

## 基于CT的影像组学特征同临床物理剂量特征预测肺癌放疗放射性肺炎研究

陈文韬<sup>1,2</sup>, 孙磊<sup>1</sup>, 谭爱斌<sup>3</sup>, 唐世强<sup>2</sup>, 陈芬<sup>2</sup>, 肖建彪<sup>2</sup>, 王志芳<sup>2</sup>, 甄鑫<sup>1</sup>

1. 南方医科大学生物医学工程学院, 广东 广州 510515; 2. 郴州市第一人民医院放疗中心, 湖南 郴州 423000; 3. 郴州市第一人民医院北院门诊, 湖南 郴州 423000

**【摘要】目的:**探讨基于CT的影像组学特征同临床物理剂量特征预测肺癌放疗放射性肺炎研究。**方法:**回顾性收集2013年1月至2017年1月进行放射治疗的83例肺癌患者的临床物理剂量参数和CT影像以及随访数据。从病例的CT图像中提取107个影像组学特征,结合对应的45个临床物理剂量特征,每例病例共收集152个特征。基于22种特征提取算法和8种分类器构建的176个鉴别模型分析152个特征预测放射性肺炎的准确性以及筛选优势特征的能力。**结果:**临床物理剂量特征和影像组学特征预测放射性肺炎的鉴别模型中AUC值最高为0.90。前5位的优势特征是:shape\_Maximum2DDiameterColumn、shape\_Maximum3DDiameter、V<sub>20</sub>、glcm\_Imc1、V<sub>45</sub>。**结论:**临床物理剂量特征和影像组学特征通过不同分类器和特征选择算法组合的鉴别模型,可以筛选出理想的鉴别模型以及优势预测特征。

**【关键词】**肺癌;放射性肺炎;特征提取;影像组学;分类器

**【中图分类号】**R312;R818.7

**【文献标志码】**A

**【文章编号】**1005-202X(2021)06-0672-05

## Prediction of radiation pneumonia in lung cancer patients by CT-based radiomics signatures and clinical physical dosimetric features

CHEN Wentao<sup>1,2</sup>, SUN Lei<sup>1</sup>, TAN Aibin<sup>3</sup>, TANG Shiqiang<sup>2</sup>, CHEN Fen<sup>2</sup>, XIAO Jianbiao<sup>2</sup>, WANG Zhifang<sup>2</sup>, ZHEN Xin<sup>1</sup>

1. School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China; 2. Radiotherapy Center, Chenzhou No.1 People's Hospital, Chenzhou 423000, China; 3. Outpatient Department, North Hospital of Chenzhou No.1 People's Hospital, Chenzhou 423000, China

**Abstract: Objective** To combine CT-based radiomics signatures with clinical physical dosimetric features for predicting radiation pneumonitis in lung cancer patients. **Methods** The clinical physical dosimetric features, CT images and follow-up data of 83 patients with lung cancer who underwent radiotherapy from January 2013 to January 2017 were retrospectively collected. A total of 152 features, including 107 radiomics signatures extracted from the CT images and 45 clinical physical dosimetric features, were collected for each case. Based on 22 feature extraction methods and 8 classifiers, 176 identification models were constructed to analyze the accuracy of 152 features in predicting radiation pneumonia and to evaluate the ability to screen dominant features. **Results** The highest AUC in the identification model for predicting radiation pneumonitis by clinical physical dosimetric parameter combined with radiomics signatures was 0.90. The top 5 dominant features included shape\_Maximum2DDiameterColumn, shape\_Maximum3DDiameter, V<sub>20</sub>, glcm\_Imc1 and V<sub>45</sub>. **Discussion** The ideal identification model and superior prediction features can be screened from identification models constructed by the combination of different classifiers and feature selection algorithms based on clinical physical dosimetric features and radiomics signatures.

