·临床研究·

基于 CT 的影像组学在肾透明细胞癌与非透明细胞癌鉴别中的应用

朱佩云，吴凡，刘虎，李天平，潘豪，陈小凤

【摘要】目的　探讨基于增强 CT 的影像组学方法在肾透明细胞癌和非透明细胞癌中的鉴别诊断价值。**方法**　收集 2020 年 10 月至 2025 年 4 月期间嘉兴市第二医院、嘉兴市第一医院收治的经病理证实的肾细胞癌患者的临床影像资料。预处理并利用 Borderline-SMOTE 解决分类不平衡问题后，采用基于极端梯度增强树（XGB）、随机森林（RF）、支持向量机（SVM）和 Logistic 回归（LR）的机器学习算法预测肾肿瘤亚型。采用受试者工作特征曲线（ROC）评价模型性能。并将影像组学模型与放射科医生的诊断结果进行比较。**结果**　训练队列包括 169 例患者（143 例透明细胞肾细胞癌，26 例非透明细胞癌），测试队列 129 例患者（97 例透明细胞癌，32 例非透明细胞癌）。在训练集中，采用 RF、SVM 和 LR 分类器建模，显示相较于单期相模型，联合模型曲线下面积（*AUC*）最高；采用 XGB 建模，静脉期（VP）模型及联合模型的 *AUC* 均较高，分别为 0.985（95%*CI*：0.939 ～ 0.987）、0.980（95%*CI*：0.966 ～ 0.991）。在外部验证集中，联合模型明显优于单期相的预测模型，其中 XGB 联合模型的 *AUC* 最高（0.965，95%*CI*：0.919 ～ 0.994），准确度、灵敏度以及特异度分别为 0.907、0.815、0.949。影像组学诊断结果优于放射科医生。**结论**　基于增强 CT 影像组学模型有助于区分肾透明细胞癌和非透明细胞癌。

【关键词】　肾肿瘤；癌；影像组学；机器学习

DOI:10.3969/j.issn.1671-0800.2026.01.008

【中图分类号】　R737.11　**【文献标志码】**　A　**【文章编号】**　1671-0800(2026)01-0032-05

肾细胞癌（renal cell carcinoma，RCC）是泌尿系统第二常见的恶性肿瘤，约占全部癌症的 3%[1]，RCC 分为透明细胞癌和非透明细胞癌，两者在发病率、临床预后和治疗选择方面都表现出特定的特征[2-3]，准确识别 RCC 亚型对于选择合适的治疗方法和改善预后非常重要。尽管临床上可以利用增强方式区分透明细胞型 RCC（clear cell renal cell carcinoma，ccRCC）和非透明细胞型 RCC（non-clear cell renal cell carcinoma，non-ccRCC），但仍存在较高的误诊率。影像组学在肿瘤成像中的应用是目前的研究热点，已成为一种从常规临床成像中获得诊断、预测和预后信息的工具[4]。为了充分挖掘增强 CT 在 RCC 亚型识别中的诊断优势，本研究探讨基于 CT 的影像组学模型在 RCC 亚型鉴别诊断中的应用价值，现报道如下。

基金项目：　嘉兴市科技计划项目（2022AD30024）

作者单位：　314000　嘉兴，嘉兴市第二医院（朱佩云，刘虎，李天平，潘豪，陈小凤）；嘉兴市第一医院（吴凡）

通信作者：　刘虎，Email：sxfyasdf_1@163.com

1　资料与方法

1.1　一般资料　回顾性收集 2020 年 10 月至 2025 年 4 月期间在嘉兴市第二医院（中心 1）、嘉兴市第一医院（中心 2）接受部分或根治性肾切除术并经病理证实的 RCC 患者的临床影像资料。纳入标准：（1）术前进行增强 CT 检查；（2）图像质量清晰、无伪影；（3）术前未接受治疗。排除标准：（1）病理证实为具有混合特征的 RCC，（2）术前接受介入等治疗，（3）同时合并其他部位肿瘤。本研究获得嘉兴市第二医院医学伦理委员会批准，免除知情同意。

