

基于影像组学和剂量组学的放射性肺炎预测研究

周露,王琳婧,张国前,李慧君,廖煜良,吴书裕
广州医科大学附属肿瘤医院放疗科,广东 广州 510095

【摘要】目的:旨在利用影像组学和剂量组学的多组学方法,建立并验证一个有效的基于CT图像的放射性肺炎(RP)预测模型。**方法:**对2019年至2021年在广州医科大学附属肿瘤医院接受放疗的91例非小细胞肺癌患者进行回顾性分析。将除去临床靶区的全肺(Lung-CTV)作为感兴趣区域,从Lung-CTV区域的CT图像和剂量分布中提取影像组学和剂量组学特征。将单独的剂量体积直方图(DVH)特征、影像组学结合DVH(radio+DVH)特征、影像组学结合剂量组学(radio+dose)特征,分别输入11个不同的分类器来构建预测模型,采用五倍交叉验证法来完成分类实验。利用接受者操作特征(ROC)曲线下的面积(AUC)、准确性、精确性、召回率和F1值来评估预测模型的性能。**结果:**与DVH模型相比,radio+DVH和radio+dose的AUC值更高,差异有统计学意义($P<0.05$)。与DVH和radio+DVH模型相比,radio+dose的准确率和F1值更高,差异有统计学意义($P<0.05$)。**结论:**使用基于机器学习的影像组学和剂量组学的多组学方法预测RP的性能更好,有望为临床治疗提供指导。

【关键词】非小细胞肺癌;影像组学;剂量组学;放射性肺炎;机器学习

【中图分类号】R312;R818.7

【文献标志码】A

【文章编号】1005-202X(2023)07-0808-06

Prediction of radiation pneumonitis based on radiomics and dosiomics

ZHOU Lu, WANG Linjing, ZHANG Guoqian, LI Huijun, LIAO Yuliang, WU Shuyu

Department of Radiation Oncology, Affiliated Cancer Hospital and Institute of Guangzhou Medical University, Guangzhou 510095, China

Abstract: Objective To establish and validate an effective CT-based radiation pneumonitis (RP) prediction model using the multi-omics approach combining radiomics with dosiomics. **Methods** A retrospective analysis was carried out on 91 non-small-cell lung cancer patients who were treated with radiotherapy from 2019 to 2021 in Affiliated Cancer Hospital and Institute of Guangzhou Medical University. The whole lung excluding clinical target volume (Lung-CTV) was taken as the region of interest (ROI), and the radiomics and dosiomics features were extracted from the CT image and dose distribution of Lung-CTV. The dose-volume histogram (DVH) features, radiomics combined with DVH (radio+DVH) features, and radiomics combined with dosiomics (radio+dose) features were imported into 11 different classifiers for constructing prediction models. The 5-fold cross-validation was used to complete the classification experiment. The area under the curve (AUC) of the receiver operating characteristics (ROC), accuracy, precision, recall and F1-score were calculated to assess the model performances. **Results** The AUC of radio+DVH and radio+dose was higher than that of DVH model ($P<0.05$); and radio+dose had higher accuracy and F1-score than DVH and radio+DVH ($P<0.05$). **Conclusion** The multi-omics approach using machine learning-based radiomics and dosiomics to predict RP exhibits better performance and is expected to provide guidance for clinical treatment.

Keywords: non-small-cell lung cancer; radiomics; dosiomics; radiation pneumonitis; machine learning

前言

【收稿日期】2023-04-07

【基金项目】广东省医学科研基金(A2023291)

【作者简介】周露,工程师,研究方向:肿瘤放射物理,E-mail: zhou-lu6105189@126.com

【通信作者】吴书裕,工程师,研究方向:肿瘤放射物理,E-mail: ws-yeasy@outlook.com

放射性肺损伤是胸部肿瘤放疗的主要限制因素和最常见的并发症,其发生率和严重程度与放疗剂量、治疗时间和肺部解剖结构等因素有关。明确肺癌胸部放疗后发生放射性肺炎(RP)的相关危险因素,指导放疗计划的制定及高危患者的提前预防,是降低此放疗并发症的关键。孙师等^[1]回顾性分析256例局部晚期非小细胞肺癌实施调强放射治疗(IMRT)的资料,进行多因素分析证实肺平均剂量、双肺