**Keywords:** lung cancer; radiation pneumonia; feature extraction; radiomics; classifier

**【收稿日期】**2021-04-08

**【基金项目】**国家自然科学基金(81874216);郴州市科技项目(jsyf2017030)

**【作者简介】**陈文韬,工程师,研究方向:肿瘤放射物理,E-mail: cwt198321@sina.com

**【通信作者】**甄鑫,副教授,研究方向:肿瘤放射物理,E-mail: xinzen@smu.edu.cn

### 前言

肺癌是常见的恶性肿瘤之一,其发病率和死亡率位居前列<sup>[1]</sup>。放射治疗是肺癌重要的治疗手段之一,特别是在不能够手术和局部晚期肺癌病人的治疗中发挥重要作用。放射性肺炎的存在会严重影响肺部肿瘤放疗效果,对治疗后患者生存时间的长短以及生活质量的好

坏有较大影响,所以降低放射性肺炎的发生率有重要的临床意义,而能够提前预测放射性肺炎的发生,就可以及时进行临床干预,降低放疗的风险<sup>[2]</sup>。

近年来随着机器学习、影像组学技术的发展,基于各种特征对疾病的预后预测越来越受到关注。本研究通过机器学习和影像组学方法对是否发生2级及以上放射性肺炎的肺癌患者的临床物理剂量学特征以及影像组学特征综合进行分析和模型构建<sup>[3]</sup>,旨在发现对预测放射性肺炎更为理想的预测模型以及相关优势因子<sup>[4]</sup>。

1 资料与方法

1.1 临床资料

回顾性收集郴州市第一人民医院2013年1月至2017年1月收治的肺癌患者83例,其中,男67例,女16例;年龄40~70岁。收集到的患者均采用调强和适形治疗方式,当次剂量180~240 cGy,剂量4 800~7 000 cGy。入组条件为:(1)经病理学或细胞学检查诊断为原发性肺癌;(2)首次接受肺部调强适形放疗;(3)接受完整放疗过程,能够从放疗计划中获取完整体积剂量数据;(4)随访时间大于3个月。

1.2 图像采集

采用CT模拟定位系统(飞利浦 Brilliance CT BigBore)扫描定位,患者以舒适的姿势自由平躺于扫描床,CT扫描条件:电压120 kV,电流300 mAs,螺旋扫描螺距为0.938,重建层厚为3 mm。扫描好的图像传至Eclipse放射治疗计划系统。在Eclipse放射治疗计划系统上靶区由高年资医师根据肺癌勾画指南勾画。肿瘤靶区(GTV)包括肺部肿瘤和CT上可见的

区域肿大淋巴结。GTV边界外放6 mm作为计划靶区(PTV)。勾画正常器官:左右肺、心脏、食管等。靶区和正常器官勾画完成后,在Eclipse放射治疗计划系统上由资深物理师完成治疗计划,再由临床医生完成审核认可后再执行。

1.3 临床数据和物理剂量参数收集

依据不良事件通用术语标准评价做放射性肺炎评级,2级及以上的为放射性肺炎并且放疗结束后CT随访3个月以上,发生放射性肺炎48例,没有发生放射性肺炎的35例。每个病人临床数据包括:性别、肿瘤位置、是否手术、临床分期。物理剂量参数41个:患侧肺的 $V_5$ 、 $V_{10}$ 、 $V_{15}$ 、 $V_{20}$ 、 $V_{25}$ 、 $V_{30}$ 、 $V_{35}$ 、 $V_{40}$ 、 $V_{45}$ 、 $V_{50}$ ,全肺的体积剂量,PTV的体积及它的最大剂量、最小剂量、平均剂量,心脏的体积、最大剂量、最小剂量、平均剂量、体积剂量,放疗计划野数、治疗技术、单次剂量等。

1.4 影像组学特征提取

使用开源的Python语言环境的pyradiomics工具包<sup>[5]</sup>,从医学成像中提取影像特征。借软件包建立放射分析的参考标准,并提供经过测试和维护的开源平台,以方便可重复地提取放射特征。pyradiomics支持2D和3D中的特征提取,可用于计算感兴趣区域每个特征的单个值或生成特征图。pyradiomics提取的各种功能的定义:一阶统计学特征、形状特征(3D)、形状特征(2D)、灰度共生矩阵(GLCM)、灰度大小区域矩阵(GLSZM)、灰度游程长度矩阵(GLRLM)、相邻灰度色调差异矩阵(NGTDM)、灰度依赖矩阵(GLDM)<sup>[6]</sup>。本研究用了107个特征如表1所示,选择的感兴趣区域为GTV。

