

◆ 生殖泌尿影像学

Establishment of machine learning models for diagnosis of clinically significant prostate cancer based on multi-parameter MRI and radiomics

PENG Tao¹, XIAO Jianming¹, ZHANG Shihui¹, PU Bingjie¹, GAO Yueqin¹,
NIU Xiangke¹, WANG Zongyong¹, ZENG Xiaohui¹, YANG Jin^{2*}, LI Ci³
(1. Department of Radiology, 2. Department of Urology, 3. Department of Pathology,
Affiliated Hospital of Chengdu University, Chengdu 610081, China)

[Abstract] **Objective** To establish machine learning models based on multi-parameter MRI (mpMRI) and radiomics features, and to investigate their value for diagnosis of clinically significant prostate cancer (CSPC). **Methods** Logistic regression (LR), stepwise regression (SR), classical decision tree (cDT), conditional inference tree (CIT), random forest (RF) and support vector machine (SVM) models were established with combining of texture analysis, dynamic contrast enhanced MRI (DCE-MRI), prostate imaging reporting and data system (PI-RADS) score and part of clinical data. ROC curve and decision curve analysis (DCA) were used to evaluate the models and the importance of variables. **Results** AUC of RF model for diagnosing CSPC in verification group was larger than that of SVM, cDT and SR model (all $P < 0.05$). There was no statistically significant difference for diagnosing CSPC in AUC of RF model and LR, CIT model ($P > 0.05$), nor of AUC for diagnosing CSPC among the other models in validation group (all $P > 0.05$). When the probability threshold was 16%—91%, the net benefit of RF model was the largest, better than other models. When the probability threshold was 23%—91%, the net benefit of SVM model was second only to the RF model, but better than other models. Prostate specific antigen density (PSAD) and some texture analysis parameters were of high importance. **Conclusion** RF model is superior to other models in diagnosis of CSPC, SVM model comes second. PSAD and some texture analysis parameters are more important than PI-RADS score and DCE-MRI quantitative parameters for diagnosis of CSPC.

[Keywords] prostatic neoplasms; magnetic resonance imaging; texture analysis; radiomics; machine learning

DOI:10.13929/j.1003-3289.201902129

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

彭 涛¹, 肖建明¹, 张仕慧¹, 蒲冰洁¹, 高月琴¹, 牛翔科¹,
王宗勇¹, 曾小辉¹, 杨 进^{2*}, 李 欣³

(1. 成都大学附属医院放射科, 2. 泌尿外科, 3. 病理科, 四川 成都 610081)

[摘要] **目的** 建立基于多参数 MRI (mpMRI) 和影像组学特征的机器学习模型, 评价其诊断临床显著性前列腺癌 (CSPC) 的价值。 **方法** 结合纹理分析、MR 动态增强定量分析、前列腺影像报告与数据系统 (PI-RADS) 评分和部分临床资料建立 Logistic 回归 (LR)、逐步回归 (SR)、经典决策树 (cDT)、条件推断树 (CIT)、随机森林 (RF) 和支持向量机 (SVM) 模型, 运用 ROC 曲线和决策曲线分析法 (DCA) 评价上述模型和变量的重要性。 **结果** 验证组中 RF 模型诊断 CSPC 的

[基金项目] 四川省卫生和计划生育委员会科研课题 (18PJ150, 17PJ430)、成都市卫生和计划生育委员会医学科研课题 (2018055)。

[第一作者] 彭涛 (1977—), 男, 四川成都人, 硕士, 副主任医师。研究方向: 腹部与骨关节系统影像学诊断。E-mail: pengtao919@163.com

[通信作者] 杨进, 成都大学附属医院泌尿外科, 610081。E-mail: bendao2000@21cn.com

[收稿日期] 2019-02-26 **[修回日期]** 2019-07-17

AUC 大于 SVM、cDT、SR 模型 (P 均 < 0.05), RF 模型与 LR、CIT 模型诊断 CSPC 的 AUC 差异无统计学意义 (P 均 > 0.05), 其余各模型间诊断 CSPC 的 AUC 差异无统计学意义 (P 均 > 0.05)。概率阈值为 16%~91% 时, RF 模型的净获益最大, 优于其他模型; 概率阈值为 23%~91% 时, SVM 模型的净获益仅次于 RF 模型而优于其他模型。前列腺特异性抗原密度 (PSAD) 和部分纹理分析参数的重要性较高。结论 RF 模型诊断 CSPC 优于其他模型, SVM 模型次之。PSAD 和纹理分析相关参数诊断 CSPC 的重要性高于 PI-RADS 评分和动态增强 MRI 定量参数。

