

· 人工智能影像学专题 ·

基于术前分期 CT 的影像组学标签预测三阴性乳腺癌

张文, 何兰, 范志豪, 黄晓媚, 杨晓君, 梁长虹, 刘再毅

【摘要】 目的:探讨基于术前分期 CT 的影像组学标签在预测三阴性乳腺癌分子分型中的附加价值。**方法:**回顾性收集 2016 年 1 月至 2018 年 5 月经手术病理证实且均为临床术前评估分期需行常规胸部 CT 增强扫描的 481 例肿块型乳腺浸润性癌患者,按照样本量 1:2 随机抽样选取三阴性乳腺癌与非三阴性乳腺癌共计 150 例患者(90 例作为训练组,60 例作为验证组)。所有患者均经免疫组织化学检测,获得乳腺癌分子分型。对所有患者基于病灶三维图像提取影像组学特征,并采用 Lasso logistic 回归模型进行特征降维及筛选,以建立影像组学标签。采用 ROC 曲线评价影像组学标签对三阴性乳腺癌的鉴别诊断效能。**结果:**由 5 个关键影像组学特征构成的影像组学标签与乳腺癌三阴性分子分型相关($P < 0.0001$)。建立的影像组学标签对于鉴别三阴性乳腺癌具有较好的预测效能,其在训练组和验证组的 ROC 曲线下面积(AUC)分别为 0.766(95% CI:0.743~0.789)和 0.758(95% CI:0.718~0.798)。**结论:**基于术前分期 CT 建立的影像组学标签有助于三阴性与非三阴性乳腺癌的鉴别,这是术前常规胸部增强 CT 扫描在辅助临床分期之外的附加价值,可为临床治疗决策提供参考。

【关键词】 乳腺肿瘤; 三阴性乳腺癌; 影像组学; 体层摄影术, X 线计算机; 分子分型; 预测

【中图分类号】 R737.9; R814.42 **【文献标识码】** A **【文章编号】** 1000-0313(2019)09-0947-05

DOI:10.13609/j.cnki.1000-0313.2019.09.003

开放科学(资源服务)标识码(OSID):



Preoperative staging CT-based radiomics signature in predicting triple-negative breast cancer ZHANG Wen, HE Lan, FAN Zhi-hao, et al. The Second School of Clinical Medicine, Southern Medical University, Guangzhou 510515, China

【Abstract】 Objective: To explore the additional value of preoperative staging CT-based radiomics signature in predicting molecular subtype of triple-negative breast cancer. **Methods:** From January 2016 to May 2018, 481 patients with pathological confirmed invasive breast cancer (mass type) were enrolled in the study. All patients were performed routine staging enhanced chest CT before surgery. A total of 150 cases were randomly selected according to the ratio of 1:2 (triple-negative vs non-triple-negative breast cancer). Among the 150 cases, 90 cases were divided into the training dataset and 60 cases were divided into the validation dataset. All patients obtained detailed information about molecular subtypes of breast cancer by immunohistochemistry. The radiomic features were extracted based on volume of the interest in all patients and Lasso logistic regression model was used for dimensionality reduction, feature selection and construction of the radiomics signature. The receiver operation curve (ROC) was used to evaluate the performance of the radiomics signature in predicting triple-negative breast cancer. **Results:** The radiomics signature composed of 5 radiomic features was associated with the triple-negative molecular subtype of breast cancer ($P < 0.0001$). The constructed radiomics signature obtained good predictive performance in identifying triple-negative breast cancer, with the area under the ROC curve (AUC) of 0.766 (95% CI:0.743~0.789) and 0.758 (95% CI:0.718~0.798) in the training and validation dataset, respectively. **Conclusion:** The constructed radiomics signature based on preoperative staging CT could be helpful in differentiating between triple-negative and non-triple-negative breast cancer, which can provide additional value of routine preoperative staging CT for clinical

作者单位: 510515 广州, 南方医科大学第二临床医学院(张文、刘再毅); 510080 广州, 广东省人民医院放射科 广东省医学科学院(张文、何兰、范志豪、黄晓媚、杨晓君、梁长虹、刘再毅)

作者简介: 张文(1988-), 男, 广东珠海人, 硕士研究生, 主治医师, 主要从事胸腹部影像诊断工作。

通讯作者: 刘再毅, E-mail: zyliu@163.com

基金项目: 国家重点研发计划(2017YFC1309100); 广东省省级科技计划项目(2017B020227012)

cal treatment decisions.

