

# CT影像组学联合形态学特征模型评估非小细胞肺癌患者预后生存期的价值

周洁<sup>1</sup>, 郑燕婷<sup>1</sup>, 江舒琪<sup>1</sup>, 安杰<sup>1</sup>, 邱士军<sup>1</sup>, SUWAL Sushant<sup>2</sup>, 黄绥丹<sup>2</sup>, 陈淮<sup>2</sup>, 李翠<sup>3</sup>, 方嘉琪<sup>3</sup>

1. 广州中医药大学第一附属医院影像科, 广东 广州 510405; 2. 广州医科大学附属第二医院放射科, 广东 广州 510260; 3. 广州中医药大学第一临床医学院, 广东 广州 510405

**【摘要】目的:**探讨CT影像组学联合形态学特征模型对非小细胞肺癌患者预后生存时间的预测价值。**方法:**从癌症影像数据库(TCIA)下载并筛选出300例非小细胞肺癌患者资料(300个病灶),随机选取210例为训练集,90例为测试集。根据预后生存期将患者分为两组,预后生存期 $\leq 3$ 年为1组,预后生存期 $> 3$ 年为2组。用3D Slicer软件在CT图像病灶中逐层勾画得到感兴趣区(ROI),再从每个ROI中提取影像组学特征,采用 $t$ 检验和最小绝对收缩与选择算子(LASSO)算法进行影像组学特征筛选。运用Logistic回归建立3种预测模型,包括影像组学模型、形态学模型和联合诊断模型。采用受试者工作特征(ROC)曲线评价3种预测模型效能。**结果:**影像组学标签、纵隔淋巴结转移在训练集和测试集的差异均具有统计学意义。影像组学模型、形态学模型和联合诊断模型,在训练集中ROC曲线下面积(AUC)分别为0.784(95% CI:0.722~0.847)、0.734(95% CI:0.664~0.804)、0.748(95% CI:0.680~0.815),在测试集中分别为0.737(95% CI:0.630~0.844)、0.665(95% CI:0.554~0.777)、0.687(95% CI:0.578~0.797)。影像组学模型诊断效能最佳。**结论:**CT影像组学模型能有效地预测非小细胞肺癌患者预后生存时间。

**【关键词】**非小细胞肺癌;影像组学;形态学特征;预后;生存

**【中图分类号】**R318;R734.2

**【文献标志码】**A

**【文章编号】**1005-202X(2024)01-0018-09

## Value of CT radiomics combined with morphological features in predicting the prognosis of patients with non-small cell lung cancer

ZHOU Jie<sup>1</sup>, ZHENG Yanting<sup>1</sup>, JIANG Shuqi<sup>1</sup>, AN Jie<sup>1</sup>, QIU Shijun<sup>1</sup>, SUWAL Sushant<sup>2</sup>, HUANG Suidan<sup>2</sup>, CHEN Huai<sup>2</sup>, LI Cui<sup>3</sup>, FANG Jiaqi<sup>3</sup>

1. Department of Imaging, the First Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou 510405, China;

2. Department of Radiology, the Second Affiliated Hospital of Guangzhou Medical University, Guangzhou 510260, China; 3. The First

Clinical Medical College, Guangzhou University of Chinese Medicine, Guangzhou 510405, China

**Abstract: Objective** To explore the predictive value of CT radiomics and morphological features for the prognosis and survival in non-small cell lung cancer (NSCLC) patients. **Methods** The clinic data of 300 NSCLC patients (300 lesions) were downloaded from the Cancer Imaging Archive, with 210 randomly selected as the training set and 90 as the test set. According to the prognosis and survival, the patients were divided into two groups with survival period  $\leq 3$  and  $> 3$  years. 3D Slicer software was used to delineate the regions of interest layer by layer in CT images, and the radiomics features were extracted from each region of interest. Both  $t$ -test and least absolute shrinkage and selection operator were utilized for radiomics feature screening. Three types of prediction models, namely radiomics model, morphological model and combined model, were constructed with Logistic regression, whose performances were evaluated using the receiver operating characteristic (ROC) curve. **Results** The differences in radiomics labels and mediastinal lymph node metastasis between the training set and the test set were statistically significant. For radiomics model, morphological model and combined model, the area under the ROC curve was 0.784 (95% CI:0.722-0.847), 0.734 (95% CI:0.664-0.804) and 0.748 (95% CI:0.680-0.815) in the training set, and 0.737 (95% CI:0.630-0.844), 0.665 (95% CI:0.554-0.777) and 0.687 (95% CI:0.578-0.797) in the test set, which demonstrated that radiomics model had the best diagnostic performance. **Conclusion** The CT radiomics model can effectively predict the prognosis and survival in NSCLC patients.