[关键词] 前列腺肿瘤; 磁共振成像; 纹理分析; 影像组学; 机器学习

[中图分类号] R737.25; R445.2 [文献标识码] A [文章编号] 1003-3289(2019)10-1526-05

前列腺癌居全球男性癌症发病率第 2 位、死亡率第 5 位, 是导致欧美男性因癌死亡的主要原因之一^[1], 近年发病率和死亡率呈迅速增长趋势^[2]。中国男性前列腺癌中, Gleason 评分 ≥ 7 分的临床显著性前列腺癌 (clinically significant prostate cancer, CSPC) 发病率居亚洲前列^[3]。由于非临床显著性前列腺癌 (non-clinically significant prostate cancer, NCSPC) 的 10 年生存率远高于 CSPC^[4], 美国泌尿外科协会临床指南推荐对于 NCSPC 患者无需穿刺活检, 仅动态随访即可^[5], 因此判断肿瘤的侵袭性成为重要诊断目的。多参数 MRI (multi-parameter MRI, mpMRI) 诊断前列腺癌现已得到广泛应用, 但仅通过形态学特征评判 CSPC 仍存在较大困难。本研究以 mpMRI 为基础, 结合临床、纹理分析和动态增强 MRI (dynamic contrast enhanced MRI, DCE-MRI) 定量参数, 建立机器学习模型, 优选 CSPC 诊断方法。

1 资料与方法

1.1 一般资料 选取 2014 年 1 月—2017 年 12 月 194 例在我院接受前列腺 mpMRI 的患者, 年龄 50~93 岁, 中位年龄 75 岁。纳入标准: ①前列腺 MRI 包括 T2WI、DWI、DCE-MRI; ②图像满足诊断, 无骨盆及股骨固定手术史; ③MR 检查前无前列腺相关手术治疗或穿刺史; ④MR 检查后 2 周内接受穿刺或根治手术并取得病理资料; ⑤前列腺特异性抗原 (prostate-specific antigen, PSA) 等临床及实验室检查资料完备。排除标准: ①mpMRI 序列不全; ②图像质量不能满足诊断; ③病理检查、实验室检查资料不完善。将 194 例患者分为 2 组: CSPC 组, 75 例, Gleason 评分 ≥ 7 分, 其中 27 例发生在移行带, 34 例在周围带, 14 例移行带和周围带均受累; NCSPC/良性病变 (benign Leision, BL) 组, 119 例, Gleason 评分 < 7 分, 其中 NCSPC 9 例。

1.2 仪器与方法 采用 Siemens Avanto 1.5T 超导型 MR 扫描仪。扫描序列及参数: T2W, TR 4 000 ms, TE 101 ms, 翻转角 150° , FOV 200 mm $\times$

200 mm, 层厚 3 mm, 层间距 0.6 mm; DWI, TR 3 700 ms, TE 80 ms, b 值取 0、100、800 s/mm², FOV 260 mm $\times$ 220 mm, 层厚 3.6 mm; DCE-MRI 时经肘前静脉注射钆喷酸葡胺 0.2 mmol/kg 体质量, 对比剂注入同时立即开始扫描, 采用 3D-flash-T1W 序列, TR 5 ms, TE 1.69 ms, FOV 260 mm $\times$ 260 mm, 层厚 3.6 mm, 动态扫描次数为 36 次。

1.3 前列腺影像报告与数据系统 (prostate imaging reporting and data system, PI-RADS) 评分 由 2 名具有 5 年以上诊断经验的前列腺 MRI 专业医师共同根据 PI-RADS v2 评分系统^[6]以盲法对前列腺病灶进行评分, 意见不同时经商议达成一致; 存在多个病灶时, 以主要病灶的评分为最终评分。评判主要病灶: 若有前列腺外侵犯, 不论 PI-RADS 评分和病灶大小, 均认定为主要病灶; 若无前列腺外侵犯, 则以 PI-RADS 评分最高的病灶为主要病灶; 若 PI-RADS 评分相同, 以体积最大的病灶为主要病灶^[6]。