【Key words】 Breast neoplasms; triple negative breast cancer; Radiomics; Tomography, X-ray computed; Molecular subtypes; Prediction

乳腺癌是女性最常见的恶性肿瘤,也是女性恶性肿瘤中病死率最高的疾病^[1]。研究表明,决定早期乳腺癌治疗方案、提示疗效与预后的并非组织病理学类型,而是其分子分型^[2,3]。三阴性乳腺癌(triple negative breast cancer, TNBC)占乳腺癌的 10%~20%^[4],是一种异质性疾病,其生物学特征不同于其他类型的乳腺癌,缺乏分子靶向治疗的有效靶点且预后不良,目前化疗仍是临床上 TNBC 唯一有效的治疗方案^[5]。因此,早期辨别 TNBC 与非三阴性乳腺癌(non triple negative breast cancer, NTNBC)对取得最佳诊疗方案具有重要临床意义。目前临床主要依靠免疫组织化学明确乳腺癌的分子分型,但由于对肿瘤组织进行活检取样是一种具有侵入性的方法,同时取样和分析时存在局限性^[6]。相比之下,影像检查能够无创且从整体上反映肿瘤特征,一方面在基因水平上对每个亚型内的差异进一步分析,另一方面能够多次动态评价治疗效果^[7]。

根据美国国立综合癌症网络(National Comprehensive Cancer Network, NCCN)指南,胸部 CT 增强扫描已被列入乳腺癌患者术前常规检查之一。不同于乳腺 X 线摄影、超声及磁共振检查对乳腺病灶的直观评估,胸部 CT 增强扫描的主要目的是用于辅助临床分期。通常认为 CT 检查对微小钙化的显示不及乳腺 X 线摄影,对乳腺囊实性病变的诊断准确性不及超声,对乳腺良、恶性病变的鉴别也无显著特异性。但随着影像组学的发展^[8,9],我们假设这种新的手段可以在乳腺癌患者术前分期 CT 中提取出以往肉眼无法直接观测到的信息,而这些信息有可能与肿瘤的分子分型、生物学特性等存在一定的关系,从而使术前常规 CT 扫描在辅助临床分期之外,提供对病灶特征评估有益的附加信息,目前国内外尚未见相关报道。

本研究拟基于影像组学的方法,在不额外增加患者辐射剂量和经济负担的前提下,进一步挖掘乳腺癌术前常规 CT 图像数据中蕴含的信息,旨在探讨术前分期 CT 在预测三阴性乳腺癌分子分型中的附加价值。

材料与方法

1. 研究对象

本研究经医院伦理委员会批准,并免除知情同意要求。回顾性收集 2016 年 6 月至 2018 年 1 月经我院病理组织活检或手术确诊的乳腺癌患者。病例纳入标准:①经手术和/或活检组织病理学证实的肿块型乳腺

浸润性癌;②具备用于术前评估乳腺癌临床分期且未经治疗的胸部增强 CT 扫描图像;③免疫组织化学资料完整,可获得分子分型。病例排除标准:①图像质量不佳,影响病灶观察及勾画;②相关临床资料或病理信息不完整;③特殊类型的乳腺癌。

2. 免疫组化获得分子分型

所有患者均进行了乳腺肿瘤活检或手术,获得的病理标本均进行免疫组织化学检测,包括雌激素受体(estrogen receptor, ER)、孕激素受体(progesterone receptor, PR)、人表皮生长因子受体 2(human epidermal growth factor receptor2, HER2)和 Ki-67。将乳腺癌分为以下 4 型:Luminal A 型、Luminal B 型、HER2 过表达型和三阴性(triple negative, TN)^[10]。