$V_5 \sim V_{20}$ 是 ≥ 2 级放射性肺损伤的独立影响因素。临床上大多采用 V_{30} 、 V_{20} 和 V_5 等物理剂量参数来预防 RP, 然而肺对放射治疗的敏感性在不同患者群体中有很大的差异, 决定患者是否患 RP 的因素是多样化的, 必须具体病人具体分析。近年来, 随着影像组学、数字图像处理和机器学习技术的发展, 可以将其应用于肺部疾病的预测和诊断方面^[2-4]。与其假设辐射所致的肺损伤在不同患者之间是一致的, 倒不如利用基于图像的方法来测量每个病人对辐射的个别反应。Cunliffe 等^[5]证明影像组学能够量化、个体化地测量患者肺组织对放疗的反应, 并评估放射性肺损伤的发展。也有一些研究利用整个肺的影像组学特征, 结合物理剂量参数和临床治疗特征来训练 RP 的预测模型^[6-12]。由于基于剂量体积直方图 (DVH) 的物理剂量参数仅利用剂量分布的部分信息, 没有关于体素的空间关系信息, 相同的物理剂量参数值可能来自具有不同空间关系的剂量分布, 所产生的生物效应可能不同^[13-15]。本研究拟将影像组学框架应用于剂量分布, 从剂量分布中提取剂量组学特征, 并与影像组学特征结合起来建立基于多模态组学的 RP 预测模型。

1 资料与方法

1.1 临床资料

本研究回顾性地收集了 2019 年至 2021 年在广州医科大学附属肿瘤医院接受放疗的 350 例诊断为非小细胞肺癌的患者, 其中 91 例患者符合入选标准。纳入标准: (1) 所有患者都经病理诊断为非小细胞肺癌, 并且有明确的肺癌 TNM 分期 (I、II、III、IV 期); (2) 每例患者都接受了两次高分辨率的诊断性 CT 扫描, 第一次是在放疗前, 第二次是在放疗后 6 个月内, 期间没有手术干预; (3) 所有患者行 IMRT, 放疗计划采用 Pinnacle 治疗计划系统 (V9.10) 设计。计划 CT 图像的切片间距为 5 mm, 剂量计算的网格间距为 3 mm×3 mm×3 mm。RP 的分级是根据 RTOG 急性放射性肺损伤的分级标准进行的。本研究的终点是 ≥ 1 级的 RP。患者的临床和治疗特征见表 1。

1.2 DVH 参数、影像组学和剂量组学特征提取以及预处理

本研究的流程如图 1 所示, 首先, 利用自行开发的放射治疗计划分析软件, 从治疗计划系统 Pinnacle 中获得除去临床靶区的全肺 (Lung-CTV) 的三维体积和原始三维剂量分布。在临床中, 常利用剂量大于 x Gy (V_x) 的肺体积参数来预防 RP 的发生。本研究提取所有患者的 V_{30} 、 V_{20} 和 V_5 参数作为 DVH 特征。将影像组学框架应用于三维剂量分布, 分别从 CT 图像和三

表 1 91 例肺癌患者的临床和治疗特征
Table 1 Clinical and treatment characteristics of 91 patients with lung cancer

临床和治疗特征	数值
年龄/岁	62.5(31~85)
性别[例(%)]	
男	83(91.21)
女	8(8.79)
吸烟史[例(%)]	
有	54 (59.34)
无	37 (40.66)
吸烟年数[例(%)]	
<30 年	8 (8.79)
≥ 30 年	46 (50.55)
肿瘤分期[例(%)]	
I	0
II	3(3.3)
III	63(69.23)
IV	25(27.47)
处方剂量/Gy	60(40~66)
分次剂量/Gy	2(1.50~2.75)
肺炎级别[例(%)]	
0	52(57.15)
1	24(26.37)
2	9(9.89)
3	6(6.59)
4	0

维剂量分布中提取 Lung-CTV 区域的影像组学和剂量组学特征。所有的特征都是由 Pyradiomics 3.0.1 提取的^[16], 分别提取 8 组特征 (共 851 个组学特征), 包括一阶统计 (18 个特征)、基于形状 (14 个特征)、灰度共生矩阵 (GLCM, 24 个特征)、灰度游程矩阵 (GLRLM, 16 个特征)、灰度区域大小矩阵 (GLSZM, 16 个特征)、灰度依赖矩阵 (GLDM, 14 个特征)、邻域灰度差矩阵 (NGTDM, 5 个特征), 并且使用小波滤波器对原始图像进行预处理后提取组学特征 (744 个特征)。

