

医工结合研究专题

基于放射组学的特征选择和亚组分析 对肺腺癌患者预后的预测价值

焦玉泉, 常艳亮, 杨春媚, 王 翔

(山西省运城市中心医院 医学影像科, 山西 运城, 044000)

摘 要: **目的** 探讨基于放射组学的特征选择和亚组分析对肺腺癌患者预后的预测价值。**方法** 选取 293 例接受放疗的肺腺癌患者, 从胸部 CT 图像中提取 107 个放射学特征(14 个形状特征、18 个一阶统计特征和 75 个纹理特征)。分析 3 种不同的特征选择(FS)方法[即重测和多重分割(FS1)、皮尔逊相关分析(FS2)以及 FS1 和 FS2 相结合的方法(FS3)]对生存预测性能的影响。对各个 T 分期进行亚组分析, 采用一致性指数(C-index)和 Kaplan-Meier 法评估预后表现。亚组分析采用 5 倍交叉验证以确保模型的可靠性。**结果** 在放射学模型的训练和测试数据集中, FS2 的 C-index 是所有选择方法中最高的(分别为 0.64、0.61)。同样, FS2 在组合模型的训练和测试数据集中的所有选择方法中具有最高的 C-index(分别为 0.65、0.63)。亚组分析表明, 基于 T 分期的预测模型对测试数据集的 C-index 高于基于全数据的预测模型。**结论** 特征选择方法在一定程度上提高了生存预测的性能, 基于 T 分期的亚组预测模型可以提高预测性能。

关键词: 放射组学; 特征选择; 预后分析; 亚组分析; 肺腺癌

中图分类号: R 734.2; R 811 文献标志码: A 文章编号: 1672-2353(2022)02-108-05 DOI: 10.7619/jcmp.20213890

Value of feature selection and subgroup analysis based on radiomics in predicting prognosis of patients with lung adenocarcinoma

JIAO Yuquan, CHANG Yanliang, YANG Chunmei, WANG Xiang

(Department of Medical Imaging, Yuncheng City Central Hospital of Shanxi Province, Yuncheng, Shanxi, 044000)

Abstract: Objective To explore the value of feature selection and subgroup analysis based on radiomics in predicting prognosis of patients with lung adenocarcinoma. **Methods** A total of 293 lung adenocarcinoma patients with radiotherapy were selected, and 107 radiological features (14 shape features, 18 first-order statistical features and 75 texture features) were extracted from chest CT images. The effects of three different feature selection (FS) methods[retest and multiple segmentation (FS1), Pearson correlation analysis (FS2) and the combination of FS1 and FS2 (FS3)] on survival prediction performance were analyzed. Subgroup analysis was performed for each T stage, and the prognostic performance was evaluated by consistency index (C-index) and Kaplan-Meier method. Quintuple cross validation was used in subgroup analysis to ensure the reliability of the model. **Results** In the training and test data sets of radiology model, the C-index of FS2 was the highest among all the selection methods (values were 0.64 and 0.61 respectively). Similarly, FS2 showed the highest C-index (values were 0.65 and 0.63, respectively) among all the selection methods in the training and test data sets of the combined model. Subgroup analysis showed that the C-index of the prediction model based on T stage was higher than that based on the full data. **Conclusion** Feature selection method improves the performance of survival prediction to a certain extent, and the subgroup prediction model based on T stage can improve the prediction performance.

Key words: radiomics; feature selection; prognosis analysis; subgroup analysis; lung adenocarcinoma

肺腺癌是非小细胞肺癌的病理亚型之一,是导致全球癌症性死亡的主要原因。尽管多年来肺

腺癌的治疗决策和预后有了明显的改善,但生存率仍有待进一步提高^[1]。肿瘤-结节-转移(TNM)分期系统是目前常用的预后评价体系^[2]。然而,即便同一分期的患者之间的生存率也可能有所不同^[3]。目前,有研究^[4]开始利用放射组学预测肺癌患者的预后。放射组学是一种高通量技术,用于量化医学图像中的表型特征,这些特征可能有助于预测生存预后、术前远处转移和进行组织学亚型的分类^[5]。近年来,研究^[6-9]显示放射组学可作为总体生存率的独立负性预测因素预测肺腺癌预后,且代表异质性的纹理特征对于准确性至关重要。目前对同一病理类型进行不同分期的亚组分析的研究较少^[10-12]。本研究探讨不同放射组学特征选择和基于分期的亚组分析对肺腺癌患者预后的预测价值,现将结果报告如下。