**Keywords:** non-small cell lung cancer; radiomics; morphological feature; prognosis; survival

**【收稿日期】**2023-09-10

**【基金项目】**广东省自然科学基金(2022A1515011028)

**【作者简介】**周洁,博士,副主任医师,主要研究方向:胸部疾病影像诊断,E-mail: zhoujie1989@gzucm.edu.cn

**【通信作者】**陈淮,博士,教授,主要研究方向:胸部疾病影像诊断,E-mail: chenhuai1977@163.com

前言

肺癌的发病率以及死亡率在全球范围内居第一位。在肺癌患者中,非小细胞肺癌(Non-Small Cell Lung Cancer, NSCLC)占80%~85%<sup>[1]</sup>,其中大部分患者发现肿瘤时已达中晚期,预后较差。临床上对肿瘤进行精准的分期及预后评估,对个性化治疗方法选择至关重要,而这些都依赖于治疗前对肿瘤进行全面的评估。胸部CT广泛地运用于肺癌的诊断,其图像可提供肿瘤的形态学信息,而影像组学作为当前的研究热点,能在CT图像上提取大量的影像特征定量信息,进一步反映肿瘤的异质性,在肺癌疗效的评估及病理预测等方面有很好的帮助<sup>[2-3]</sup>。前期有较多学者通过肺癌影像组学模型和临床模型相结合预测NSCLC淋巴结转移<sup>[4-5]</sup>、骨转移<sup>[6]</sup>、脑转移<sup>[7]</sup>;通过影像组学方法预测孤立性肺结节的侵袭性、浸润情况<sup>[8-9]</sup>、肺癌的分期<sup>[10]</sup>,预测晚期肺癌联合放化疗治疗效果<sup>[11]</sup>,预测四期NSCLC化疗后骨骼肌反应、肌肉萎缩<sup>[12]</sup>。通过影像组学结合基因表达谱的方法成功预测肺癌患者的化疗、靶向治疗效果<sup>[13]</sup>。通过影像组学机器学习方法预测NSCLC肿瘤EGFR和KRAS基因突变<sup>[14-15]</sup>。然而,目前通过肺癌影像组学及形态特征预测患者的生存期相对较少,本研究尝试通过对肿瘤CT影像组学、形态特征进行建模分析,探讨不同预测模型在评估NSCLC患者预后生存期的应用价值。

1 材料与方法

1.1 研究对象

实验中使用的数据来自癌症影像数据库(the Cancer Imaging Archive, TCIA) NSCLC-Radiomics 项

目的lung 1数据集。TCIA是癌症研究的公共数据库,包含病例的医学影像资料及临床资料。lung 1数据集中包含422例NSCLC患者的资料,包括CT图像、截止生存状态和截止生存时间(发现病灶到统计时的时间),还包含病理结果、年龄、性别和TNM分期等临床信息。病例的CT影像数据图像的矩阵是512×512,层厚3 mm。纳入标准:(1)病理结果为NSCLC,有完整胸部CT图像以及基本临床信息;(2)肺内单发肿瘤病灶;(3)扫描范围覆盖整个瘤体。数据排除标准:(1)仅进行胸部CT增强扫描的患者;(2)肺内多发肿瘤病灶的患者。最后筛选出了300例NSCLC患者数据进行试验,根据预后生存期将患者分为两组,预后生存期≤3年为1组,预后生存期>3年为2组。

1.2 图像分割与特征提取

本试验从患者CT影像资料上进行病灶分割及图像特征提取,使用医学图像处理软件3D Slicer(版本5.0.3, <http://www.slicer.org>)。将研究病例CT图像导入该软件,先由一名具有7年放射诊断经验的医师在横断面上选择肿瘤病灶的3个层面(包括最大横断面及其上、下层面)勾划其边缘轮廓,进行三维半自动分割,逐层人工修改勾划的ROI轮廓,在这个过程中尽量避开支气管及血管等结构。再请一名具有17年放射诊断经验的医师对病灶轮廓勾划范围进行审核、修改(图1、2)。在提取肿瘤病灶的影像组学特征前,为了减少图像采集异质性影响,图像需要进行预处理,采用像素大小归一化和重采用的方法,把体素归一化为1 mm×1 mm×1 mm。本研究从每个病灶提取影像组学特征共1 037个,包括形状特征14个、一阶直方图特征162个、纹理特征861个。