表1 提取的候选特征  
Tab.1 Extracted candidate features

特征类型	提取的候选特征
形态学特征 (n=14)	Elongation, Flatness(3D), Least Axis Length(3D), Major Axis Length, Maximum2DDiameterColumn, Maximum2DDiameterRow, Maximum2DDiameterSlice, Maximum3DDiameter, MeshVolume, MinorAxisLength, Sphericity, SurfaceArea, SurfaceVolumeRatio, VoxelVolume
一阶统计学特征 (n=18)	10Percentile, 90Percentile, Energy, Entropy, InterquartileRange, Kurtosis,Maximum, MeanAbsoluteDeviation, Mean, Median, Minimum, Range, RobustMeanAbsoluteDeviation, RootMeanSquared, Skewness, TotalEnergy, Uniformity, Variance
纹理特征(n=75)	GLCM(n=24), GLSZM(n=16), GLRLM(n=16), NGTDM(n=5), GLDM(n=14)

1.5 鉴别模型构建

基于22种特征选择算法和8个分类器构建出176个鉴别模型,将上述临床物理剂量和影像组学特征作为模型的输入,对鉴别模型进行训练。由于数据量较小,没有分单独的测试集,在模型训练过程中,采用了五折交叉验证来评估模型的鉴别性能。

由于是否有放射性肺炎的两类数据不平衡,因此在模型的训练过程中,采用了SMOTE算法对数据进行增强。该研究中采用了22种特征选择算法和8个分类器。一个模型由1个特征选择算法和1个分类器组成,共构建176个不同的鉴别模型。22种特征选择算法分别为:CIFE、CMIM、DISR、ICAP、JMI、MIFS、

MIM、MRMR、Fisher\_score、Lap\_score、ReliefF、SPEC、Trace\_ratio、II\_I21、Is\_I21、MCFS、NDFS、RFS、UDFS、F\_score、Gini\_index、T\_score<sup>[7-12]</sup>。8个不同的分类器分别是：Logistics Regression、SVM、Naïve Bayes、KNN、Decision Tree、Bagging、Random Forest、AdaBoosting。

1.6 统计学方法

采用SPSS20.0统计分析软件,符合正态分布的计量资料使用独立t检验分析,不符合正态分布的计量资料用非参数检验的方法即Mann-Whitney U。P<0.05为差异有统计学意义。

评价上述176个鉴别模型对放射性肺炎和非放

射性肺炎的鉴别能力的指标,包括ROC下面积(Area Under Curve, AUC)。

2 结果

将临床物理剂量特征和影像组学特征作为22种特征选择算法的输入,利用特征选择算法选取最优的特征子集并输入到分类器中进行训练,实验可得选取前30个特征的鉴别效果最佳。根据模型的训练结果作AUC值热量图(图1)。每个模型的AUC值是5折交叉的平均值。图1是各种特征选择算法与分类器所对应的AUC值,颜色越深,AUC值越大,颜色越浅,AUC值越小。