1 资料与方法

1.1 一般资料

选取2016年1月—2018年1月在本院接受治疗的293例肺腺癌患者为研究对象。纳入标准:①病理诊断为肺腺癌者;②影像学资料完整者;③未合并其他恶性肿瘤者。排除标准:①病理诊断不明确者;②全身重要脏器功能衰竭者。将患者分为训练组235例和测试组58例,患者一般资料见表1。

表1 2组患者一般资料比较(x±)[n(%)]

一般资料		训练组 (n=235)	测试组 (n=58)
年龄/岁		70.23±16.34	68.44±15.87
性别	男	195(82.99)	47(81.03)
	女	40(17.02)	11(18.97)
T分期	T ₁	74(31.49)	19(32.20)
	T ₂	77(32.77)	19(32.76)
	T ₃	39(16.60)	10(17.24)
	T ₄	45(19.15)	10(17.24)
	N ₀	87(37.02)	22(37.93)
N分期	N ₁	29(12.34)	7(12.07)
	N ₂	82(34.89)	20(34.48)
	N ₃	37(15.74)	9(15.52)
	M ₀	201(85.53)	50(86.21)
M分期	M ₁	34(14.47)	8(13.79)
	I期	66(28.09)	17(29.31)
	II期	20(8.51)	5(8.62)
	III期	116(49.36)	28(48.28)
化疗状态	是	107(45.53)	26(44.83)
	否	128(54.47)	32(55.17)
生存时间/d		591.38±156.21	595.16±153.47
生存状态	生存	101(42.97)	24(41.38)
	死亡	134(57.02)	34(58.62)

1.2 放射组学特征提取与选择

每例患者共提取107个特征,其中包括14个形状特征、18个一阶统计学特征和75个纹理特征。形状特征量化了感兴趣区域(ROI)的直径、体积以及不规则程度;一阶统计特征用于创建像素值的直方图,并定义与该直方图有关的特征;纹理特征用于将像素值之间的关系转换为矩阵,以衡量图像的均匀性和异质性。纹理特征包括灰度共生矩阵、灰度依赖矩阵、灰度运行长度矩阵、灰度大小矩阵和邻域灰度差异矩阵。

107个从肿瘤区中提取的放射学特征用3种特征选择方法进行了还原,分别为:①特征选择1(FS1)为文献报道的测试-重复测试和多重分割来选择稳健的特征^[13-15]。使用文献报道的数据集来评估当天重复CT扫描的肿瘤单维、二维和体积测量的变异性。该数据集可从癌症成像档案馆公开的在线参考图像数据库评估治疗反应测试-重测数据集中下载。测试-重测法将肿瘤的放射学分析应用于每例患者的2张图像,并用一致性指数(C-index)评估2个特征值之间的一致性,在C-index>0.85时进行特征选择。②特征选择2(FS2)是根据所有特征的皮尔逊相关分析计算出的相关系数,将其中1个相关的特征作为多余的特征从分析中排除^[12]。相关系数的绝对值≥0.8提示2个特征之间强相关的阈值^[16]。③特征选择3(FS3)是结合FS1和FS2^[17],在使用测试检验和多重分割选择出稳健的特征后,使用皮尔逊相关分析选择非冗余的特征,阈值为0.8。

1.3 构建LASSO Cox回归模型

本实验构建了2个不同的模型:一个是使用FS1、FS2和FS3中选定的特征的单纯放射学特征模型(放射学模型),另一个是使用选定的特征加上临床预测因子的放射学和临床特征相结合的模型(组合模型)。LASSO Cox回归模型被用来构建预测生存预后的模型。作者采用了5倍交叉验证来防止模型的简化和过拟合,并为数据选择最佳的λ。