将患者数据随机分为两组, 其中 70% 为训练组, 30% 为测试组。使用合成少数过采样技术来防止训练集中由于数据不平衡导致的过拟合。随机森林算法是一种基于决策树的集成机器学习算法, 该算法具有分类精度高、运算速度快、能够从具有小边际效应和复杂相互作用的特征中识别主相关特征的优点, 在医学数据分析中得到广泛应用^[17]。本研究采

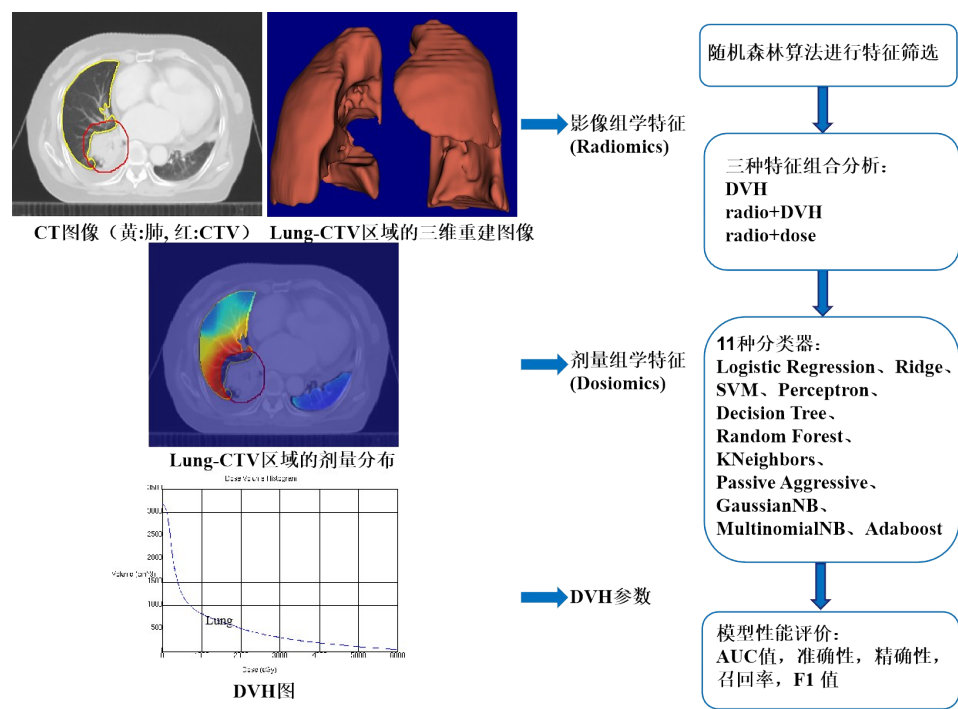


图 1 多组学放射性肺炎预测模型建立流程图

Figure 1 Construction of a multi-omics prediction model for radiation pneumonitis

用随机森林算法进行特征筛选,过滤冗余信息。使用五折交叉验证法来减少非线性回归中的过拟合。该方法可以有效地减少模型评估中的随机误差,提高评估结果的准确性。同时,也可以充分利用数据集,减少因数据分割不合理而引入的偏差。本研究利用验证集的 AUC 值、准确性、精确性、召回率和 F1 值来评估预测模型的性能。

1.3 模型构建

本研究利用 Python 3.7.6 和 scikit-learn 0.24.2 进行特征提取、数据预处理和机器学习建模。使用 11 个不同的分类器,包括 Logistic Regression、Ridge、SVM、Perceptron、Decision Tree、Random Forest、KNeighbors、Passive Aggressive、GaussianNB、MultinomialNB 和 Adaboost 来建立鉴别模型。将 DVH、影像组学结合 DVH(radio+DVH)特征、影像组学结合剂量组学(radio+dose)特征分别作为模型的输入,对鉴别模型进行训练。

1.4 统计学方法

利用 R 语言进行数据的统计分析,符合正态性分布的计量资料用均数±标准差表示,采用独立样本 *t* 检验,计数资料用率(%)表示。小提琴图的绘制采用 R 语言的 Vioplot 包。*P*<0.05 为差异有统计学意义。

