

Application of Radiomics Machine Learning in Differentiating Gleason Grade Group for Prostate Cancer Based on Biparametric Magnetic Resonance Imaging*

SUN Yu-yue¹, ZHANG Jun-tao², SU Yue-ting³, TIAN Wei-zhong^{3,*}.

1. Graduate School of Dalian Medical University, Dalian 116044, Liaoning Province, China

2. GE Healthcare, GE Healthcare PDX AA team, Shanghai 210000, China

3. Department of Imaging, Taizhou People's Hospital, Taizhou 225300, Jiangsu Province, China

基于双参数MRI影像组学机器学习鉴别前列腺癌Gleason评分分级分组中的应用*

孙于越¹ 张濬韬² 苏月婷³

田为中^{3,*}

1. 大连医科大学研究生院 (辽宁 大连 116044)

2. 通用电气医疗通用电器药业高级应用团队 (上海 210000)

3. 泰州市人民医院影像科 (江苏 泰州 225300)

【摘要】目的 探讨基于双参数MRI影像组学机器学习模型对前列腺癌(PCa) Gleason评分分级分组中的价值。**方法** 回顾性收集经病理证实为PCa的患者138例, 其中GG>2患者83例, GG≤2患者55例。所有病人术前均进行MRI检查。按7:3将患者分为训练集和测试集, 分别用于影像组学模型的机器学习和验证, 采用RF、SVM和XGboost构建3组模型(ADC、T₂WI、ADC+T₂WI), 采用受试者操作特征(ROC)曲线评估各模型鉴别GG≤2与GG>2 PCa的诊断效能。**结果** 测试集中应用SVM算法的T₂WI+ADC模型诊断效能最高, AUC为0.896。其次为T₂WI+ADC中的RF模型, AUC为0.871。在各特征集中, RF和SVM算法的模型的AUC均高于XGboost算法。**结论** 基于双参数MRI影像组学机器学习模型可较好地鉴别GG≤2与GG>2 PCa。

【关键词】 前列腺癌; Gleason评分分级分组; 影像组学; 机器学习

【中图分类号】 R697+.3

【文献标识码】 A

【基金项目】 泰州市科技支撑计划(社会发展)项目(TS201906)

DOI:10.3969/j.issn.1672-5131.2023.08.038

ABSTRACT

Objective To investigate the value of biparametric MRI radiomics machine learning model in Gleason grade group and grouping of prostate cancer (PCa). **Methods** 138 patients with PCa confirmed by pathology were retrospectively collected, including 83 patients with GG > 2 and 55 patients with GG ≤ 2. All patients underwent preoperative biparametric (T₂WI+ADC) MRI examination. Patients were randomly divided into training set and test set at 7:3 for machine learning and validation of imaging radiomics models, respectively. RF, SVM and XGboost were used to construct three groups of models (ADC, T₂WI and ADC+T₂WI). Receiver operating characteristic (ROC) curve was used to evaluate the diagnostic efficiency of each model in distinguishing GG ≤ 2 from GG > 2 PCa. **Results** In the test set, the T₂WI+ADC model using SVM algorithm had the highest diagnostic efficiency, with an AUC of 0.896. Secondly, the AUC of RF model in T₂WI+ADC was 0.871. In each feature set, the AUC of RF and SVM algorithm is higher than XGboost algorithm. **Conclusion** The radiomics machine learning model based on biparametric MRI can distinguish GG ≤ 2 from GG > 2 PCa.

Keywords: Prostate Cancer; Gleason Grade Group; Radiomics; Machine Learning

前列腺癌 (prostate cancer, PCa) 是全球男性中第二常见的癌症^[1]。根据美国癌症协会的最新统计数据, 2022年仅PCa就占了新诊断癌症的近27%, 占男性癌症死亡的11%。PCa的发病率和死亡率在中国迅速上升。由于PCa的预后与其生物学攻击行为密切相关, 因此早期发现和准确的风险分层在确保患者获得最佳结果方面发挥着关键作用^[2]。

由于使用血清中的前列腺特异性抗原水平 (prostate specific antigen, PSA) 进行筛查, 现在通常可以在早期进行诊断和治疗。根据PSA水平、肿瘤局部扩散程度和组织学分化程度 (Gleason评分) 对前列腺癌进行风险分组。如果在分期期间未检测到转移, 且患者年龄、共病和预期寿命符合治疗条件, 则需要行局部治愈性治疗。根据患者的风险分组, 选择根治性前列腺切除术, 经皮放射治疗, 近距离放射治疗或积极监测。如果在初步诊断时已存在转移, 则可采用雄激素剥夺治疗和化疗进行姑息性全身治疗。在晚期和激素难治期, 用嗜骨放射性示踪剂或姑息性放射治疗可以减少骨转移和缓解各自的症状。