1.4 亚组分析

肺腺癌患者依据不同的T分期创建亚组数据集。T₁~T₄期分别有93、96、49和55例患者。亚组分析中使用5倍交叉验证来验证构建的模型。使用分层抽样将每个亚组数据集分为5个部分,同时保持死亡患者和存活患者的比率不变;然后,将4个部分作为训练数据集,1个部分作为

测试数据集。采用 5 倍交叉验证,以确保用数据数量较少的子组数据集构建的模型的可靠性。

1.5 统计学分析

采用 R 软件 3.6.1 进行数据的统计分析, Kaplan-Meier 生存分析评估放射组学特征和生存率之间的关系。 $P<0.05$ 为差异有统计学意义。

2 结果

2.1 不同模型的预测性能

在放射学模型的训练和测试数据集中, FS2 的 C-index 是所有选择方法中最高的(分别为 0.64、0.61)。同样, FS2 在组合模型的训练和测

试数据集中的所有选择方法中具有最高的 C-index(分别为 0.65、0.63)。因此, FS2 被应用于亚组分析,见表 2。

2.2 基于 T 分期的亚组预测性能

表 3 显示了各亚组的预后表现,为了避免复杂化,显示了在 5 倍交叉验证中最接近测试数据集的平均 C-index 时的情况。所有 T 分期放射学模型和组合模型产生的 C-index 都高于所有数据,特别是在放射学模型中, T₁ 组的测试数据集的 C-index 提高最多,组合模型中 T₄ 组的指数提高最多。见图 1。

表 2 训练组和测试组中不同模型的预测性能

模型构建		特征总数	训练数据集		测试数据集	
			C-index	HR(95% CI)	C-index	HR(95% CI)
放射学模型	FS1	23	0.63	1.55(1.30~1.85)	0.60	0.95(0.82~1.10)
	FS2	28	0.64	3.96(2.43~6.45)	0.61	1.87(0.88~3.99)
	FS3	9	0.62	1.84(0.17~2.19)	0.60	1.06(0.01~2.08)
组合模型	FS1+临床	31	0.64	2.22(0.58~3.22)	0.62	1.20(0.45~2.87)
	FS2+临床	36	0.65	4.75(2.99~7.56)	0.63	2.24(1.13~4.36)
	FS3+临床	17	0.64	2.62(0.90~3.96)	0.62	0.94(0.19~2.32)

表 3 各亚组预后预测表现($\bar{x}\pm s$)

组别	放射学模型		组合模型	
	训练模型	测试模型	训练模型	测试模型
全组($n=293$)	0.63±0.01	0.62±0.03	0.65±0.01	0.64±0.04
T ₁ 组($n=93$)	0.66±0.03	0.66±0.04	0.70±0.02	0.68±0.03
T ₂ 组($n=96$)	0.64±0.03	0.63±0.05	0.66±0.02	0.65±0.02
T ₃ 组($n=49$)	0.68±0.02	0.65±0.03	0.68±0.04	0.66±0.06
T ₄ 组($n=55$)	0.65±0.02	0.63±0.04	0.72±0.02	0.70±0.06