2 结果

2.1 筛选的特征

采用随机森林算法能够筛选出大量特征数据中对于回归预测最重要的特征,并对各特征进行重要性排序。本研究分别从影像组学和剂量组学特征中按重要性选择 3 个特征。如表 2 所示,所选特征大多为肺部纹理特征,代表图像空间域中的周期性变化和结构规律,能充分反映组织的异质性。GLSZM 特征对图像中连续像素的区域进行量化,包括描述小面积或大面积、低灰度或高灰度区域分布的特征,极大地影响了纹理一致性、非周期性或斑点状纹理的表示,在图像纹理分析中具有很好的性能。

表 2 筛选的影像组学和剂量组学特征

Table 2 Selected radiomics and dosiomics features

类别	特征	重要性
影像组学	Wavelet-HHH_GLSZM_Gray Level	0.027
	Non Uniformity	
	GLSZM_Zone Entropy	0.026
	Wavelet-LLH_Firstorder_Variance	0.025
剂量组学	GLSZM_Small Area Low Gray Level Emphasis	0.041
	GLSZM_Size Zone Non Uniformity Normalized	0.031
	GLSZM_Zone Entropy	0.025

2.2 模型性能

图 2 和表 3 是 11 个分类器下 3 种不同特征组合模

型的性能统计分析,统计分析是为了证明本研究提出的方法性能不会受不同分类器的影响。本研究利用测试集的 AUC 值、准确性(Accuracy)、精确性(Precision)、召回率(Recall)和F1 值来评价。与 DVH 模型相比,radio+DVH 和 radio+dose 的 AUC 值更高,且差异有统计学意义($P<0.05$),这证明相对于传统的剂量体积参数,利用反映细节特征的组学模型可以提高对 RP 的预测能力。与 DVH 和 radio+DVH 模型相比,radio+dose 的准确率更高,差异有统计学意义

($P<0.05$),说明 radio+dose 模型能更准确地预测患者发生 RP 的风险。精确率反映了模型区分阴性样本的能力,召回率展示了模型识别阳性样本的能力,而F1 值同时考虑了分类模型的精确率和召回率。与 DVH 和 radio+DVH 模型相比,radio+dose 模型的 F1 值更高,且差异有统计学意义($P<0.05$)。因此,相对于反映统计剂量的 DVH 参数,利用了剂量空间分布信息的剂量组学特征能够更好地预测 RP。