目前, Gleason评分 (Gleason score, GS) 系统是世界范围内应用最广泛的评估PCa侵袭性的病理评分系统。GS评分越高, 代表侵袭性越强^[3]。虽然Gleason评分3+4=7分和4+3=7分总分一致, 但患者有不同的预后^[4], 且在积分系统上无法将其区分开来。为了解决上述Gleason评分系统存在的缺陷, 2014年国际泌尿病理协会 (ISUP) 提出Gleason评分分级分组 (Gleason grade group, GG), 根据GS将GG分为五组, GG1 (GS ≤ 6)、GG2 (GS 3+4=7)、GG3 (GS 4+3=7)、GG4 (GS = 4+4、3+5、5+3)、GG5 (GS = 9-10)^[5]。然而, 经直肠超声引导下穿刺活检是一种侵入性外科手术, 前列腺穿刺活检为有创检查, 存在发生活检后血精、血尿和败血症等风险的可能。同时其取材具有随机性, 故存在一定漏检率和误诊率。因此, 迫切需要一种能够准确无创地评估前列腺癌侵袭性替代方法, 这将有助于指导临床治疗。

近年来, 越来越多的兴趣集中在识别成像替代物和开发用于癌症表征的非侵入性诊断工具^[6]。影像组学从医学图像衍生的特征推断肿瘤分子状态, 并在空间和时间上研究肿瘤异质性。磁共振成像 (magnetic resonance imaging, MRI) 作为一种可靠的无创性影像学检查方法, 用于识别和确定前列腺内的病灶区域, 被广泛应用于PCa的诊断^[7]。影像组学分析利用先进的图像分析方法, 快速训练和验证医学成像数据, 基于医学图像的特征进行精确诊断和治疗, 为现代医学提供了强大的工具。目前, 影像组学作为一种新兴的方法, 通过高通量提取丰富的定量特征, 并通过机器学习的方法建立预测模型, 被证明可以有效地为泌尿外科的临床实践提供更多潜在的有用信息。PCa的影像组学已广泛应用于肿瘤识别、分期和预后评估。多项研究表明, 从T₂WI和ADC中提取的特征有助于Gleason评分的分类^[8-10]。本研究基于双参数磁共振成像 (bi-parametric magnetic resonance imaging, bpMRI) 建立影像组学机器学习模型, 鉴别GG ≤ 2与GG > 2的PCa, 协助临床为不同患者制定个性化治疗方案。

1 资料与方法

1.1 研究对象 回顾性分析 2021年1月至2022年5月于泰州市人民医院行前列腺MRI检查

【第一作者】 孙于越, 女, 硕士研究生在读, 主要研究方向: 腹部影像诊断。E-mail: 1109494000@qq.com

【通讯作者】 田为中, 男, 主任医师, 主要研究方向: 腹部影像诊断。E-mail: jstztwz@163.com

的患者病例资料。

纳入标准：根据PI-RADS V2 T₂WI和ADC图像上边界明确的前列腺病变；经超声引导下穿刺活检或手术，病理组织学确诊为PCa的患者；在MRI检查之前未进行前列腺穿刺活检和手术、放射或内分泌治疗。排除标准：确诊前列腺癌以外的肿瘤；患者放置导尿管后；病理信息不全者。

1.2 检查方法 所有患者采用Siemens Skyra 3.0T超导全身磁共振扫描仪，采用标准8通道体部线圈，线圈两侧用腹带固定减少移动伪影的影响病人采取仰卧位，头先进，检查前嘱咐患者适当膀胱，扫描时线圈中心对准耻骨联合。横断面T₂WI参数为重复时间/回波时间(TR/TE)=7500/97ms，视野(FOV)为200mm²×200mm²，矩阵512×512，层厚3.5mm，层间距为1，采集时间为2min 39s。横断面DWI参数为重复时间/回波时间(TR/TE)=5100/64ms，视野(FOV)为308mm²×380mm²，矩阵208×256，层厚4mm，层间距为1，b值=50，800s/mm²，采集时间为1min 57s。