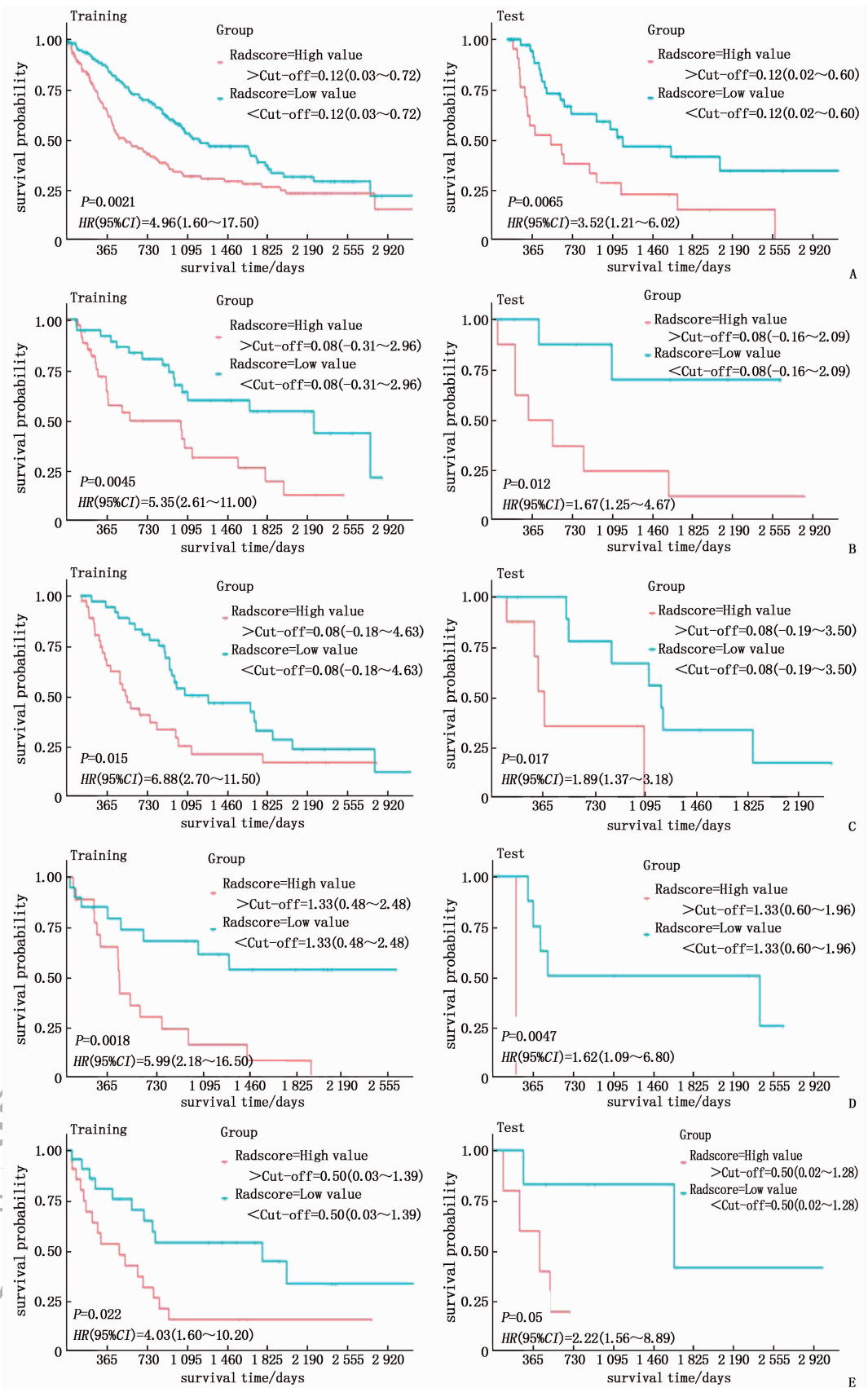
3 讨论

本研究探讨了基于放射组学特征选择和 T 分期的亚组数据集的肺腺癌患者的生存预测,对所有数据的分析并未显示出高的预后性能。然而,亚组的分析比所有数据的分析有更好的预后性能。这一结果表明,按特定的 T 分期分析肺腺癌可明显改善生存预测。

本研究对所有数据应用了 3 种独立的特征选择方法,以确定预后性能方面的最佳方法。在测试数据集中,放射组和组合模型使用 FS2 时获得最高的预后性能。SUN W 等^[18]研究表明, Cox 模型中 Pearson 的特征选择方法在 5 种选择方法中产生了第 2 高值的 C-index。LEGER S 等^[19]研究也表明, Cox 模型中同样的 Pearson 特征选择