3. CT 扫描方案

采用 Brilliance iCT 256 (Philips Healthcare, Cleveland OH, USA)、Somatom Definition Flash (Siemens Healthcare, Forchheim, Germany)、Philips Ingenuity CT (Philips Healthcare, Cleveland, Ohio, USA)及 LightSpeed VCT (GE Medical systems, Milwaukee, Wis, USA)扫描仪进行术前胸部 CT 增强扫描。CT 扫描参数:管电压 120 kVp,管电流 150 mAs;Brilliance iCT 256、Somatom Definition Flash 和 Philips Ingenuity CT 旋转时间为 0.5 s,LightSpeed VCT 旋转时间为 0.4 s;探测器准直分别为 128×0.625 mm、2×64×0.6 mm、64×0.625 mm 及 64×0.625 mm;视野为 350 mm×350 mm,矩阵为 512×512,重建层厚为 1.0 mm 或 1.25 mm。CT 扫描方案:首先行常规 CT 平扫,采集平扫 CT 图像;其次行增强 CT 扫描,采用高压注射器(Ulrich CT Plus 150, Ulrich Medical, Ulm, Germany)经静脉注射非离子型对比剂优维显 80~100 mL(370 mg I/mL),流率 4 mL/s,分别于注射对比剂后 18 s、35 s 采集肺动脉期、主动脉期 CT 图像。

4. 影像组学特征提取

采用基于 MATLAB 2014a (Mathworks, Natick, MA, USA)软件进行影像组学分析。兴趣区(region of interest, ROI)的放置由两位放射科医生(分别具有 3 年和 7 年的乳腺影像诊断经验)使用 itk-SNAP 软件(<http://www.itksnap.org/pmwiki/pmwiki.php>)完成。选取主动脉期图像,剔除病灶显示不清或非肿块型病灶,沿肿块病灶边缘逐层手动勾画 ROI,ROI 应尽量包括整个病灶,注意避开病灶内坏死、钙化及气体

影等,最终得到整个瘤体感兴趣体积(volume of interest, VOI)。第一位医生(医生 1)放置 ROI 并进行影像组学特征提取一次,1 周后再进行第二次 ROI 放置及特征提取,用于测量者自身一致性的评估;另一位医生(医生 2)仅进行 ROI 放置及特征提取一次,与医生 1 的第一次测量进行测量者间的一致性评估。

所提取的影像组学特征主要分为三类:①一阶统计特征;②形态学特征;③纹理特征。每例患者的术前 CT 图像共提取 182 个影像组学特征(表 1)。

表 1 提取的影像组学特征

特征类型	特征名称
一阶灰度直方图特征	能量、熵、最小值、最大值、平均值、中位数、平均绝对偏差、均方差、标准偏差、偏度、峰度、方差、均匀度
形态学特征	表面积、周长、凹度、体素量、最大 3D 直径、球度
灰度纹理特征	灰度共生矩阵 GLCM、灰度游程矩阵 GLRLM、灰度区域大小矩阵 GLSZM、灰度差异矩阵 GLDM、邻域灰度差异矩阵(NGTDM)

5. 统计学分析

患者临床病理基线资料的比较:对患者临床病理基线资料训练组与验证组间的差异性比较依据数据特征采用卡方检验(年龄)、独立样本 t 检验或 Mann-Whitney U 检验进行统计学分析。

观察者自身及之间的一致性评估:采用组内和组间相关系数(intra and interclass correlation coefficients, ICCs)对医生 1 自身和两位测量者间的一致性进行评价。采用医生 1 的两次测量数据计算测量者自身的 ICC;采用医生 1 的第一次测量结果与医生 2 的测量结果计算测量者间的 ICC。ICC>0.75 认为其一致性好。

特征筛选和影像组学标签的建立:采用 Lasso-logistic 回归模型筛选基于动脉期 VOI 提取的对乳腺癌三阴性状态具有预测意义的影像组学特征。通过选取的特征与对应加权系数乘积的线性组合依次形成每例患者的影像组学标签,并计算相应值。

影像组学标签预测效能的评价:应用受试者工作特征(receiver operator characteristic, ROC)曲线的曲线下面积(area under curve, AUC)评估建立的影像组学标签对乳腺癌三阴性状态的鉴别预测效能, AUC 越大证明其诊断效能越高。计算相应的 ROC 曲线下面积(AUC)、敏感度、特异度,均包括 95%置信区间的计算。以上分析的界值(cut-off points)以最大约登指数(youden index)(敏感度+特异度-1)确定。

采用 R 软件(版本:3.0.1; <http://www.Rproject.org>)进行统计学分析。加载相应的程序包或函数完成相应的统计检验。以 $P<0.05$ 为差异有统计学意义。