1.3 图像分割及影像组学特征提取 将入组患者的MR平扫图像以DICOM格式从PACS系统导出，再导入ITK-SNAP开源软件(版本3.6.0)，一位有三年MR诊断经验且研究方向为腹部影像诊断的医生手动勾画感兴趣区(region of interest, ROI)，勾画ROI时应尽量避开尿道、出血及钙化(图1)，并由一名有10年腹部影像诊断经验的影像诊断医师进行视觉验证。如医师2对医师1勾画的感兴趣区有质疑，则医师2与医师1协商达成一致结果或由医师2对医师1的勾画结果做出相应修改。如一个患者存在多处病灶且Gleason评分不同时，选择Gleason评分最高且直径最大的病灶层面进行勾画。通过Python(<https://www.python.org>)中的Pyradiomics包提取勾画得到的ROI中的影像组学特征。

1.4 机器学习建模与统计学分析 为保证特征值尺度统一，所有影像组学特征均采用z-score归一化进行标准化，计算公式如下： $y=(x-\mu)/\sigma$ 其中x是特征的原始值， μ 和 σ 分别是x的均值和标准差值，y是变换后的特征值。本研究使用了最大相关-最小冗余算法(maximum correlation minimum redundancy, mRMR)来选择特征，对冗余和不相关的特征进行筛选后用于构建最终的影像组学模型。

采用R软件完成所有影像组学的统计学分析。将数据按7:3随机分成训练组和测试组，采用随机森林(RF)、支持向量机(SVM)和极端梯度提升(XGboost)机器学习算法对上述筛选的特征参数构建模型。共构建3组模型：1)ADC特征集；2)T₂WI特征集；3)T₂WI+ADC特征集。上述机器学习算法的超参数通过10倍交叉验证和启发式算法进行调整。应用受试者工作特征(receiver operating characteristic, ROC)曲线评价测试数据集中不同模型效能，并计算AUC、敏感性、特异性、准确性、假阳性率、假阴性率。

使用SPSS(版本, 22.0)计算和分析临床数据。符合正态分布的计量资料以($\bar{x} \pm s$)表示，2组间比较采用t检验；非正态分布的计量资料以中位数(四分位间距)[M(P25, P75)]表示，2组间比较采用Mann-Whitney U检验。P<0.05为差异有统计学意义。

2 结果

2.1 临床资料 总共有138名患有PCa患者被确定并纳入我们的研究。在138名患者中，83名患者(60.1%)经组织学证实为GG>2，年龄平均为(73.17±6.99)岁，PSA中位数为20.45(10.47, 69.75)ng/mL。55名患者(39.9%)经组织学证实为GG≤2，年龄平均为(71.38±8.03)岁，PSA中位数为9.02(6.60, 13.65)ng/mL。GG>2组的PSA值高于GG≤2组(P<0.05, Z=-5.06)。2组病人年龄的差异均无统计学意义(P>0.05, t=-1.39)。

2.2 机器学习结果 利用mRMR算法对所得影像组学特征进行筛选，基于ADC筛选出12个特征，基于T₂WI筛选出20个特征，基于T₂WI+ADC筛选出17个特征。根据箱线图显示基于ADC、T₂WI和T₂WI+ADC训练组中9种机器学习模型具体表现(图2)。测试集中各模型的预测性能ROC曲线(图3)。在测试集中，T₂WI+ADC特征集中的机器学习性能最好，其中SVM模型的AUC为0.896，准确率为76.2%，优于其他模型。在测试集中，基于T₂WI+ADC特征集影

像组学机器学习模型预测GG≤2与GG>2的准确率、特异度、敏感度、假阳性率及假阴性率见表1。

T₂WI+ADC特征集机器学习模型预测GG≤2与GG>2PCa的校准曲线见(图4)，结果提示预测结果与病理结果一致性良好。决策曲线(图5)表明，使用RF、SVM模型预测GG≤2与GG>2PCa比使用XGboost模型有更高的临床净收益。利用Delong法比较测试集中T₂WI+ADC影像组学机器学习模型之间AUC的差异。XGboost模型与RF、SVM模型预测GG≤2与GG>2的AUC差异有统计学意义(P均<0.05)，RF、SVM模型间预测GG≤2与GG>2的AUC差异无统计学意义(P均>0.05)。