方法在 12 种选择方法中产生了最高的 C-index。这些结果与本研究结果一致(即 FS2 的 C-index 最高)。因为 Cox 模型用一个简单的回归方程直接预测事件发生的时间,这个模型经常产生过拟合的结果^[19]。Pearson 的特征选择方法可以通过去除多余的特征交互作用来减少过拟合,而且计算效率很高。基于 Cox 模型和 Pearson 特征选择方法的这些特点,可以说明这种选择方法是 Cox 模型预测预后的最有用的方法。

既往研究^[20]通过将每个子组数据集应用于整个数据集训练模型来验证预测性能,与应用整个数据集相比,预测性能得到提高。然而,这些研究没有使用子组数据集构建训练模型,只进行了最小的子组分析。本研究同时构建全部数据和子组模型进行肺腺癌患者预后分析。结果显示,



A: 总数据的训练模型和生存模型曲线; B: T₁ 期的训练模型和生存模型曲线; C: T₂ 期的训练模型和生存模型曲线; D: T₃ 期的训练模型和生存模型曲线; E: T₄ 期的训练模型和生存模型曲线。

图 1 Kaplan-Meier 曲线基于每个亚组的放射模型中的 Rad 得分

与所有数据相比,基于 T 分期亚组分析的预后性能提高。高预后性能是通过消除具有不同预后和异质性的亚组之间的放射学特征趋势的差异而产生的。因此,本研究采用的方法,即为每个亚组构建训练模型,可以准确反映每个组的放射学特征,并可以提高预后预测的性能。

本研究旨在使用一种方法来实现足够的预后性能,以实现临床效用,该方法侧重于对具有相同特征的亚组进行预后分析。然而,在未来的临床应用之前,有一个问题必须解决,即使用免疫检查点抑制剂和分子靶向药物的治疗方法的出现,这些疗法大大改善了肺癌患者的预后^[21]。因此,有必要建立一个考虑这些因素的预后模型。最近,在接受这些疗法治疗的患者的数据集中,已有研究^[22]报道了与放射组学的高度关联和高度预后预测的潜力。未来的一个挑战是揭示该模型是否能适用于接受过上述治疗的患者的数据。此外,本研究是基于相对较少的患者数量,由于一些亚组的数据数量相当少,因此本研究得到的结果需要基于更多数据的研究来进一步验证。

综上所述,本研究调查了基于放射组学的特征选择和 T 分期的亚组分析对肺腺癌患者的生存预测,基于每个 T 分期组的模型较基于所有数据的模型具有更高的 C-index。因此,对特定亚组的预后分析或可改善预后性能。