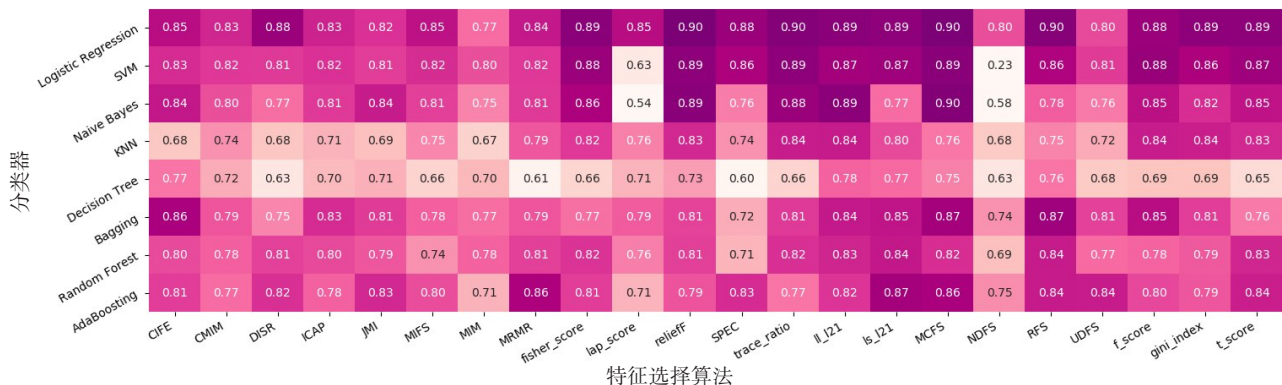


图1 临床剂量特征和影像组学特征的AUC值(选取30个特征)  
Fig.1 AUC of the combination of radiomics signatures with clinical dosimetric features (30 features were selected)

选择AUC大于0.7的所有模型的特征进行出现频率统计,图2中标蓝部分代表了在临床物理剂量和影像组学特征中前10位的优势特征,分别为Shape-based特征中的shape\_Maximum2DDiameterColumn、shape\_Maximum3DDiameter、shape\_Sphericity, First-

Order based特征中的firstorder\_90Percentile, Texture-based特征中的glcm\_Imc1,还有临床物理剂量特征中的V<sub>20</sub>、V<sub>45</sub>、V<sub>50</sub>、V<sub>50a</sub>、V<sub>25</sub>。标蓝柱子的高低代表了特征被优势模型选中的频率和次数,标蓝柱子越高代表特征越有优势。