3 讨论

2019年PI-RADS 2.1提出bpMRI概念，以简化前列腺MRI序列^[11]。bpMRI的诊断准确性及效能与包括T₂WI、DWI和动态增强的多参数磁共振成像相当，阅片者间的一致性良好，且无须使用造影剂。然而，尽管高年资影像科医师在检出癌灶和边界方面具有较高的可靠性，但判断PCa的侵袭性仍具有挑战性。在评估潜在病变时，临床医师受到个人知识和既往经验的限制，可能导致不同临床医师之间的差异。基于影像组学方法的机器学习可以提取临床医生视觉分析之外的大量有特色的影像特征，并构建客观的预测模型。

近年来，影像组学应用的增加使得影像学研究能够通过客观和定量的方法预测肿瘤的侵袭性。目前，MRI影像组学已在PCa的诊疗方面表现出巨大优势^[12]。以往研究多集中于区分有临床意义(GS≥7)和无临床意义的PCa(GS<6)，但在临床工作中，PCa(GS3+4)多采用“随诊观察”或“主动监测”等治疗方式。新的分级系统避免了对惰性病例的过度治疗，能更好的指导临床工作。因此鉴别GG≤2与GG>2的PCa，采取正确的治疗方式被认为是至关重要的。

在这项研究中，我们建立了九个基于MRI影像组学机器学习模型来鉴别GG≤2与GG>2PCa。在比较了影像组学中常用的三种分类算法后，我们发现RF模型及SVM模型在判别PCa恶性潜能方面表现出最佳性能。不同的机器学习算法各有优缺点。在研究之前无法预测算法在特定机器学习任务中的性能。大多数以前的影像组学研究使用单一算法进行建模，并且没有说明选择该模型的具体原因。我们的研究表明，在ADC特征集中RF优于SVM和XGboost，在T₂WI、T₂WI+ADC特征集中SVM优于RF和XGboost。RF是一种集成机器学习，它基于决策树方法，可以在不显著增加计算量的情况下提高预测精度。Donisi等^[13]基于公开的数据集，证明了影像组学机器学习模型可以成功鉴别有临床意义的PCa，其中，RF模型获得了最高的准确性。彭涛等^[14]建立基于mpMRI机器学习模型，评价其鉴别临床显著性PCa的诊断价值，结果表明RF模型净收益最高，SVM模型次之。SVM在处理小样本和非线性、高维数据方面有很多优势，Chen等^[15]结合PI-RADS V2.1评分和MR影像组学特征所开发的SVM分类器在诊断前列腺癌侵袭性显示出显著的诊断效能，并表明基于bpMRI和mpMRI的影像组学模型的性能没有显著差异。在本项研究中，通过10倍交叉验证和启发式算法对模型进行训练，最终发现发现RF和SVM表现稳定。由于本研究中分析的样本量较小，GG≤2的比例相对较小，数据分布不平衡可能是导致XGboost分类精度弱于RF、SVM的原因之一。

PCa患者不可避免地会合并一些前列腺疾病包括良性前列腺增生、慢性前列腺炎、前列腺囊肿等，我们的模型仍然取得了良好的区分能力。在本研究中，我们提取了尽可能多的特征。然后，我们采用了一种高效的特征选择方法-mRMR，该方法已被大多数研究^[16]证明是先进的，以获得最相关和最小冗余的特征。其原因可能是我们筛选方法选择的特征对相应的生物学特性具有较强的特异性，其中许多特征反映了病变的复杂性，如纹理特征。ADC值和T₂WI与病理前列腺组织中细胞核或细胞质面积百分比呈负相关，与腔隙百分比呈正相关^[17]。Gleason评分越高，腺体结构破坏导致腺腔外观越混乱和分散^[18]。以上结果可作为PCa与ADC及T₂WI征象相关性的组织学基础。

我们的研究存在一些局限性。首先我们的研究是回顾性的，

表1 测试集T₂WI+ADC影像组学机器学习模型ROC分析结果

	AUC	敏感度(%)	特意度(%)	准确率(%)	假阴性率(%)	假阳性率(%)
RF	0.871	0.600	0.941	0.738	0.400	0.059
SVM	0.896	0.640	0.941	0.762	0.360	0.059
XGboost	0.731	0.720	0.706	0.714	0.280	0.294