1.2　扫描仪器及检查方法　使用机型包括西门子 SOMATOM Definition Flash、SOMATOM Definition AS 及东芝 Aquilion One 320。扫描参数：管电压 120 kV，管电流 250 mA，层厚 5 mm。使用非离子型对比剂（碘海醇，碘含量 320 mg/ml），采用静脉团注对比剂法和团注追踪技术，注射速率为 2.5 ～ 3.0 ml/s，于注射对比剂 30 s、60 s 后进行动脉期、静脉期扫描。

1.3　图像处理及分割　为了减少不同扫描仪间的影响，对图像进行归一化以及灰度离散化处理，确保结

果的稳健性和可重复性。首先对图像强度值进行标准化处理，标准化尺度为 500，然后使用B样条插值法对图像进行重采样，重采样体素间距为 1 mm×1 mm×1 mm，灰度值以 25 个灰度级进行离散化。使用开源软件ITK-SNAP 进行图像分割，首先由 1 名放射科住院医师对肿瘤全层进行逐层手动勾画，勾画过程中尽量包含肿瘤所有区域，避开瘤周邻近血管，以此获得肿瘤的感兴趣区域（region of interest，ROI）。再由另 1 名高年资放射科医生进行复核、修正，若有分歧，协商讨论后得出一致意见。为了评估影像组学特征的再现性，1 个月后随机抽取 30 例患者，由这两名放射科医生独立绘制 ROI。计算每个特征的组内相关系数（intraclass correlation coefficient，ICC），只有ICC值≥0.75 的特征被保留并纳入进行进一步的分析。

1.4　样本扩增及特征提取　为处理训练集中类别不平衡问题，采用Borderline-SMOTE方法对少数类样本进行合成，并用五折交叉验证划分数据。未对验证集进行样本平衡处理，以评估模型的泛化性能。使用开源 PyRadiomics 框架从每位患者的每个模态影像（平扫、动脉期、静脉期）中提取影像组学特征。提取特征包括一阶统计特征、形状特征及基于灰度共生矩阵（GLCM）、灰度游程长度矩阵（GLRLM）、灰度区域大小矩阵（GLSZM）、灰度依赖矩阵（GLDM）、邻域灰度差矩阵（NGTDM）的纹理特征。本研究获取的影像组学特征遵循成像生物标志物标准化倡议（imaging biomarker standardization initiative，IBSI）[5]概述的特征定义。

1.5　特征筛选和模型构建　中心 1 的数据作为训练数据集，中心 2 的数据作为独立测试数据集。使用最小绝对收缩和选择算子（the least absolute shrinkage and selection operator，LASSO）对训练队列进行特征选择，使用极端梯度增强树分类模型（eXtreme gradient boosting，XGB）、随机森林（random forest，RF）、支持向量机（support vector machine，SVM）和 Logistic 回归（logistic regression，LR）4 种分类器来区分 ccRCC 和 non-ccRCC。每一种分类器均得到以下 4 个模型：（1）平扫（PL）；（2）动脉期（arterial phase，AP）；（3）静脉期（venous phase，VP）；（4）平扫+动脉期+静脉期的联合模型。

1.6　放射诊断医师诊断性能　分别邀请 1 位具有 3 年工作经验的放射诊断医师（放射诊断 1）以及另 1 位具有 7 年以上工作经验的医师（放射诊断 2）在对临床、病理信息均不知晓的情况下对 CT 结果进行判读，并与影像组学模型进行比较。

1.7　统计方法　采用 SPSS 26.0 版软件和 Python 3.10 版本进行统计分析。连续变量采用Kolmogorov-Smirnov 检验正态性，符合正态分布的变量进行独立样本 t 检验；否则用 Mann-Whitney U 检验。分类变量用 χ^2 检验或 Fisher 精确检验。通过受试者工作特征曲线（receiver operating characteristic，ROC）曲线下面积（area under curve，*AUC*）、准确性（accuracy，ACC）、敏感性（sensitivity，SEN）、特异性（specificity，SPE）、f1score（F1）。并使用 bootstrap 方法计算 95%置信区间（confidence interval，*CI*）。应用 DeLong 检验比较不同模型 ROC 曲线下面积。$P < 0.05$ 表示差异有统计学意义。