参考文献

- [1] MOLINA J R, YANG P, CASSIVI S D, *et al.* Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship[J]. *Mayo Clin Proc*, 2008, 83(5): 584–590.
- [2] LIANG W, ZHANG L, JIANG G, *et al.* Development and validation of a nomogram for predicting survival in patients with resected non-small-cell lung cancer[J]. *J Clin Oncol*, 2015, 33(8): 861–869.
- [3] RAMI-PORTA R, ASAMURA H, GOLDSTRAW P. Predicting the prognosis of lung cancer: the evolution of tumor, node and metastasis in the molecular age-challenges and opportunities[J]. *Transl Lung Cancer Res*, 2015, 4(4): 415–423.
- [4] AERTS H J, VELAZQUEZ E R, LEIJENAR R T, *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach[J]. *Nat Commun*, 2014, 5: 4006–4015.
- [5] GILLIES R J, KINAHAN P E, HRICAK H. Radiomics: Images Are More than Pictures, They Are Data[J]. *Radiology*, 2016, 278(2): 563–568.
- [6] LIU J, CUI J, LIU F, *et al.* Multi-subtype classification model for non-small cell lung cancer based on radiomics: SLS model[J]. *Med Phys*, 2019, 46(7): 3091–3098.
- [7] ZHU X, DONG D, CHEN Z, *et al.* Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer[J]. *Eur Radiol*, 2018, 28(7): 2772–2784.
- [8] ABEL S, HASAN S, WHITE R, *et al.* Stereotactic ablative radiotherapy (SABR) in early stage non-small cell lung cancer: Comparing survival outcomes in adenocarcinoma and squamous cell carcinoma[J]. *Lung Cancer*, 2019, 128: 127–135.
- [9] FUKUI T, TANIGUCHI T, KAWAGUCHI K, *et al.* Comparisons of the clinicopathological features and survival outcomes between lung cancer patients with adenocarcinoma and squamous cell carcinoma[J]. *Gen Thorac Cardiovasc Surg*, 2015, 63(9): 507–517.
- [10] YANG L, YANG J, ZHOU X, *et al.* Development of a radiomics nomogram based on the 2D and 3D CT features to predict the survival of non-small cell lung cancer patients[J]. *Eur Radiol*, 2019, 29(5): 2196–2205.
- [11] CHADDAD A, DESROSIERS C, TOEWS M, *et al.* Predicting survival time of lung cancer patients using radiomic analysis[J]. *Oncotarget*, 2017, 8(61): 104393–104399.
- [12] YANG J, WU Q, XU L, *et al.* Integrating tumor and nodal radiomics to predict lymph node metastasis in gastric cancer[J]. *Radiother Oncol*, 2020, 150: 89–95.
- [13] KADOYA N, TANAKA S, KAJIKAWA T, *et al.* Homology-based radiomic features for prediction of the prognosis of lung cancer based on CT-based radiomics[J]. *Med Phys*, 2020, 47(5): 2197–2203.
- [14] VAN BAARDWIJK A, BOSMANS G, BOERSMA L, *et al.* PET-CT-based auto-contouring in non-small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumor and involved nodal volumes[J]. *Int J Radiat Oncol Biol Phys*, 2007, 68(3): 771–786.
- [15] MORI M, PASSONI P, INCERTI E, *et al.* Training and validation of a robust PET radiomic-based index to predict distant-relapse-free-survival after radio-chemotherapy for locally advanced pancreatic cancer[J]. *Radiother Oncol*, 2020, 153: 258–264.
- [16] LI H, ZHANG R, WANG S, *et al.* CT-Based Radiomic Signature as a Prognostic Factor in Stage IV ALK-Positive Non-small-cell Lung Cancer Treated With TKI Crizotinib: A Proof-of-Concept Study[J]. *Front Oncol*, 2020, 10: 57–65.
- [17] LIN P, YANG P F, CHEN S, *et al.* A Delta-radiomics model for preoperative evaluation of Neoadjuvant chemotherapy response in high-grade osteosarcoma [J]. *Cancer Imaging*, 2020, 20(1): 7–12.
- [18] SUN W, JIANG M, DANG J, *et al.* Effect of machine learning methods on predicting NSCLC overall survival time based on Radiomics analysis[J]. *Radiat Oncol*, 2018, 13(1): 197–204.
- [19] LEGER S, ZWANENBURG A, PILZ K, *et al.* A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling[J]. *Sci Rep*, 2017, 7(1): 13206–13212.
- [20] ZHOU X, YI Y, LIU Z, *et al.* Radiomics-Based Preoperative Prediction of Lymph Node Status Following Neoadjuvant Therapy in Locally Advanced Rectal Cancer [J]. *Front Oncol*, 2020, 10: 604–611.
- [21] OUYANG M L, XIA H W, XU M M, *et al.* Prediction of occult lymph node metastasis using SUV, volumetric parameters and intratumoral heterogeneity of the primary tumor in T₁₋₂ N₀M₀ lung cancer patients staged by PET/CT[J]. *Ann Nucl Med*, 2019, 33(9): 671–682.
- [22] LIU C, GONG J, YU H, *et al.* A CT-Based Radiomics Approach to Predict Nivolumab Response in Advanced Non-Small-Cell Lung Cancer[J]. *Front Oncol*, 2021, 11: 544339.

(本文编辑:梁琥)