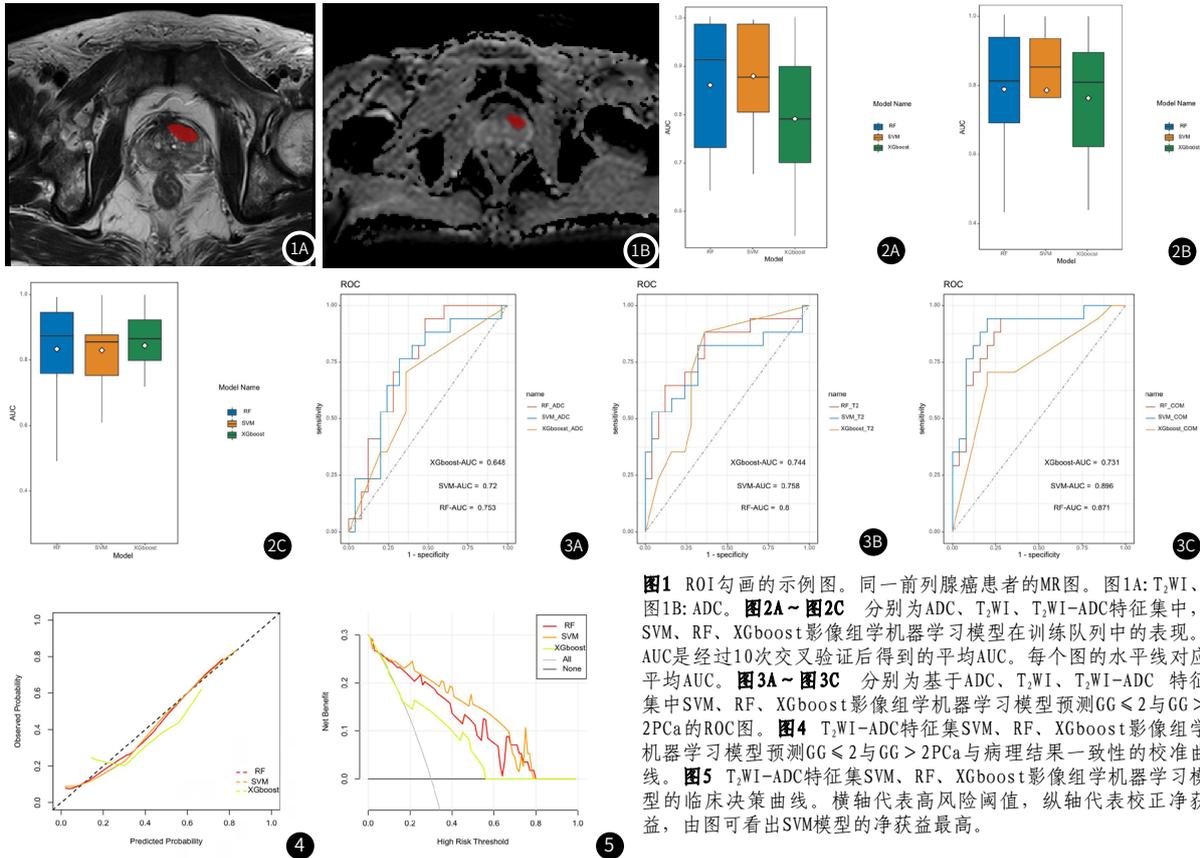


图1 ROI勾画的示例图。同一前列腺癌患者的MR图。图1A: T₂WI、图1B: ADC。图2A~图2C 分别为ADC、T₂WI、T₂WI+ADC特征集中, SVM、RF、XGboost影像组学机器学习模型在训练队列中的表现。AUC是经过10次交叉验证后得到的平均AUC。每个图的水平线对应平均AUC。图3A~图3C 分别为基于ADC、T₂WI、T₂WI+ADC特征集中SVM、RF、XGboost影像组学机器学习模型预测GG≤2与GG>2PCa的ROC图。图4 T₂WI+ADC特征集SVM、RF、XGboost影像组学机器学习模型预测GG≤2与GG>2PCa与病理结果一致性的校准曲线。图5 T₂WI+ADC特征集SVM、RF、XGboost影像组学机器学习模型的临床决策曲线。横轴代表高风险阈值,纵轴代表校正净获益,由图可看出SVM模型的净获益最高。

需要通过外部队列进行验证,以确定我们的模型在临床实践中的价值并提高性能的信心。其次,仅使用MRI图像放射组学特征进行分析。结合其他医学图像的多组学研究,如超声图像可能会提高我们预测模型的性能。第三,我们只是使用经典的监督分类算法来构建模型,基于深度学习的MRI图像特征需要进一步研究。我们下一步是对预测前列腺风险分层的多组学研究进行前瞻性和标准化研究。我们未来的研究将考虑来自多中心的外部验证。

综上所述,基于bp-MRI影像组学机器学习鉴别了GG≤2与GG>2PCa,最后SVM和RF模型取得了较好的结果,从而对PCa病变侵袭性进行定量评估,协助临床为不同患者制定个性化治疗方案。

参考文献

[1] Hernández G, Ramírez JL, Pedroza-Torres A, et al. The Secret Life of Translation Initiation in Prostate Cancer[J]. Front Genet, 2019, 10: 14.