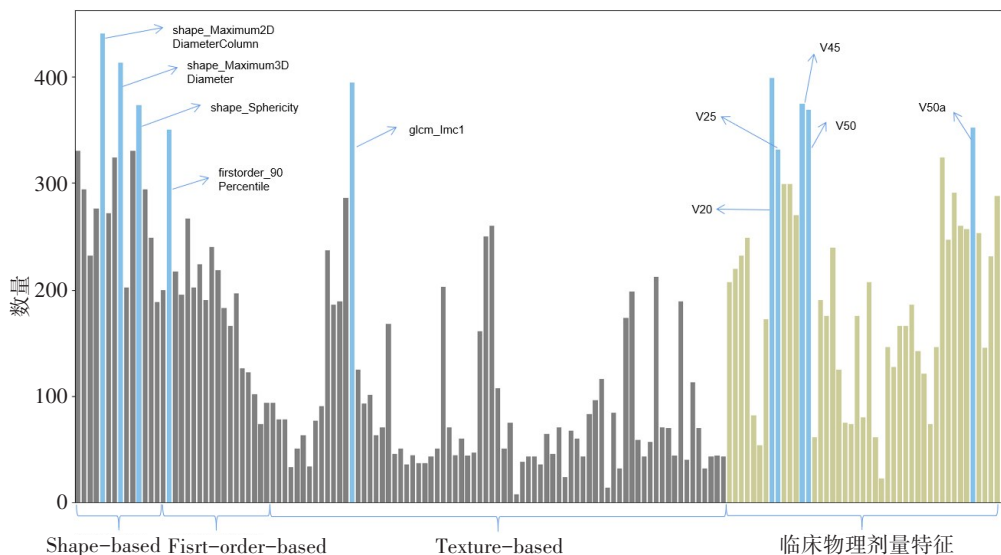


图2 特征统计条形图  
Fig.2 Features statistical bar chart

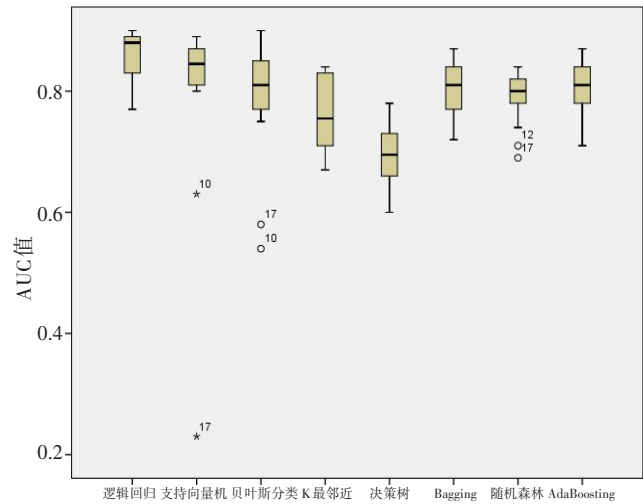


图3 8个分类器的AUC箱线图  
Fig.3 AUC boxplot of 8 classifiers

图3为8种分类器针对22种特征选择算法的每种鉴别模型的AUC值箱线图,图中箱体上边框表示上四分位数即75%,箱体下边框表示下四分位数即25%,箱体内横线表示中位数,箱体上面横线表示95%数据,下面横线表示5%数据。从图3可以看出逻辑回归、支持向量机的分类效果比较理想。逻辑回归的AUC值中位线最高,支持向量机的AUC值次之。

表2是临床物理剂量和影像组学优势特征的统计学分析,前10位的特征中只有 firstorder\_90Percentile的P值大于0.05,在放射性肺炎(RP)和非放射性肺炎(NRP)中没有显著性差异,其他9个特征P值均小于0.05,有显著性差异,其中临床物理剂量特征P值均小于10<sup>-5</sup>。

表2 10个关键特征的统计分析  
Tab.2 Statistical analysis of 10 key features

特征类别	特征排名	RP	NRP	P值	M <sup>†1</sup>	(<M >M) <sup>‡2</sup>
临床物理特征 (n=5)	V <sub>20</sub> (3 <sup>rd</sup> )/%	53.39±15.74	36.60±9.01	10 <sup>-7,a</sup>	45.00	RP(32.5% 67.5%) NRP(81.5% 18.5%)
	V <sub>45</sub> (5 <sup>th</sup> )/%	23.05(4.90, 60.90)	12.08(5.20, 24.20)	10 <sup>-6,b</sup>	17.56	RP(37.5% 62.5%) NRP(85.2% 14.8%)
	V <sub>50</sub> (7 <sup>th</sup> )/%	19.37(3.20, 56.60)	8.96(2.00, 22.00)	10 <sup>-6,b</sup>	14.16	RP(35.0% 65.0%) NRP(85.2% 14.8%)
	V <sub>50a</sub> (8 <sup>th</sup> )/%	9.26(2.00, 18.50)	5.06(0.70, 13.00)	10 <sup>-5,b</sup>	7.16	RP(30.0% 70.0%) NRP(81.5% 18.5%)
	V <sub>25</sub> (10 <sup>th</sup> )/%	46.91±14.15	28.96±6.72	10 <sup>-9,a</sup>	37.94	RP(30.0% 70.0%) NRP(92.6% 7.4%)
影像组学特征 (n=5)	shape_Maximum2D	95.43±33.24	134.82±43.79	10 <sup>-4,a</sup>	115.13	RP(65.0% 35.0%) NRP(33.3% 66.7%)
	DiameterColumn(1 <sup>st</sup> )	113.47±39.52	153.78±48.00	10 <sup>-4,a</sup>	133.62	RP(65.0% 35.0%) NRP(33.3% 66.7%)
	shape_Maximum3D	-0.21±0.04	-0.26±0.04	10 <sup>-6,a</sup>	-0.24	RP(32.5% 67.5%) NRP(70.4% 29.6%)
	Diameter(2 <sup>nd</sup> )	0.51±0.11	0.43±0.08	10 <sup>-3,a</sup>	0.47	RP(32.5% 67.5%) NRP(74.1% 25.9%)
	glcm_Imc1(4 <sup>th</sup> )	78.15(-118.00, 267.00)	100.63(70.00, 250.00)	0.107 <sup>b</sup>	89.39	RP(50.0% 50.0%) NRP(51.9% 48.1%)
	shape_Sphericity(6 <sup>th</sup> )					
	firstorder_90 Percentile(9 <sup>th</sup> )					

<sup>†1</sup>表示RP组的平均特征值与NRP组的平均特征值的均值;<sup>‡2</sup>表示RP和NRP组中特征值小于或大于M值的患者百分比;a:独立t检验;b:Mann Whitney U检验