2　结果

2.1　临床资料　中心 1 共纳入 169 例（训练集），其中透明细胞癌 143 例，非透明细胞癌 26 例（包括乳头状肾癌 12 例，嫌色细胞癌 14 例）。中心 2 共纳入 129 例（外部验证集），其中透明细胞癌 97 例，非透明细胞癌 32 例（包括乳头状肾癌 10 例，嫌色细胞癌 22 例）。两组性别、年龄、肿瘤位置及肿瘤最大径差异均无统计学意义（均 $P > 0.05$），两组肿瘤病理亚型差异有统计学意义（$P < 0.05$），见表 1。

2.2　特征选择　从每位患者的每个模态影像（PL、AP、VP）中各自提取了 105 个影像组学特征，共 315 个特征。使用 LASSO 对训练队列进行特征筛选，根

表 1　研究队列的临床信息

指标	训练集（n=169）	外部验证集（n=129）	$\chi^2(t)$值	P值
性别[例（%）]			1.07	＞0.05
男	127（75.15）	90（69.77）		
女	42（24.85）	39（30.23）		
年龄（岁）	62.4±12.1	59.3±11.9	2.04	＞0.05
肿瘤位置[例（%）]			0.50	＞0.05
左肾	96（56.80）	68（52.71）		
右肾	73（43.20）	61（47.29）		
肿瘤病理亚型[例（%）]			4.14	＜0.05
透明细胞癌	143（84.62）	97（75.20）		
非透明细胞癌	26（15.38）	32（24.80）		
肿瘤最大径（mm）	4.23±2.56	3.92±2.12	1.11	＞0.05

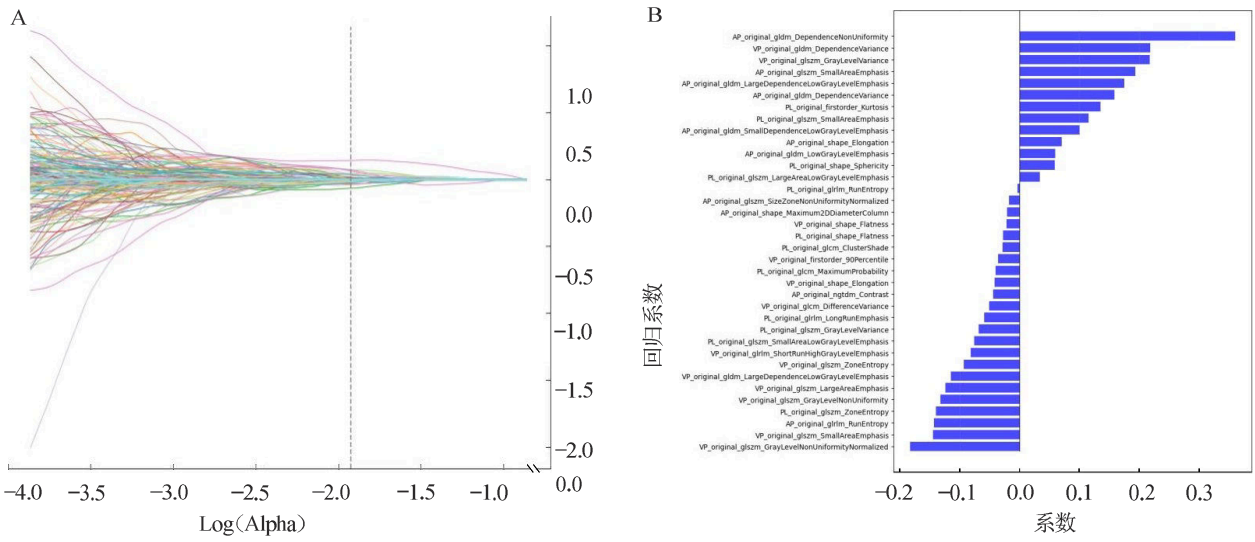
据正则化原则，最后筛选保留 39 个特征，其中 12 个来自 PL，13 个来自 AP，14 个来自 VP，见图 1。