[2] Matoso A, Epstein JI. Defining clinically significant prostate cancer on the basis of pathological findings[J]. Histopathology, 2019, 74(1): 135-145.

[3] Epstein JI, Egevad L, Amin MB, et al. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma[J]. Am J Surg Pathol, 2015, 40(2): 758-759.

[4] Sato S, Kimura T, Yorozu T, et al. Cases Having a Gleason Score 3+4=7 With <5% of Gleason Pattern 4 in Prostate Needle Biopsy Show Similar Failure Survival and Adverse Pathology Prevalence to Gleason Score 6 Cases in a Radical Prostatectomy Cohort[J]. Am J Surg Pathol, 2019, 43(11): 1560-1565.

[5] Epstein JI, Egevad L, Amin MB, et al. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System[J]. Am J Surg Pathol, 2016, 40(2): 244-252.

[6] Nketiah GA, Elschof M, Scheenen TW, et al. Utility of T2-weighted MRI texture analysis in assessment of peripheral zone prostate cancer aggressiveness: a single-arm, multicenter study[J]. Sci Rep, 2021, 11(1): 2085.

[7] 穆玖汐, 李琳, 曹绍东, 等. MRI在前列腺癌诊断和判断侵袭性中的进展[J]. 临床放射学杂志, 2018, 37(08): 1409-1412.

[8] Donisi L, Cesarelli G, Castaldo A, et al. A Combined Radiomics and Machine Learning Approach to Distinguish Clinically Significant Prostate Lesions on a Publicly Available MRI Dataset[J]. J Imaging, 2021, 7(10): 215.

[9] Kang Z, Xu A, Wang L. Predictive role of T2WI and ADC-derived texture parameters in differentiating Gleason score 3+4 and 4+3 prostate cancer[J]. J Xray Sci Technol, 2021; 29(2): 307-315.

[10] Bertelli E, Mercatelli L, Marzi C, et al. Machine and Deep Learning Prediction Of Prostate Cancer Aggressiveness Using Multiparametric MRI[J]. Front Oncol, 2022, 11: 802964.

[11] Turkbey B, Rosenkrantz AB, Haider MA, Padhani AR, Villeirs G, Macura KJ, Tempany CM, Choyke PL, Cornud F, Margolis DJ, Thoeny HC, Verma S, Barentsz J, Weinreb JC. Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2. Eur Urol, 2019, 76(3): 340-351.

[12] Nketiah GA, Elschof M, Scheenen TW, et al. Utility of T2-weighted MRI texture analysis in assessment of peripheral zone prostate cancer aggressiveness: a single-arm, multicenter study[J]. Sci Rep, 2021, 11(1): 2085.

[13] Donisi L, Cesarelli G, Castaldo A, et al. A Combined Radiomics and Machine Learning Approach to Distinguish Clinically Significant Prostate Lesions on a Publicly Available MRI Dataset[J]. J Imaging, 2021, 7(10): 215.

[14] 彭涛, 肖建明, 张仕慧, 等. 基于多参数MRI及影像组学建立机器学习模型诊断临床显著性前列腺癌[J]. 中国医学影像技术, 2019, 35(10): 1526-1530.

[15] Chen T, Zhang Z, Tan S, et al. MRI Based Radiomics Compared With the PI-RADS V2.1 in the Prediction of Clinically Significant Prostate Cancer: Biparametric vs Multiparametric MRI[J]. Front Oncol, 2022, 11: 792456.

[16] Meng X, Xia W, Xie P, et al. Preoperative radiomic signature based on multiparametric magnetic resonance imaging for noninvasive evaluation of biological characteristics in rectal cancer[J]. Eur Radiol, 2019, 29(6): 3200-3209.

[17] Langer DL, van der Kwast TH, Evans AJ, et al. Prostate tissue composition and MR measurements: investigating the relationships between ADC, T2, K(trans), v(e), and corresponding histologic features[J]. Radiology, 2010, 255: 485-494.

[18] Epstein JI. An update of the Gleason grading system[J]. J Urol, 2010, 183: 433-440.

(收稿日期: 2022-09-25)

(校对编辑: 谢诗婷)