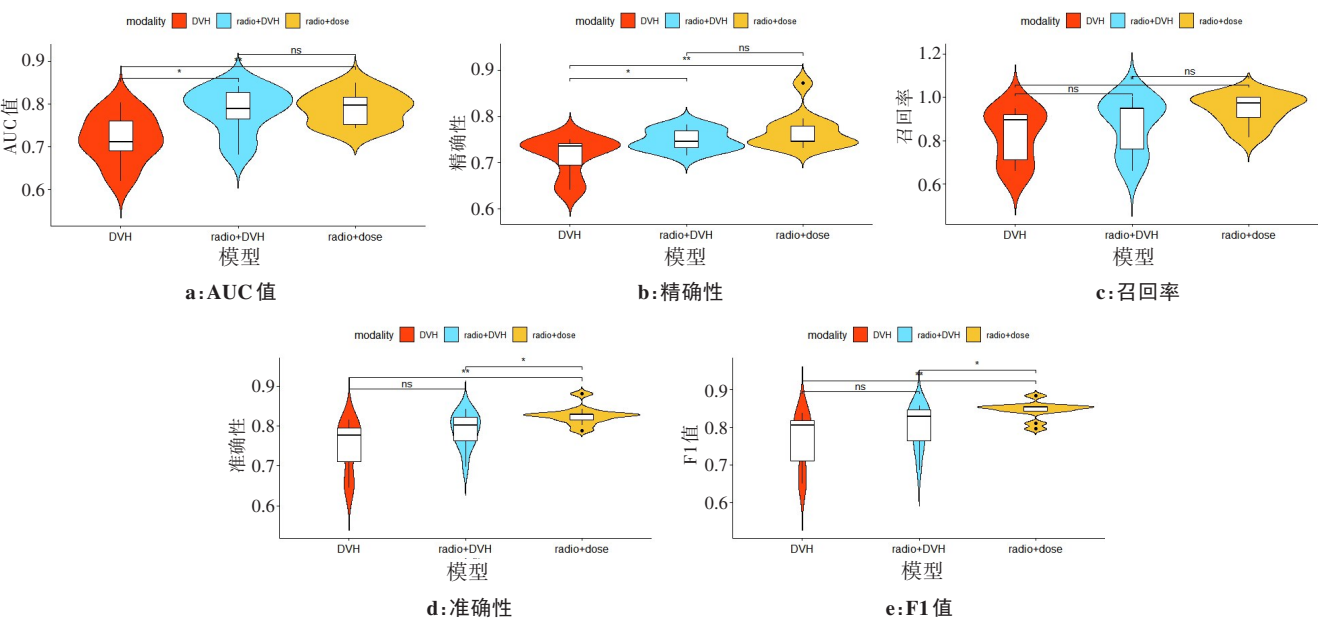


图2 3种特征组合模型在11个分类器下的性能指标统计(* $P<0.05$, ** $P<0.01$, *** $P<0.005$, ns $P>0.05$)

Figure 2 Performances of 3 feature combination models using 11 classifiers (* $P<0.05$, ** $P<0.01$, *** $P<0.005$, ns $P>0.05$)

表3 11个分类器下3种特征组合模型的评价指标统计($\bar{x} \pm s$)

Table 3 Statistics of evaluation indicators of 3 feature combination models using 11 classifiers (Mean±SD)

模型	AUC 值	准确性	精确性	召回率	F1 值
DVH	0.718±0.056	0.748±0.064	0.714±0.041	0.822±0.121	0.762±0.073
radio+DVH	0.784±0.053	0.789±0.042	0.748±0.022	0.873±0.124	0.802±0.056
radio+dose	0.789±0.037	0.826±0.023	0.768±0.040	0.942±0.068	0.844±0.023

3 讨论

在临床中找到影响 RP 的因素以及影响的程度非常重要,能够提前预估 RP 的发生几率,指导临床工作,尽量避免 RP 的发生或者做好预防措施。有关 RP 发生相关因素的研究很多,大多数研究已将 V_{20} 和平均肺部剂量做为最常见的预测参数,同时发现除了肺部剂量外,心脏剂量参数也有相关性。Liang 等^[18-19]相关研究证明相对于传统的正常组织控制率模型及 DVH 模型,剂量组学可以提高对 RP 的预测能

力。以往的研究大多单独采用 DVH 参数、影像组学、剂量组学或者影像组学结合 DVH 参数等进行 RP 的预测。本研究将影像组学和剂量组学特征结合起来,利用随机森林方法筛选相关性强的特征来进行预测。本研究筛选出来的特征大多是纹理特征,其中 GLSZM 特征描述了剂量分布纹理的区域一致性,该计数矩阵保存了感兴趣区域中所有灰度和剂量的连通域大小和个数信息^[20-22]。Wavelet-HHH_GLSZM_Gray Level Non Uniformity 特征是指图像中灰度级分布不均匀的程度,值越大表示灰度

级分布越不均匀。Wavelet-LLH_Firstorder_Variance 主要描述图像中像素灰度值的差异程度,在影像组学中,该特征常用于描述肿瘤组织与正常组织的灰度分布差异。影像组学可以提取与RP相关的特征信息,如肺部病变的形态、大小、密度等。GLSZM_Small Area Low Gray Level Emphasis特征测量的是剂量分布中具有较低剂量值的小连通区域的联合分布比例,值越大表示低剂量的小连通区域分布越密集。GLSZM_Size Zone Non Uniformity Normalized特征衡量了整个剂量分布中归一化区域大小不均匀性,数值越低表示区域之间剂量越均匀。GLSZM_Zone Entropy 计算了每个灰度和剂量级别的区域大小和出现次数,并将它们作为概率分布计算熵值,熵值越高表示图像中的灰度级别和剂量分布越均匀,纹理特征越复杂^[23]。因此,本研究推测RP的发生可能与肺组织的低剂量区域以及局部剂量变化有关。Liang等^[24]研究表明,同侧肺内局部剂量变化越大,总肺的低剂量区越大,RP发生的概率越大。DVH参数,如 V_5 、 V_{20} 和 V_{30} ,只是特定体积的肺组织的累积剂量,不能反映局部剂量的变化。然而,纹理特征显示了有关剂量分布空间关系的信息,可用于放疗反应预测^[25]。将影像组学和剂量组学结合起来,可以提取更多更全面特征信息,充分利用它们的优势,实现更准确的预测。