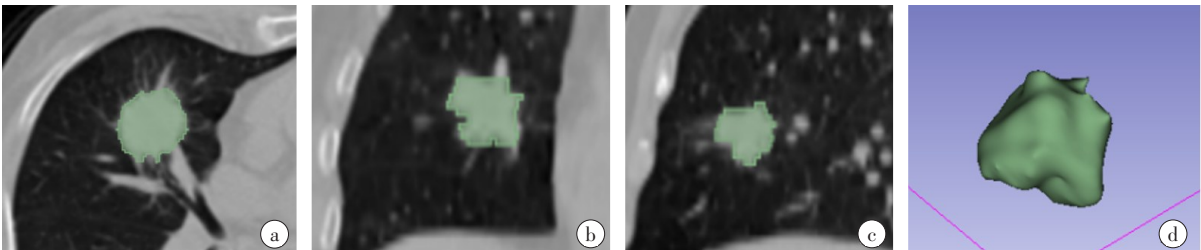


图1 运用3D Slicer软件对周围型肺癌病灶ROI勾画及三维重建

Figure 1 ROI delineation and 3D reconstruction of peripheral lung cancer lesions using 3D Slicer software

a:原始CT横断面肺窗图像,层厚3 mm;b:冠状面重建图像;c:矢状面重建图像;d:对病灶轮廓勾画完成后重建出的三维病灶图像

1.3 可重复性分析

用随机的方法从300个肿瘤病灶中抽取60个,请上述两位医师分别对病灶ROI轮廓进行勾画并提

取影像组学特征。用组内相关系数(Intraclass Correlation Coefficient, ICC)对他们提取的特征进行一致性分析。以ICC>0.75定义为一致性较好。

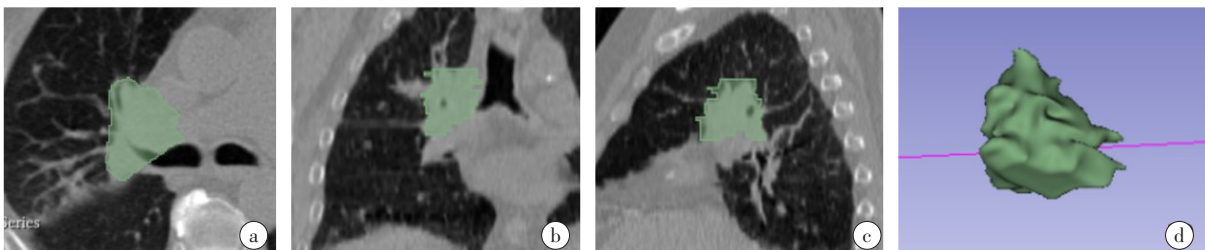


图 2 运用 3D Slicer 软件对中央型肺癌病灶 ROI 勾画及三维重建

Figure 2 ROI delineation and 3D reconstruction of central lung cancer lesions using 3D Slicer software

a:原始 CT 横断面肺窗图像,层厚 3 mm;b:冠状面重建图像;c:矢状面重建图像;d:对病灶轮廓勾画完成后重建出的三维病灶图像

1.4 CT 形态学特征的评估

所有病例的 CT 图像由两名分别具有 7 年和 17 年放射诊断经验的医师在肺窗(窗宽 1 400 HU,窗位 -500 HU)下独立观察,需要评估的内容包括:病灶最大径、深分叶征、胸膜凹陷征、毛刺征、棘状突起、支气管气相、空洞、空泡征、钙化、肺门淋巴结肿大、纵隔淋巴结肿大等。病灶最大径的测量取肿瘤横截面的最长径。深分叶征定义为肿块单个分叶的弧弦距与弦长比值大于 0.3。胸膜凹陷征定义为病灶与胸膜之间的细小条样影。毛刺征定义为病灶边缘伸出 3 条以上毛刺影。棘状突起为介于分叶与毛刺之间的一种较粗大而钝的结构。支气管气相定义为病灶内细条状,直径约 1 mm 的空气密度影,或连续层面上呈小泡状(直径<1 mm)的气体密度影。空洞定义为大于 5 mm 的圆形或类圆形气体密度影。空泡征定义为病灶内<5 mm 的点状透亮影。淋巴结肿大定义为短径≥10 mm 的淋巴结。如果两位医师评估的结果不一致,该病例图像将由一位具有 22 年放射诊