1.4 纹理分析与 DCE-MRI 定量分析 采用 GE Omni Kinetics 软件进行后处理。根据轴位 T2WI、DWI 及 DCE-MRI 病灶信号特点, 对照病理按病变所在分区^[6]逐层勾画 ROI。对 CSPC 根据病理反映的肿瘤所在分区结合信号表现确定 ROI; 对 NCSPC 或 BL 选取低级别癌、高级别上皮内瘤或最接近肿瘤表现区域作为 ROI, 通过融合 3D-ROI 计算纹理参数 (图 1)。DCE-MRI: 选取达峰期相图像, 以 Extended Tofts 双室模型进行定量分析。存在多个病灶时, 针对主要病灶进行评估。

1.5 数据处理、机器学习建模与统计学分析 建立数据库, 包括 T2WI 及 ADC 的纹理特征、DCE-MRI 定量参数、PI-RADS 评分、临床参数 (包括年龄、PSA 及衍生参数) 等共 237 个参数, 利用 R 语言 (V 3.5.1) 的 Caret 软件包对综合数据库变量进行降维和特征选择, 建立机器学习模型; 于 CSPC 组和 NCSPC/BL 组分别随机抽取 70% 作为训练组, 剩余 30% 作为验证组。在训练组中建立 Logistic 回归 (logistic regression,

LR)、逐步回归(stepwise regression, SR)、经典决策树(classical decision tree, cDT)、条件推断树(conditional inference tree, CIT)、随机森林(random forest, RF)和支持向量机(support vector machine, SVM)模型,RF 模型使用 500 棵树,SVM 模型采用 RBF(radial basis function)核函数。在验证组中采用 ROC 曲线和混淆矩阵比较不同模型的诊断效能,以 Delong's 法比较 ROC 曲线 AUC 差异, $P < 0.05$ 为差异有统计学意义。对上述模型进行决策曲线分析(decision curve analysis, DCA)。

2 结果

降维和特征选择后得到预测精度最高的 10 个自变量为 T2WI、T2. Relative Deviation、前列腺特异性抗原密度(prostate specific antigen density, PSAD)、T2. Quantile5、Vp0.1、ADC. Quantile95、ADC. MinIntensity、ADC. uniformity、T2. Variance 和 TTPMax。训练组中 CSPC 51 例,NCSPC/BL 84 例;验证组中 CSPC 24 例,NCSPC/BL 35 例。LR 模型筛选出的独立预测因子为 PSAD、T2. Quantile5、ADC. Quantile95、ADC. MinIntensity 和 T2. Variance,其中 ADC. MinIntensity 的 $P < 0.01$;SR 模型筛选出的独立预测因子为 PSAD、T2. Quantile5、ADC. MinIntensity、T2. Variance 和 TTPMax,其中 T2. Quantile5、ADC. MinIntensity、T2. Variance 的 $P < 0.01$ 。建立 cDT 模型,通过 plotcp 函数绘制“交叉验证误差与复杂度参数关系图”,选得最优树为 2 次分割(3 个终端节点)对应的树。建立 CIT 模型和 RF 模型,RF 模型自变量重要性排序见图 2。

将 LR、SR、cDT、CIT、RF、SVM 各模型用于验证组,验证组中各模型诊断 CSPC 的 ROC 曲线见图 3,混淆矩阵及 ROC 曲线的 AUC 见表 1,诊断预测特征见表 2。