3 讨论

放射性肺炎是肺癌放疗的主要并发症之一,影响病人放疗后的生存质量。如果能够预测放射性肺炎的发生,就可以提前避免或者干预,使病人生存质量提高。放射性肺炎的发生与诸多因素有关,如患者相关的因素年龄、性别、吸烟状况和肺部疾病等,治疗方面原因如化疗药物和化疗疗程、免疫治疗药物和疗程、手术方式,还有放射治疗方面的剂量、治疗技术、野数、体积剂

量参数等<sup>[13-16]</sup>。在临床中找到影响放射性肺炎的因素以及影响的程度很重要,可以指导临床工作,能够提前预估放射性肺炎的发生几率,避免放射性肺炎的发生或者做好预防措施。在以往的研究中大多是通过统计的方法来研究单个因素或者多个因素的影响。放射性肺炎相关因素的研究很多,大多数研究已将V<sub>20</sub>和平均肺部剂量做为最常见的相关参数,同时发现除了肺部剂量外,心脏剂量参数也有相关性<sup>[17-20]</sup>。



如果想建立一个判别模型来准确地区分两种事物,那么判别精度是最先考虑的,而判别模型达到公认的判别精度,则可能不再需要进行综合分析。如果只是探索两种事物中起重要作用的潜在因素,则应需要进行综合分析,通过采用不同分类器和不同的特征选择算法构建鉴别模型来筛选出被选中的最频繁的关键特征即最可能的关键因素。放射性肺炎的影响因素众多,如何选出更相关的因素很重要,同时也是本研究的目的。本研究采用多种鉴别模型考虑了不同分类器和特征选择算法之间存在算法差异。不同的分类器有不同的判别能力,而不同的特征选择算法选择不同的最优特征子集并输入到分类器中,所以我们采用多个不同的分类器和多个不同的特征选择方法的组合来构建模型进行更加全面评估。

临床物理剂量特征和影像组学特征用不同特征选择算法和不同分类器组合的176个鉴别模型进行预测放射性肺炎训练,模型中选次数在前面5位的是:shape\_Maximum2DDiameterColumn、shape\_Maximum3DDiameter、 $V_{20}$ 、glcm\_Imc1、 $V_{45}$ 。从图2可以看出各种特征选出次数的高低,预测效果理想的特征是shape\_Maximum2DDiameterColumn、shape\_Maximum3DDiameter、 $V_{20}$ ,前面两个跟CT的微观形态结构相关,而患侧肺 $V_{20}$ 是普遍认可的预测特征。Logistic regression分类器与ReliefF、Trace\_ratio、MCFS、RFS 4种特征选择算法得到的模型AUC值都是0.9,分类效果理想。

以往的研究大多集中在医学统计的单个因素的分析,或者是单一模型的设计分析,而本研究利用了多个模型分析,不但可以筛选出好的模型而且可以选出好的特征。从上面的研究可以看出在目前的数据和样本情况下,临床物理剂量特征和影像组学特征预测肺癌放疗放射性肺炎比较理想。本研究的局限性:首先,临床治疗特征少,没有收集到化疗药物、免疫治疗药物、疗程数等,没有量化临床治疗对放射性肺炎的影响;其次,基于CT影像组学特征进行模型构建,并未纳入相关剂量云图特征与之配合,有待在后续研究中增加患者临床信息,并且完善分类模型。

综上所述,基于CT的影像组学特征和临床物理剂量特征预测肺癌放疗放射性肺炎效果理想。

## 【参考文献】

[1] TORRE LA, BRAY F, SIEGEL R L, et al. Global cancer statistics, 2012 [J]. *Ca Cancer J Clin*, 2015, 65(2): 87-108.