本研究中,相比于传统DVH模型,radio+DVH和radio+dose的AUC值更高,且差异有统计学意义($P<0.05$)。由于影像组学可以直观地反映肺部病变的情况,剂量组学可以直观反映肺部的受照剂量和剂量分布的空间关系,因此多组学方法的优势是显而易见的。但是相对于radio+DVH,radio+dose模型的AUC值没有表现出明显的优势,可能的原因是测试集数据相对较少,因此未来的研究将包括更多的数据,使性能评价更加客观。与DVH和radio+DVH模型相比,radio+dose的准确率和F1值更高,差异有统计学意义($P<0.05$)。首先说明radio+dose模型在预测RP方面的正确率更高,而且F1值是精确度和召回率的调和平均值,反映模型的稳健性。研究结果表明基于影像组学和剂量组学的多组学预测模型具有更准确、更全面、更可靠的优势,可为临床医生提供更好的参考信息,帮助患者更好地进行放射治疗。

本研究还有两个主要的局限性,一是符合研究要求的病人数量相对较少,为了避免由于设备型号、扫描条件、治疗计划等方面的一些误差,本研究只使用了本院的病人,此外,本院放疗后RP ≥ 2 级的病例较少(15例),因此本研究的终点是RP ≥ 1 级,这可能限制了其在实际临床中的作用。在后期研究中,有

必要在本院和多个中心招募更多的患者来验证本文提出的方法;二是本研究只考虑了放射学和剂量学特征,然而,RP还与一些临床特征高度相关,如吸烟史、疾病分期和肿瘤位置等,在接下来研究中,将增加临床特征以进一步讨论。为了预防RP的发生,如何利用放射学、剂量学和临床特征来指导放疗治疗计划的设计是接下来研究的目标。

【参考文献】

- [1] 孙帅,陈幸元,王静波,等. 256例局部晚期NSCLC调强放疗后放射性肺损伤预测因素分析[J]. 中华放射肿瘤学杂志, 2015, 24(5): 479-483.
Sun S, Chen XY, Wang JB, et al. Predictive factors for radiation-induced lung injury in locally advanced non-small cell lung cancer after intensity-modulated radiotherapy[J]. Chinese Journal of Radiation Oncology, 2015, 24(5): 479-483.
- [2] Krafft SP, Rao A, Stingo F, et al. The utility of quantitative CT radiomics features for improved prediction of radiation pneumonitis[J]. Med Phys, 2018, 45(11): 5317-5324.
- [3] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data[J]. Radiology, 2016, 278(2): 563-577.
- [4] Van Timmeren JE, Leijenaar RT, Van Elmpt W, et al. Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images[J]. Radiother Oncol, 2017, 123(3): 363-369.
- [5] Cunliffe A, Armato SG, Castillo R, et al. Lung texture in serial thoracic computed tomography scans: correlation of radiomics-based features with radiation therapy dose and radiation pneumonitis development[J]. Int J Radiat Oncol Biol Phys, 2015, 91(5): 1048-1056.
- [6] Zhang Z, Wang Z, Luo T, et al. Computed tomography and radiation dose images-based deep-learning model for predicting radiation pneumonitis in lung cancer patients after radiation therapy[J]. Radiother Oncol, 2023, 5(182).
- [7] 于佳琦,张臻,任凯,等. 利用放射组学模型预测肺癌和食管癌患者发生放射性肺炎的研究[J]. 中华放射肿瘤学杂志, 2021, 30(11): 1111-1116.
Yu JQ, Zhang Z, Ren K, et al. Study of application of radiomics model in predicting radiation pneumonitis in patients with lung cancer and esophageal cancer[J]. Chinese Journal of Radiation Oncology, 2021, 30(11): 1111-1116.
- [8] 孔燕,吴佳,魏贤顶,等. 肺癌放疗患者症状性放射性肺炎预测的CT影像组学研究[J]. 中华放射医学与防护杂志, 2022, 42(2): 115-120.
Kong Y, Wu J, Wei XD, et al. Application of CT radiomics analysis to predict symptomatic radiation pneumonitis for lung cancer[J]. Chinese Journal of Radiological Medicine and Protection, 2022, 42(2): 115-120.
- [9] Walls GM, Osman SO, Brown KH, et al. Radiomics for predicting lung cancer outcomes following radiotherapy: a systematic review[J]. Clin Oncol, 2022, 34(3): 107-122.
- [10] 闫萌,张臻,于佳琦,等. 基于剂量组学预测肺癌患者放射性肺炎发生的研究[J]. 中华放射肿瘤学杂志, 2022, 31(8): 698-703.
Yan M, Zhang Z, Yu JQ, et al. Dosimetrics-based prediction of incidence of radiation pneumonitis in lung cancer patients[J]. Chinese Journal of Radiation Oncology, 2022, 31(8): 698-703.
- [11] 张臻,赵路军,王伟,等. 基于放射组学预测放射性肺炎的初步研究[J]. 中华放射肿瘤学杂志, 2020, 29(6): 427-431.
Zhang Z, Zhao LJ, Wang W, et al. Preliminary study of predicting radiation pneumonitis based on radiomics technology[J]. Chinese Journal of Radiation Oncology, 2020, 29(6): 427-431.
- [12] Kraus KM, Oreshko M, Bernhardt D, et al. Dosimetrics and radiomics to predict pneumonitis after thoracic stereotactic body radiotherapy and immune checkpoint inhibition[J]. Front Oncol, 2023, 13: 1124592.
- [13] Adachi T, Nakamura M, Shintani T, et al. Multi-institutional dose-segmented dosimetric analysis for predicting radiation pneumonitis after lung stereotactic body radiation therapy[J]. Med Phys, 2021, 48(4): 1781-1791.

