veo in renal and adrenal computed tomography[J]. *J Xray Sci Technol*, 2022, 30(3): 409-418.
- [10] 刘娜娜, 吕培杰, 刘星, 等. 深度学习图像重建算法对低辐射剂量下乏血供肝转移瘤 CT 图像质量及检出效能的影响[J]. *中华放射学杂志*, 2022, 56(11): 1175-1181.
- [11] Wang S, Tong X, Fan Y, et al. Combining deep learning and radiomics for automated, objective, comprehensive bone mineral density assessment from low-dose chest computed tomography[J]. *Acad Radiol*, 2023, 31(3): 1180-1188.
- [12] Pan W, Zhang H, Jin S, et al. Development and validation of a clinical-image model for quantitatively distinguishing uncertain lipid-poor adrenal adenomas from nonadenomas[J/OL]. *Front Oncol*, 2022, 12: e902991 [Published 2022 Jul 13]. DOI: 10.3389/fonc.2022.902991.
- [13] 李宁, 张彪, 王紫君, 等. 并行采集技术对颅脑 MRI 影像组学特征稳定性的影响[J]. *放射学实践*, 2023, 38(4): 514-521.
- [14] Michallek F, Genske U, Niehues SM, et al. Deep learning reconstruction improves radiomics feature stability and discriminative power in abdominal CT imaging: a phantom study[J]. *Eur Radiol*, 2022, 32(7): 4587-4595.
- [15] Gao Y, Hua M, Lv J, et al. Reproducibility of radiomic features of pulmonary nodules between low-dose CT and conventional-dose CT[J]. *Quant Imaging Med Surg*, 2022, 12(4): 2368-2377.
- [16] Solomon J, Lyu P, Marin D, et al. Noise and spatial resolution properties of a commercially available deep learning-based CT reconstruction algorithm[J]. *MED PHYS*, 2020, 47(9): 3961-3971.
- [17] Nam JG, Ahn C, Choi H, et al. Image quality of ultralow-dose chest CT using deep learning techniques: potential superiority of vendor-agnostic post-processing over vendor-specific techniques[J]. *Eur Radiol*, 2021, 31(7): 5139-5147.
- [18] Wang H, Lin M, Chen Y, et al. A radiomics model can distinguish solitary pulmonary capillary haemangioma from lung adenocarcinoma[J]. *Interact Cardiothorac Vasc*, 2022, 34(3): 369-377.
- [19] Moawad AW, Ahmed A, Fuentes DT, et al. Machine learning-based texture analysis for differentiation of radiologically indeterminate small adrenal tumors on adrenal protocol CT scans[J]. *Abdom Radiol*, 2021, 46(10): 4853-4863.
- [20] Zhao B. Understanding sources of variation to improve the reproducibility of radiomics[J/OL]. *Front Oncol*, 2021, 11: e633176 [Published 2021 Mar 29]. DOI: 10.3389/fonc.2021.633176.