- [2] MARKS L B, BENTZEN S M, DEASY J O, et al. Radiation dose-volume effects in the lung[J]. *Int J Radiat Oncol Biol Phys*, 2010, 76 (Suppl 3): S70-S76.
- [3] GIULIANI M E, LINDSAY P E, KWAN J Y, et al. Correlation of dosimetric and clinical factors with the development of esophagitis and radiation pneumonitis in patients with limited-stage small-cell lung carcinoma[J]. *Clin Lung Cancer*, 2015, 16(3): 216-220.
- [4] PARMAR C, GROSSMANN P, BUSSINK J, et al. Machine learning methods for quantitative radiomic biomarkers[J]. *Sci Rep*, 2015, 5(1): 1-11.
- [5] VAN GRIETHUYSEN J J, FEDOROV A, PARMAR C, et al. Computational radiomics system to decode the radiographic phenotype [J]. *Cancer Res*, 2017, 77(21): e104-e107.
- [6] THIBAULT G, ANGULO J, MEYER F. Advanced statistical matrices for texture characterization: application to cell classification[J]. *IEEE Trans Biomed Eng*, 2013, 61(3): 630-637.
- [7] BROWN G, POCOCCOCK A, ZHAO M J, et al. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection[J]. *J Mach Learn Res*, 2012, 13(1): 27-66.
- [8] LI J, CHENG K, WANG S, et al. Feature selection: a data perspective [J]. *ACM Comput Surv*, 2017, 50(6): 1-45.
- [9] HE X, CAI D, NIYOGI P. Laplacian score for feature selection[C]. Vancouver: MIT Press, 2005.
- [10] ROBNIK-AIKONJA M, KONONENKO I. Theoretical and empirical analysis of ReliefF and RReliefF[J]. *Mach Learn*, 2003, 53(1-2): 23-69.
- [11] LIU J, JI S, YE J. Multi-task feature learning via efficient l2, 1-norm minimization[C]. Montreal: AUAI Press, 2009.
- [12] LI Z C, TANG J H. Unsupervised feature selection via nonnegative spectral analysis and redundancy control[J]. *IEEE Trans Image Process*, 2015, 24(12): 5343-5355.
- [13] KONG F S, WANG S. Nondosimetric risk factors for radiation-induced lung toxicity[J]. *Semin Radiat Oncol*, 2015, 25(2): 100-109.
- [14] LEPRIEUR E G, FERNANDEZ D, CHATELLIER G, et al. Acute radiation pneumonitis after conformational radiotherapy for nonsmall cell lung cancer: clinical, dosimetric, and associated-treatment risk factors[J]. *J Cancer Res Ther*, 2013, 9(3): 447-451.
- [15] TANG X, LI Y, TIAN X, et al. Predicting severe acute radiation pneumonitis in patients with non-small cell lung cancer receiving postoperative radiotherapy: development and internal validation of a nomogram based on the clinical and dose-volume histogram parameters[J]. *Radiation Oncol*, 2019, 132: 197-203.
- [16] DANG J, LI G, ZANG S, et al. Comparison of risk and predictors for early radiation pneumonitis in patients with locally advanced non-small cell lung cancer treated with radiotherapy with or without surgery[J]. *Lung Cancer*, 2014, 86(3): 329-333.
- [17] TAKAHASHI S, GO T, KASAI Y, et al. Relationship between dose-volume parameters and pulmonary complications after neoadjuvant chemoradiotherapy followed by surgery for lung cancer[J]. *Strahlenther Onkol*, 2016, 192(9): 658-667.
- [18] ALBAIN K S, SWANN R S, RUSCH V W, et al. Radiotherapy plus chemotherapy with or without surgical resection for stage III non-small-cell lung cancer: a phase III randomised controlled trial [J]. *Lancet*, 2009, 374(9687): 379-386.
- [19] TUCKER S L, LIAO Z, DINH J, et al. Is there an impact of heart exposure on the incidence of radiation pneumonitis? Analysis of data from a large clinical cohort[J]. *Acta Oncol*, 2014, 53(5): 590-596.
- [20] TSUJINO ACCELERATED K, HIROTA S, KOTANI Y, et al. Radiation pneumonitis following concurrent hyperfractionated radiotherapy and chemotherapy for limited- small-cell lung cancer: dose-volume histogram analysis and comparison stage with conventional chemoradiation [J]. *Int J Radiat Oncol Biol Phys*, 2006, 64(4): 1100-1105.

(编辑:陈丽霞)