- [14] Ramella S, Trodella L, Mineo TC, et al. Adding ipsilateral V_{20} and V_{30} to conventional dosimetric constraints predicts radiation pneumonitis in stage IIIA-B NSCLC treated with combined-modality therapy[J]. Int J Radiat Oncol Biol Phys, 2010, 76(1): 110-115.
- [15] Boonyawan K, Gomez DR, Komaki R, et al. Clinical and dosimetric factors predicting grade ≥ 2 radiation pneumonitis after postoperative radiotherapy for patients with non-small cell lung carcinoma[J]. Int J Radiat Oncol Biol Phys, 2018, 101(4): 919-926.
- [16] Griethuysen JJ, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype[J]. Cancer Res, 2017, 77(21): 104-107.
- [17] Genuer R, Poggi JM, Tuleau-Malot C. VSURF: variable selection using random forests[J]. Pattern Recognit Lett, 2016, 31(14): 2225-2236.
- [18] Liang B, Tian Y, Su ZH, et al. A deep learning-based dual-omics prediction model for radiation pneumonitis[J]. Med Phys, 2021, 48(10): 6247-6256.
- [19] Liang B, Yan H, Tian Y, et al. Dosimetrics: extracting 3D spatial features from dose distribution to predict incidence of radiation pneumonitis[J]. Front Oncol, 2019, 9: 269.
- [20] Haralick R, Shanmugam K, Dinstein I. Textural features for image classification[J]. IEEE Trans Syst Man Cybernet, 1973, 3(6): 610-621.
- [21] Huynh E, Controller TP, Narayan V, et al. CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer[J]. Radiother Oncol, 2016, 120(2): 258-266.
- [22] Thibault G, Fertil B, Navarro C, et al. Texture indexes and gray level size zone matrix application to cell nuclei classification[C]//10th International Conference on Pattern Recognition and Information Processing. 2009.
- [23] Van Griethuysen JJ, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype[J]. Cancer Res, 2017, 77(21): 104-107.
- [24] Liang B, Tian Y, Chen XY, et al. Prediction of radiation pneumonitis with dose distribution: a convolutional neural network (CNN) based model[J]. Front Oncol, 2020, 9(1500).
- [25] Rossi L, Bijman R, Schillemans W, et al. Texture analysis of 3D dose distributions for predictive modeling of toxicity rates in radiotherapy[J]. Radiat Oncol, 2018, 129(3): 548-553.

(编辑:陈丽霞)