结 果

1. 人口学及临床特征

共纳入 481 例肿块型乳腺浸润性癌病例形成研究队列,其中 TNBC 50 例,根据病例号,采用随机数字法,按照样本量 1:2 随机抽样选取三阴性乳腺癌与非三阴性乳腺癌共计 150 例患者(90 例作为训练组,60 例作为验证组)进行分析,患者均为女性。TNBC 患者的年龄为 23~81 岁,中位年龄 49 岁;NTNBC 患者的年龄为 26~73 岁,中位年龄 49 岁,两组患者的年龄差异无统计学意义($P>0.05$,表 2)。

表 2 乳腺癌患者的临床资料及病理结果比较

临床特征	训练组	验证组	t/χ^2 值	P 值
年龄(岁)	49.93±11.90	49.95±11.05	-0.009	0.993
月经状态(n,%)			1.394	0.238
绝经前	61(67.8)	35(58.3)		
绝经后	29(32.2)	25(41.7)		
病理分级(n,%)			4.197	0.123
I	8(8.9)	1(1.7)		
II	44(48.9)	27(45)		
III	38(42.2)	32(53.3)		
分子分型(n,%)			1.227	0.754
Luminal A 型	11(12.2)	6(10)		
Luminal B 型	30(33.3)	17(28.3)		
HER2 过表达型	19(21.2)	17(28.3)		
三阴性(TNBC)	30(33.3)	20(33.4)		

2. 观察者自身及之间的一致性评估

基于医生 1 与医生 2 提取的特征计算得出测量者间的一致性较好(ICC 为 0.785~0.998),基于医生 1 两次提取的特征计算得出测量者内的一致性较好(ICC 为 0.812~0.998)。因此,下述分析均基于医生 1 第一次提取的特征。

3. 影像组学标签的建立

通过 Lasso-Logistic 回归模型共筛选出具有非零系数的 5 个影像组学特征(图 1)。基于这些特征与之相应的系数乘积的线性组合构建影像组学标签(公式 1)。训练组中, TNBC 组患者的影像组学标签值为 -0.496(四分位数间距 -0.707, -0.222),高于 NTNBC 组患者的 -0.864(四分位数间距 -1.031, -0.634),差异有统计学意义($P<0.001$);验证组中, TNBC 组和 NTNBC 组患者的影像组学标签值分别为 -0.603(四分位数间距 -0.902, -0.346)和 -0.899(四分位数间距 -1.010, -0.768),差异也具有统计学意义($P<0.001$)。

$$\text{影像组学标签值} = -2.61548 + 0.010178 \times \text{compactness1} - 50.3070 \times \text{GLZM_ZP} + 0.11981 \times \text{GLSZM_SAHGLE} + 0.26238 \times \text{band_max} + 1.43099 \times \text{band_mean} \quad (1)$$

4. 预测效能评估

本研究建立的影像组学标签对乳腺癌分子分型三阴性具有较好的预测效能。训练组中的 AUC 为 0.766(95% CI:0.743~0.789),敏感度为 0.667,特异