断经验的医师重新评估。

1.5 Logistic 回归模型分析

用分层抽样的方法将 300 例患者数据以 7:3 比例划分为训练集和测试集,训练集 210 例,测试集 90 例。在训练集中使用 *t* 检验和最小绝对收缩与选择算子 (Least Absolute Shrinkage and Selection Operator, LASSO) 算法,通过 10 折交叉验证,筛选出最能体现数据集特点的影像组学特征,建立影像组学特征 Logistic 回归模型。患者临床资料和 CT 形态学特征使用单因素分析(*t* 检验和卡方检验)进行筛选,建立形态学特征 Logistic 回归模型。计算影像组学标签值 Radscore(RS),并将 RS 和筛选后的形态学特征合并,建立影像组学与形态学联合的 Logistic 回归模型。运用受试者工作特征 (Receiver Operating Characteristic, ROC) 曲线的曲线下面积 (Area Under Curve, AUC) 来分别评价上述 3 种模型在训练集和测试集的效能。详细过程如图 3 所示。

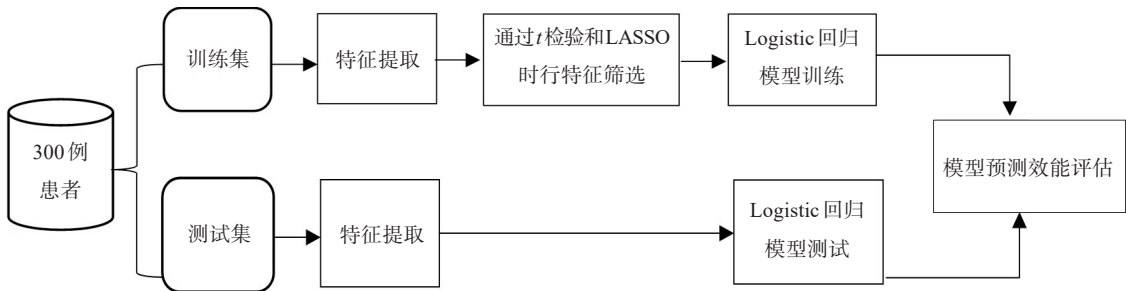


图 3 本文模型建立流程图

Figure 3 Model establishment flowchart

1.6 统计学方法

本研究数据统计使用 SPSS 27.0 和 Python 软件 (version 3.9.13) 进行分析。计数资料用 *n*(%) 的形式表示,计量资料用均值±标准差的方式表示。运用独立样本 *t* 检验对计量资料进行统计学分析;运用卡方

检验对计数资料进行统计学分析。使用 Python 软件 LASSO 回归对影像组学特征进行筛选。使用 SPSS 27.0 软件建立 Logistic 回归模型并绘制 ROC 曲线、计算 AUC 值。*P*<0.05 为差异有统计学意义。



2 结果

2.1 临床资料和形态学征象的组间比较

把收集的300个肺癌病灶,按预后生存期划分为1组198个,2组102个,基本资料如表1所示,CT影像

学征象资料如表2所示。训练集患者的病灶最大径、胸膜凹陷征、毛刺征、支气管气相、钙化、纵隔淋巴结肿大征象,1组和2组差异具有统计学意义( $P<0.05$ )(表2)。

表1 训练集和测试集基本临床信息  
Table 1 Basic clinical information in the training set and the test set

参数	训练集( $n=210$ )				测试集( $n=90$ )			
	1组( $n=139$ )	2组( $n=71$ )	$t/\chi^2$ 值	$P$ 值	1组( $n=59$ )	2组( $n=31$ )	$t/\chi^2$ 值	$P$ 值
性别[例(%)]			0.259	0.611			0.261	0.609
男	93(66.9)	45(63.4)			43(72.9)	21(67.7)		
女	46(33.1)	26(36.6)			16(27.1)	10(32.3)		
年龄( $\bar{x} \pm s$ , 岁)	69.56 $\pm$ 10.07	67.74 $\pm$ 8.80	1.260	0.209	69.51 $\pm$ 10.96	66.84 $\pm$ 8.45	1.183	0.120
病理类型[例(%)]			3.999	0.262			4.501	0.212
鳞癌	58(41.7)	21(29.6)			21(35.6)	9(29.0)		
腺癌	15(13.7)	10(14.1)			9(15.3)	2(6.5)		
大细胞癌	41(29.5)	21(29.6)			13(22.0)	13(42.0)		
未分类	25(18.0)	19(26.8)			16(27.1)	7(22.6)		

2.2 ICC分析

对前述两位具有7年、17年放射诊断经验的医师提取病灶ROI影像组学特征进行可重复性分析,两位医师对同一病灶(抽取60个)进行ROI边缘勾画后提取的各项影像组学特征进行比较,结果显示ICC>0.75,两者一致性较好。