2.3　模型构建及诊断效能　采用 4 种分类器分别基于 CT 平扫、动脉期、静脉期以及平扫+动脉期+静脉期的联合模型进行建模，共构建 16 个模型。在训练集中，采用 RF、SVM 和 LR 分类器建模，联合模型 *AUC* 均表现最优。采用 XGB 建模，VP 模型以及联合模型的 *AUC* 分别为 0.985（95%*CI*：0.939 ～ 0.987）、0.980（95%*CI*：0.966 ～ 0.991），VP 模型的 SEN（0.953）高于联合模型（0.913），而联合模型的特异度、ACC、F1 值均高于 VP 模型，AP 模型的 *AUC* 为 0.966（95%*CI*：0.939 ～ 0.987），略低于 VP 和联合模型，但其 SEN、SPE 及 ACC 高于 VP 和联合模型。在外部验证集中，无论采用何种分类器，联合模型均优于单期相模型，其中 XGB 联合模型的 *AUC* 最高，达 0.965（95%*CI*：0.919 ～ 0.994），其 ACC、SEN 以及特异度分别为 0.907、0.815、0.949。训练集及外部验证集中各模型的 ROC 曲线见图 2 ～ 5。DCA 显示，利用影像组学模型可以获得良好的临床收益，见封三图3。

2.4　不同样本扩增技术的效果评估　为评估不同样本合成技术对模型性能的影响，本研究基于 XGBoost 构建的全期相联合模型，对比了未进行样本平衡、标准 SMOTE 及 ADASYN 三种处理方式。在外部测试集中，未使用样本扩增的模型 *AUC* 为 0.582；而采用样本平衡后，模型性能均显著提升。其中，Borderline-SMOTE 模型表现最佳，其 *AUC* 达 0.965（95% *CI*：0.919 ～ 0.994），见表 2、见封三图 4。

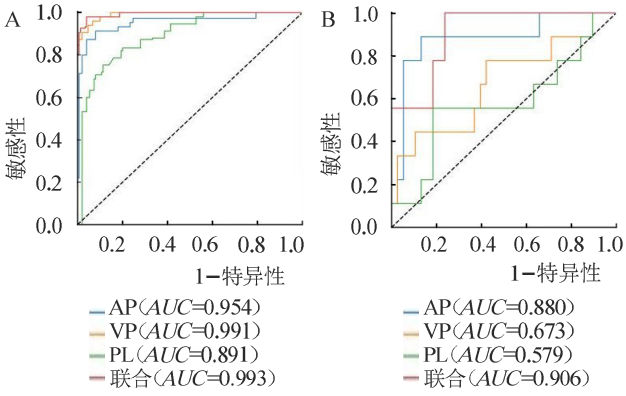
2.5 影像组学模型与放射诊断医生诊断性能比较

诊断医生 1 的 SEN 为 0.833，SPE 为 0.534，*AUC* 值为 0.684；诊断医生 2 的 SEN 为 0.879，SPE 为 0.672，*AUC* 值为 0.776，影像组学模型优于放射科医生，见封三图 5。