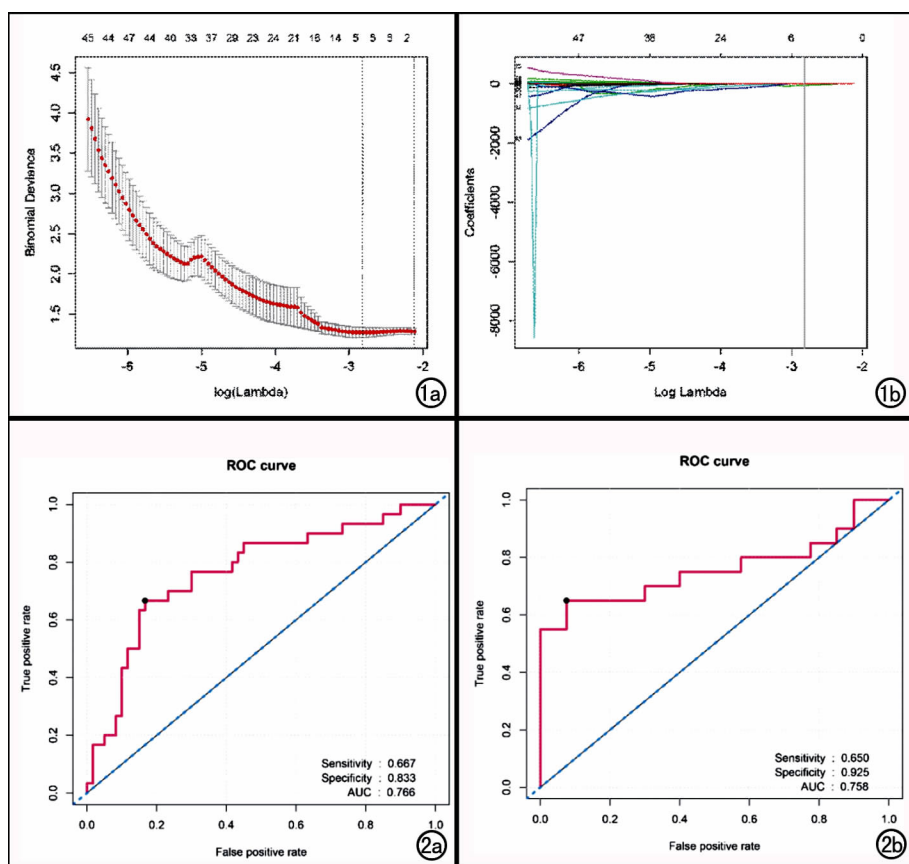


图 1 采用 Lasso-Logistic 回归模型对影像组学特征进行筛选。a) 纵坐标为二项式偏差(binomial deviance), 横坐标为 $\log(\lambda)$ 。模型中使用十倍交叉验证方法, 通过调节不同参数 λ 实现模型的二项式偏差最小, 从而筛选出与三阴性乳腺癌相关联的特征集合。本研究共筛选出 5 个特征; b) 为特征筛选过程的系数收敛图。纵坐标表示特征在模型中的各自系数, 横坐标为 $\log(\lambda)$ 。垂直线对应采用十折交叉验证后拟合筛选出的非零特征, 共筛选出 5 个特征。图 2 影像组学标签对乳腺癌三阴性状态预测的 ROC 曲线。a) 训练组的 ROC 曲线; b) 验证组的 ROC 曲线。

度为 0.833, 阳性预测值为 0.655, 阴性预测值为 0.820; 在验证组中的 AUC 为 0.758 (95% CI: 0.718~0.798), 敏感度为 0.65, 特异度为 0.925, 阳性预测值为 0.800, 阴性预测值为 0.822(图 2)。

讨 论

通过本次研究, 我们首次探讨了基于术前分期 CT 的影像组学标签在术前鉴别 TNBC 与 NTNBC 的价值, 可以为临床治疗决策提供参考信息。本研究回顾性分析的图像, 均为我院已有的用于乳腺癌术前分期的胸部增强 CT 图像, 并没有额外增加患者的辐射剂量与经济负担, 这对于术前分期 CT 获得的主观特征, 影像组学方法对图像信息的挖掘, 具有附加的临床价值。

目前, 乳腺癌临床治疗方案的制定主要依据分子

分型, TNBC 的生物学特征不同于其他类型的乳腺癌, 相同治疗方案的 TNBC 患者其反应及预后也不尽相同, 原因是由于肿瘤的异质性^[11]。由于肿瘤生长过程中时间和空间存在差异性, 病理活检有时并不能代表完全的肿瘤组织特点, 而影像组学标签由图像中提取得到的定量化特征构成, 与肿瘤病灶的主观定性特征相比, 可以更全面、无创地定量描述肿瘤异质性^[12-14]。本研究联合了 5 个 CT 图像特征构建影像组学标签, 作为预测乳腺癌分子分型的影像生物标志, 成功将 TNBC 与 NTNBC 患者进行分层 ($P < 0.0001$)。此外, 本研究提取得到大量影像组学特征, 在解决高通量计算中普遍面临的预测因子筛选问题中, 采用了一种惩罚估计技术——Lasso logistic 回归模型, 以实现变量筛选, 最终得到了由 5 个影像组学特征组成的影像组学标签, 显示了较好的预测效能。在此之前, 相关学者基于乳腺 X 线摄影、磁共振的影像组学鉴别 TNBC 与 NTNBC 做了较多研究^[15-17], 但都没有验证。本研究通过将训练组得到的预判能力阈值用于验证组进行验证, 使实验结果更可靠。同时, 我们发现用于术前评估患者临床分期的 CT 图像, 可以通过影像组学的方法进一步挖掘并获得图像特征, 这些特征构建的标签在预测三阴性乳腺癌中有较高的特异度(训练组为 0.833, 验证组为 0.925)。对比乳腺癌其他分子分型, TNBC 对靶向治疗及内分泌治疗效果差, 对化疗敏感性较高, 而化疗具有潜在的毒性, 为了使化疗得到合理的运用, 降低潜在风险, 影像组学标签的高特异度可能会为乳腺癌个体化治疗带来额外的价值。分期 CT 属于临床常规检查, 不会增加患者的辐射剂量和经济负担, 可为临床诊疗提供额外的信息, 因此具有一定的临床应用价值。