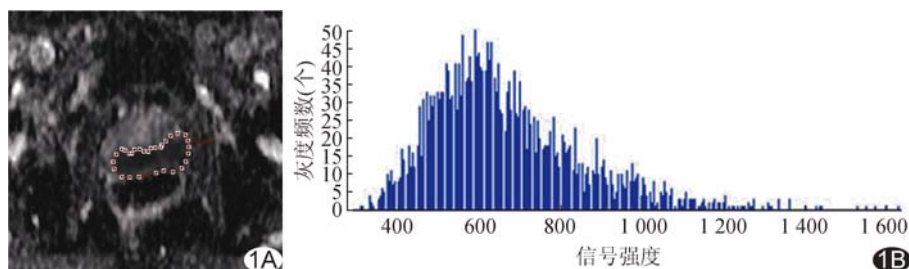


图 1 ROI 勾画及直方图 A. 在 ADC 图像中勾画 ROI; B. ROI 信号强度直方图

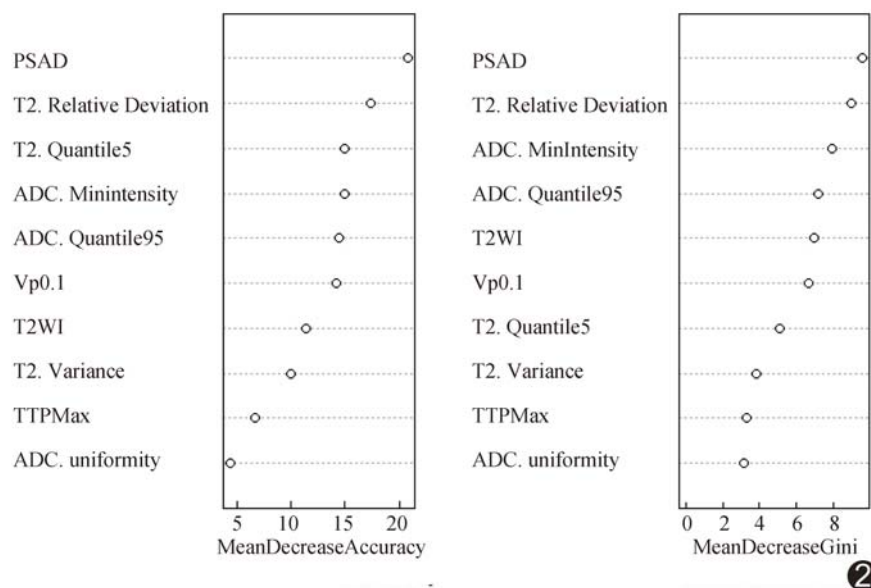


图 2 RF 模型自变量重要性示意图 MeanDecreaseAccuracy 和 MeanDecreaseGini 表示变量的重要性,值越大表示其重要性越高

表 1 验证组各模型混淆矩阵及 ROC 曲线 AUC

模型	PA00(例)	PA01(例)	PA10(例)	PA11(例)	AUC
LR	31	4	4	20	0.932
SR	32	4	3	20	0.924
cDT	29	5	6	19	0.860
CIT	25	3	10	21	0.924
RF	33	2	2	22	0.977
SVM	33	4	2	20	0.888

注:PA00:真阴性;PA01:假阴性;PA10:假阳性;PA11:真阳性

表 2 验证组各模型诊断预测特征

模型	敏感度	特异度	阳性预测值	阴性预测值	准确率
LR	0.83	0.89	0.83	0.89	0.86
SR	0.83	0.91	0.87	0.89	0.88
cDT	0.79	0.83	0.76	0.85	0.81
CIT	0.88	0.71	0.68	0.89	0.78
RF	0.92	0.94	0.92	0.94	0.93
SVM	0.83	0.94	0.91	0.89	0.90

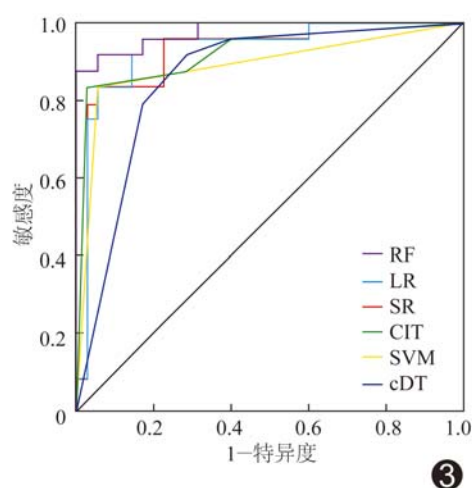


图3 验证组各模型诊断 CSPC 的 ROC 曲线

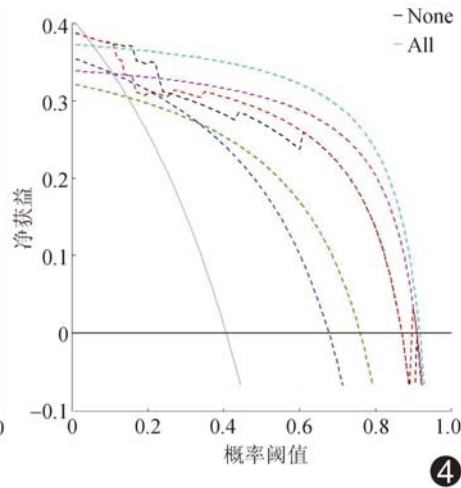


图4 决策曲线 概率阈值为 16%~91%时, RF 模型净获益最大, 优于其他模型; 概率阈值为 23%~91%时, SVM 模型的净获益仅次于 RF 模型而优于其他模型 (All: 所有受试者均为 CSPC 患者; None: 所有受试者均非 CSPC 患者; 浅蓝虚线: RF; 粉红虚线: SVM; 红虚线: SR; 黑虚线: LR; 绿虚线: cDT; 深蓝虚线: CIT)