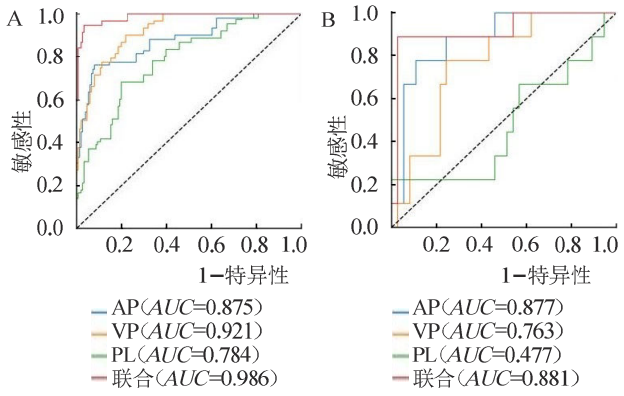
注：A 为 LASSO 模型参数λ调节；B 为特征筛选结果及其权重系数

图 1　LASSO 模型特征筛选



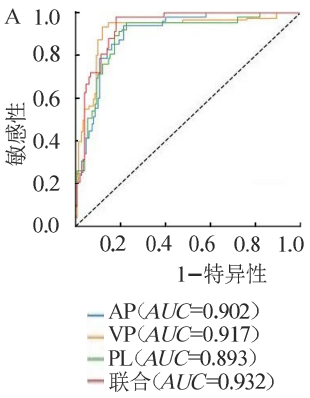
注：A 为训练集，B 为验证集；AP 为动脉期，VP 为静脉期，PL 为平扫

图 2　SVM 模型在训练集和验证集中的 ROC 曲线



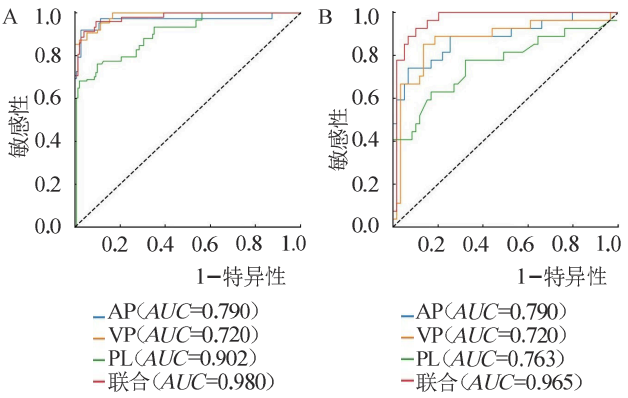
注：A 为训练集，B 为验证集；AP 为动脉期，VP 为静脉期，PL 为平扫

图 3　LR 模型在训练集和验证集中的 ROC 曲线



注：A 为训练集，B 为验证集；AP 为动脉期，VP 为静脉期，PL 为平扫

图 4　RF 模型在训练集和验证集中的 ROC 曲线



注：A 为训练集，B 为验证集；AP 为动脉期，VP 为静脉期，PL 为平扫

图 5　XGB 模型在训练集和验证集中的 ROC 曲线

表 2　XGB 联合模态验证集结果对比

模型	*AUC*（95%*CI*）	敏感性	特异性	准确性	精确率	F1 值
未扩增	0.582（0.357 ～ 0.810）	0.333	0.789	0.702	0.273	0.300
ADASYN	0.860（0.732 ～ 0.957）	0.667	0.816	0.787	0.462	0.545
SMOTE	0.857（0.635 ～ 0.992）	0.667	0.947	0.894	0.750	0.706
Borderline-SMOTE	0.965（0.919 ～ 0.994）	0.815	0.949	0.907	0.880	0.846

3　讨论

根据 2022 年出版的世界卫生组织泌尿生殖系统肿瘤分类肾脏肿瘤的分类标准，RCC 有多种亚型，其中最常见的是 ccRCC，其次是乳头状癌及嫌色细胞癌，分别占 80 ～ 90%、10 ～ 15%、5%[3]。ccRCC 是一种侵袭性和致死性癌，占转移性 RCC 的 94%，5 年生存率为 44%～ 69%，而 non-ccRCC 的 5 年生存率为 78%～ 92%[6]。RCC 的治疗具有挑战性，即使在分期和分级控制后，ccRCC 的预后仍比 non-ccRCC 差[7]。因此术前确定 RCC 亚型很有必要[8]。