本研究有一定的局限性: 首先, 这是对单个机构获得的图像进行回顾性分析。在后续研究中, 需增大样本量, 利用多中心数据对本研究结果进行验证; 其次, 本研究基于临床分期 CT 进行数据挖掘, 未针对乳腺

病灶进行数据采集优化。

综上所述,基于术前分期 CT 建立的影像组学标签有助于乳腺癌分子分型及 TNBC 与 NTNBC 的鉴别,这是术前常规胸部增强 CT 扫描在辅助临床分期之外的附加临床价值,可为临床治疗决策提供参考。

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[J]. CA Cancer J Clin, 2018, 68(6): 394-424.
- [2] Harbeck N, Gnant M. Breast cancer [J]. Lancet, 2017, 389(10074): 1134-1150.
- [3] Arvold ND, Taghian AG, Niemierko A, et al. Age, breast cancer subtype approximation, and local recurrence after breast-conserving therapy[J]. J Clin Oncol, 2011, 29(29): 3885-3891.
- [4] Reis-Filho JS, Tutt AN. Triple negative tumours: a critical review [J]. Histopathology, 2008, 52(1): 108-118.
- [5] Denkert C, Liedtke C, Tutt A, et al. Molecular alterations in triple-negative breast cancer-the road to new treatment strategies [J]. Lancet, 2017, 389(10087): 2430-2442.
- [6] Guo Y, Hu Y, Qiao M, et al. Radiomics analysis on ultrasound for prediction of biologic behavior in breast invasive ductal carcinoma [J]. Clin Breast Cancer, 2018, 18(3): 335-344.
- [7] 高微波, 朱海涛, 孙应实. 乳腺癌影像基因组学研究现状与进展 [J]. 中华放射学杂志, 2017, 51(12): 990-992.
- [8] Valdora F, Houssami N, Rossi F, et al. Rapid review: radiomics and breast cancer [J]. Breast Cancer Res Treat, 2018, 169(2): 217-229.
- [9] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data[J]. Radiology, 2016, 278(2): 563-577.
- [10] Park S, Koo JS, Kim MS, et al. Characteristics and outcomes according to molecular subtypes of breast cancer as classified by a panel of four biomarkers using immunohistochemistry [J]. Breast, 2012, 21(1): 50-57.
- [11] 马文娟, 季宇, 郝玉娟, 等. 乳腺癌 X 线摄影的影像组学特征 [J]. 中国医学影像学杂志, 2018, 26(7): 515-518.
- [12] 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.
- [13] Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach [J]. Nat Commun, 2014, 6(5): 4006.
- [14] Liang C, Huang Y, He L, et al. The development and validation of a CT-based radiomics signature for the preoperative discrimination of stage I ~ II and stage III ~ IV colorectal cancer [J]. Oncotarget, 2016, 7(21): 31401-31412.
- [15] Agner SC, Rosen MA, Englander S, et al. Computerized image analysis for identifying triple-negative breast cancers and differentiating them from other molecular subtypes of breast cancer on dynamic contrast-enhanced MR images: a feasibility study [J]. Radiology, 2014, 272(1): 91-99.
- [16] Ma W, Zhao Y, Ji Y, et al. Breast cancer molecular subtype prediction by mammographic radiomic features [J]. Acad Radiol, 2019, 26(2): 196-201.
- [17] Wang J, Kato F, Oyama-Manabe N, et al. Identifying triple-negative breast cancer using background parenchymal enhancement heterogeneity on dynamic contrast-enhanced MRI: A pilot radiomics study [J]. PLoS One, 2015, 10(11): e0143308.

(收稿日期: 2019-07-24 修回日期: 2019-08-16)