利用 Delong's 法比较 AUC 的差异。验证组 RF 模型与 SVM、cDT、SR 模型诊断 CSPC 的 AUC 差异有统计学意义 (P 均 < 0.05), RF 模型与 LR、CIT 模型诊断 CSPC 的 AUC 差异无统计学意义 (P 均 > 0.05), 其余各模型间诊断 CSPC 的 AUC 差异无统计学意义 (P 均 > 0.05)。DCA 显示, 概率阈值为 16%~91%时, RF 模型的净获益最大, 优于其他模型; 概率阈值为 23%~91%时, SVM 模型的净获益仅次于 RF 模型而优于其他模型(图 4)。

3 讨论

提高 CSPC 的检出率是降低死亡率和减少不必要穿刺活检的关键。PI-RADS v2 中指出, 前列腺 MR 检查的主要目的是鉴别和定位 CSPC, CSPC 定义为 Gleason 评分 ≥ 7 分(包括 $3+4=7$ 分)、和/或病灶体积 ≥ 0.5 ml、和/或前列腺外侵犯^[6], 一定程度上提高了 CSPC 的检出率, 但具有局限性, 如鉴别诊断小体积 CSPC、DWI 阴性的周围叶 CSPC、T2WI 阴性的移行带 CSPC、间质增生结节及前列腺炎等较困难。

本研究在 mpMRI 基础上结合影像组学、PI-RADS 评分和部分临床资料, 建立机器学习模型并进行评价, ROC 曲线和 DCA 均显示 RF 为最佳模型, RF 模型诊断 CSPC 的 AUC 最大, 且与 SVM、cDT、SR 模型差异有统计学意义, 但与 LR、CIT 模型间差异无统计学意义。DCA 表明, 概率阈值为 16%~91%时, RF 模型的净获益最大, 优于其他模型, 与 ROC 曲线结果一致。RF 模型是一种包含多个决策树的分类器, 其

从原始训练样本集中有放回地随机抽取样本, 生成新的训练样本集合, 重复多次以训练多个决策树, 采用多数投票法进行分类, 对均衡样本集的处理具有更好的分类性能, 且调整参数少、容噪能力好、不易出现过拟合, 并能高效处理大量特征^[7]。Lay 等^[8] 建立基于信号、MRI 纹理特征的计算机辅助诊断 RF 模型和 SVM 模型, 其 RF 模型(AUC 为 0.93) 优于 SVM(AUC 为 0.86), 本研究结果与之相似。

本研究中, DCA 的概率阈值为 23%~91%时, 使用 SVM 模型的净获益仅次于 RF 模型而优于其他模型, 但验证组 SVM 的

AUC 低于除 cDT 以外其他模型, 此时 DCA 的结果更有意义。SVM 在解决小样本、高维度、非线性等方面具有较高性能, 可利用较少训练样本获得与大量训练样本几乎相同的精度^[9], 其基本思想是寻找使两类变量之间的分类间隔最大的最优分类超平面^[10]。目前结合 PI-RADS 评分和 MR 影像组学特征所开发的 SVM 分类器诊断前列腺癌已取得显著进步^[11]。ROC 曲线和 DCA 均属于预测模型, 可用来评价不同诊断方法或模型的优劣, 但二者又有本质区别, 因为 ROC 曲线并未关注预测信息所致后续影响, 故其 AUC 最大模型并不一定最优, 而 DCA 则更兼顾临床可用性及效益^[12]。

LR 是二分类任务中最常用的机器学习算法之一, 属于有监督的机器学习模型, 应用较广^[13]。SR 模型在 LR 基础上去掉了未通过显著性检验的预测变量, 得到一个精简模型, 其在验证集上的误差相对于全变量模型更小。本研究先行对数据库进行降维和特征选择, 仅保留少数预测精度高的变量, 故 LR 和 SR 模型诊断效能差异并不显著。cDT 模型是数据挖掘常用模型, 以二元输出变量和一组预测变量为基础。CIT 是 cDT 的变体, 基于显著性检验选取变量和分割, 而非选择组内同质性数据。cDT 由 rpart 包建立, 而 CIT 用 party 包来建立且不需要剪枝, 以阈值决定模型的复杂程度。cDT 易出现过拟合^[14], 而 RF 模型则可基本避免这个问题。