在临床实践中，肾脏 CT 增强扫描是 RCC 诊断的主要依据。既往研究表明，CT 增强模式可以区分 ccRCC 和其他亚型[9-10]。但存在不可避免的主观性。近年来，影像组学在肿瘤成像领域的应用日益受到关注，在肾脏病变研究中已被证实可增强诊断、预后评估及预测效能[11-13]。本研究结果表明基于“平扫+动脉期+静脉期”的全期相联合模型优于单期相模型，这可能得益于全期相融合了更完整的肿瘤血流动态信息，能更全面地反映肿瘤内部异质性。Zhang 等[14]研究结果显示在 ccRCC 与 non-ccRCC 的 CT 影像组学模型分类中，全期相模型的性能最好（*AUC* 为 0.89），其 SEN、SPE、ACC 均优于其他单期相模型，与本文结果一致。同时本研究显示影像组学模型的诊断性能优于放射科医师的独立评估，这表明影像组学模型可作为有效的辅助工具，提升医师对常见肾癌亚型的鉴别诊断能力。Wang 等[15]基于增强 CT 建立了三种影像组学模型（RF 模型、SVM 模型和 LR 模型）来区分 ccRCC 和 non-ccRCC，LR 模型表现最优，*AUC* 为 0.906，SEN 达 0.956，SPE 为 0.692，其 SEN 与 SPE 均高于放射科医师（SEN 为 0.85，SPE 为 0.581，*AUC* 为 0.69），该研究虽取得了较好的区分效果，但缺乏外部独立测试。而本文通过外部验证表明了模型具有良好的泛化能力。

放射科医生在临床实践中主要依据动脉期 CT 表现评估肾肿瘤。Li 等[16]选取动脉期纹理特征构建区分 ccRCC 与 non-ccRCC 机器模型取得了较好性能（*AUC*=0.95），但本研究中动脉期影像组学特征并未显著提升肾癌亚型识别的 ACC。另外一项研究[17]仅基于静脉期影像组学特征构建的 XGBoost 模型在测试集上取得了 0.75 的 *AUC*，而动脉期 *AUC* 为 0.67，提示静脉期可能更能反映不同肾癌亚型的关键影像学差异。

本研究存在样本量与类别不平衡问题，可能会影响模型的性能，为此，在训练集中采用 Borderline-SMOTE 方法对少数类样本进行合成扩增。引入样本平衡方法后，模型的性能显著提升。此外，本研究亦对比了标准 SMOTE 与 ADASYN 方法，结果显示 Borderline-SMOTE 模型表现最优（*AUC*=0.965），其

在处理边界样本上更具优势。

值得注意的是，模型在训练集中的表现优于独立测试集，该差异提示模型存在过拟合问题，而使用的交叉验证方法未能完全克服这一局限。此外，样本量有限、类别分布不均衡以及不同中心的 CT 扫描协议差异等因素，均可能影响模型的泛化性能。本研究的训练集与外部独立验证集在肾癌亚型分布上存在统计学差异，这为评估模型的泛化能力提供了一个严格而真实的测试环境，训练集中 ccRCC 的高占比可能使模型对 ccRCC 的特征更敏感，而降低了识别非透明细胞癌亚型的能力；尽管如此，模型在验证集上仍保持了稳健的鉴别效能，证明所提取的影像组学特征捕获了跨中心的、与肾癌亚型生物学基础相关的部分稳定信息。然而，分布差异也警示未来面向临床应用的模型开发必须致力于解决数据不平衡问题，并通过多中心协作构建更具代表性的数据集来提升模型的普适性。尽管存在这些挑战，通过将影像组学模型与放射科医师的诊断经验相结合，有望进一步提升肾癌亚型鉴别诊断的稳定性和可靠性。

本研究有几个局限性：(1)样本量有限且分布不平衡；(2) 部分患者缺乏排泄期扫描，因此本研究未将该期相纳入分析；(3)当前模型未结合临床数据及 CT 参数与影像组学特征进行联合建模；(4) 未对肾细胞癌亚型展开进一步的亚组分析。

综上所述，本文基于增强CT影像特征构建了可靠的预测模型来区分 ccRCC 和 non-ccRCC，与单期相模型相比，全期相联合模型的性能最好。利用这种无创预测模型进行 RCC 亚型鉴别有助于指导治疗决策和精确治疗。

利益冲突　所有作者声明无利益冲突

作者贡献声明　朱佩云：实验操作，论文撰写；朱佩云，李天平，吴凡，潘豪，陈小凤：数据整理，统计学分析；刘虎：研究指导

参　考　文　献

[1] Siegel R L, Miller K D, Jemal A. Cancer statistics, 2018[J]. CA Cancer J Clin, 2018, 68(1): 7-30.