本研究中 RF 模型对参数的重要性进行了排序, MeanDecreaseAccuracy 和 MeanDecreaseGini 图中均

排在前 5 位的为 PSAD、T2. Relative Deviation、ADC. MinIntensity、ADC. quantile95, 前两者值越大越支持诊断 CSPC, 后两者相反。PSAD 为 PSA 的衍生参数, 代表单位体积前列腺 PSA 的含量。Verma 等^[15]认为 PSAD 可显著增加肿瘤检出率, 且与肿瘤侵袭性关系密切。PI-RADS 评分受主观因素影响较大, 其在 ROC 曲线评价中的 AUC 为 0.904, 敏感度为 0.76、特异度为 0.94、约登指数为 0.7, 特征选择后, 其中的 T2WI 评分在 MeanDecreaseAccuracy 和 MeanDecreaseGini 图中分别排在第 7 和第 5 位, 其重要性低于相对客观的 PSAD 和纹理特征。DCE-MRI 定量参数 Vp0.1 和 TTPMax 的重要性分别排在第 6 和第 9 位, Vp 代表血管内对比剂所占容积百分比, 反映平均血管密度增加、灌注血管增多、血管通透性高的特点; TTPMax 越小越支持诊断 CSPC, 反映肿瘤快速强化的特点。

本研究局限性: ①因样本量相对较小, 故未选择卷积神经网络模型进行评价; ②PI-RADS 评分虽为 2 名医师共同进行, 但仍可能存在主观偏倚, 此为 PI-RADS 评分系统的固有缺陷; ③部分病理结果来自于超声引导下穿刺活检取材, 不能避免穿刺漏诊可能。

综上所述, RF 模型在影像学诊断 CSPC 方面具有较好的分类性能, SVM 模型仅次于 RF 模型。PSAD 和纹理分析相关参数的重要性高于 PI-RADS 评分和 DCE-MRI 定量参数, 其相对客观性可为人工智能诊断提供依据。

[参考文献]

- [1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 2018, 68(6):394-424.
- [2] Center MM, Jemal A, Lortet-Tieulent J, et al. International variation in prostate cancer incidence and mortality rates. *Eur Urol*, 2012, 61(6):1079-1092.
- [3] Chen R, Ren S. Prostate cancer in Asia: A collaborative report. *Asian J Urol*, 2014, 1(1):15-29.
- [4] Albertsen PC, Hanley JA, Fine J. 20-year outcomes following conservative management of clinically localized prostate cancer. *JAMA*, 2005, 293(17):2095-2101.
- [5] Thompson I, Thrasher JB, Aus G, et al. Guideline for the management of clinically localized prostate cancer: 2007 update. *J Urol*, 2007, 177(6):2106-2131.
- [6] Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS prostate imaging -reporting and data system: 2015, version 2. *Eur Urol*, 2016, 69(1):16-40.
- [7] Qian C, Wang L, Gao Y, et al. In vivo MRI based prostate cancer localization with random forests and auto-context model. *Comput Med Imaging Graph*, 2016, 52:44-57.
- [8] Lay N, Tsehay Y, Greer MD, et al. Detection of prostate cancer in multiparametric MRI using random forest with instance weighting. *J Med Imaging (Bellingham)*, 2017, 4(2):024506.
- [9] Fasshauer GE, Hickernell FJ, Ye Q. Solving support vector machines in reproducing kernel Banach spaces with positive definite functions. *Appl Comput Harmon Anal*, 2015, 38(1):115-139.
- [10] 李梅, 张伟, 李永忠, 等. SVM 神经网络在判别前列腺癌中的应用. *四川大学学报(医学版)*, 2013, 44(4):666-668.
- [11] Wang J, Wu CJ, Bao ML, et al. Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. *Eur Radiol*, 2017, 27(10):4082-4090.
- [12] 王启明, 周苏娜, 邵秋菊, 等. 基于决策曲线分析和生存分析评估 Her-2 和 Ki-67 对结直肠癌预后的价值. *山西医科大学学报*, 2018, 49(5):513-517.
- [13] Eftekhari B, Mohammad K, Ardebili HE, et al. Comparison of artificial neural network and Logistic regression models for prediction of mortality in head trauma based on initial clinical data. *BMC Med Inform Decis Mak*, 2005, 5:3.
- [14] Tong W, Xie Q, Hong H, et al. Using decision forest to classify prostate cancer samples on the basis of SELDI-TOF MS data: Assessing chance correlation and prediction confidence. *Environ Health Perspect*, 2004, 112(16):1622-1627.
- [15] Verma A, St Onge J, Dhillon K, et al. PSA density improves prediction of prostate cancer. *Can J Urol*, 2014, 21(3):7312-7321.