[2] Ljungberg B, Albiges L, Abu-Ghanem Y, et al. European association of urology guidelines on renal cell carcinoma: The 2022 update[J]. Eur Urol, 2022, 82(4): 399-410.

[3] Moch H, Cubilla A L, Humphrey P A, et al. The 2016 WHO classification of tumours of the urinary system and male genital organs: Part A: Renal, penile, and testicular tumours[J]. Eur Urol, 2016, 70(1): 93-105.

[4] Lubner M G, Smith A D, Sandrasegaran K, et al. CT texture analysis: Definitions, applications, biologic correlates, and challenges[J]. Radiographics, 2017, 37(5): 1483-1503.

[5] van Griethuysen J J M, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype[J]. Cancer Res, 2017, 77(21): e104-e107.

[6] Truong L D, Shen S S. Immunohistochemical diagnosis of renal neoplasms[J]. Arch Pathol Lab Med, 2011, 135(1): 92-109.

[7] Keegan K A, Schupp C W, Chamie K, et al. Histopathology of surgically treated renal cell carcinoma: Survival differences by subtype and stage[J]. J Urol, 2012, 188(2): 391-397.

[8] Farber N J, Kim C J, Modi P K, et al. Renal cell carcinoma: The search for a reliable biomarker[J]. Transl Cancer Res, 2017, 6(3): 620-632.

[9] 陆艳艳,陈宏伟,鲍健,等. 双源 CT 对肾透明与非透明细胞癌的鉴别诊断价值[J]. CT 理论与应用研究,2018,27(1): 19-25.

[10] 朱庆强,王中秋,王守安,等. 肾透明细胞癌与少见的肾癌亚型 CT 表现的对照研究[J]. 中华泌尿外科杂志, 2014,35(3): 168-173.

[11] Ursprung S, Beer L, Bruining A, et al. Radiomics of computed tomography and magnetic resonance imaging in renal cell carcinoma-a systematic review and meta-analysis[J]. Eur Radiol, 2020, 30(6): 3558-3566.

[12] Roussel E, Capitanio U, Kutikov A, et al. Novel imaging methods for renal mass characterization: A collaborative review[J]. Eur Urol, 2022, 81(5): 476-488.

[13] Campi R, Stewart G D, Staehler M, et al. Novel liquid biomarkers and innovative imaging for kidney cancer diagnosis: What can be implemented in our practice today? a systematic review of the literature[J]. Eur Urol Oncol, 2021, 4(1): 22-41.

[14] Zhang H J, Yin F, Chen M L, et al. A reliable prediction model for renal cell carcinoma subtype based on radiomic features from 3D multiphase enhanced CT images[J]. J Oncol, 2021, 2021: 6595212.

[15] Wang P, Pei X, Yin X P, et al. Radiomics models based on enhanced computed tomography to distinguish clear cell from non-clear cell renal cell carcinomas[J]. Sci Rep, 2021, 11(1): 13729.

[16] Li Z C, Zhai G T, Zhang J H, et al. Differentiation of clear cell and non-clear cell renal cell carcinomas by all-relevant radiomics features from multiphase CT: A VHL mutation perspective[J]. Eur Radiol, 2019, 29(8): 3996-4007.

[17] Uhlig A, Uhlig J, Leha A, et al. Radiomics and machine learning for renal tumor subtype assessment using multiphase computed tomography in a multicenter setting[J]. Eur Radiol, 2024, 34(10): 6254-6263.

收稿日期：2025-08-26

（本文编辑：吴迪汉）