2.3 影像组学分析

训练集包含210个病灶,笔者运用Python软件进行 $t$ 检验和LASSO回归进行特征筛选,从1037个影像组学特征中筛选出19个对预测肺癌患者预后生存时间最有价值的影像组学特征(图4~6),最佳的 $\lambda$ 值是0.0356。19个特征的名称和权重系数显示在表3中,包括1个原始形状特征,14个小波纹理特征,4个拉普拉斯纹理特征。LASSO回归函数公式为: $y_i=w^T x_i+b$ 。其中 $y_i$ 为预测结果, $x_i$ 为对应 $y_i$ 的每一个特征的值, $w$ 为每个特征的系数,常数项 $b=0.3381$ 。最后用上述数据来计算影像组学标签值Radscore(RS)与形态学特征建立联合模型。其中RS为 $y_i$ 的平均值。

训练集1组影像组学标签值为-8.61 $\pm$ 3.12,2组为-6.81 $\pm$ 2.09;测试集1组影像组学标签值为-8.77 $\pm$ 3.98,2组为-7.61 $\pm$ 2.05。训练集和测试集中1组的影像组学标签值的平均值高于2组。运用 $t$ 检验分析,训练集1组与2组相比较, $t$ 值为-4.945,差异有统计学意义( $P<0.001$ );测试集1组与2组相比较, $t$ 值为-1.824, $P=0.072$ 。

2.4 Logistic回归模型建立

使用经过 $t$ 检验和LASSO回归筛选出的19个影像组学特征,建立影像组学Logistic回归模型。预测模型在训练集的AUC为0.784(敏感度0.831,特异度0.655),在测试集的AUC为0.737(敏感度0.871,特异度0.492)。使用经过单因素分析筛选出的6个具有统计学意义的CT征象(病灶最大径、胸膜凹陷征、毛刺征、支气管气相、钙化、纵隔淋巴结肿大),建立形态学Logistic回归模型,预测模型在训练集的AUC为0.734(敏感度0.704,特异度0.640),在测试集的AUC为0.665(敏感度0.871,特异度0.542)。将6个CT征象联合影像组学标签值建立Logistic回归模型,预测模型在训练集的AUC为0.748(敏感度0.592,特异度0.820),测试集的AUC为0.687(敏感度0.677,特异度0.695)。影像组学模型优于形态学模型和联合模型。3种预测模型的结果见表4,3种预测模型的ROC曲线见图7、图8。

3 讨论

目前,肺癌是癌症相关死亡发生的主要原因,准确分期及疗效预测对于治疗方法选择至关重要<sup>[16]</sup>。大量临床研究表明,肺癌患者预后生存时间超过3年,复发几率将大大降低,此生存时间节点对临床医生设计患者治疗和复查方案有重大的指导意义<sup>[17]</sup>。因此,本研究以3年生存期作为时间节点,把数据集

表2 训练集和测试集CT影像学征象  
Table 2 CT imaging signs in the training set and the test set

参数	训练集( <i>n</i> =210)				测试集( <i>n</i> =90)			
	1组( <i>n</i> =139)	2组( <i>n</i> =71)	<i>t</i> / $\chi^2$ 值	<i>P</i> 值	1组( <i>n</i> =59)	2组( <i>n</i> =31)	<i>t</i> / $\chi^2$ 值	<i>P</i> 值
病灶最大径( $\bar{x} \pm s$ , mm)	45.73±22.62	33.51±17.86	3.965	0.000	43.73±20.17	44.87±21.12	0.251	0.402
类型[例(%)]			4.450	0.035			0.004	0.953
中央型	30(21.6)	7(9.9)			13(22.0)	7(22.6)		
周围型	109(78.4)	64(90.1)			46(78.0)	24(77.4)		
深分叶征[例(%)]			0.209	0.647			2.468	0.116
无	79(56.8)	38(53.5)			35(59.3)	13(41.9)		
有	60(43.2)	33(46.5)			24(40.7)	18(58.1)		
胸膜凹陷征[例(%)]			6.315	0.012			0.492	0.483
无	86(61.9)	31(43.7)			35(59.3)	16(51.6)		
有	53(38.1)	40(56.3)			24(40.7)	15(48.4)		
毛刺征[例(%)]			5.723	0.017			0.140	0.708
无	81(58.3)	29(40.8)			28(47.5)	16(51.6)		
有	58(41.7)	42(59.2)			31(52.5)	15(48.4)		
棘状突起[例(%)]			3.80	0.051			0.001	0.980
无	54(38.8)	18(25.4)			23(39.0)	12(38.7)		
有	85(61.2)	53(74.6)			36(61.0)	19(61.3)		
支气管气相[例(%)]			5.560	0.018			0.040	0.842
无	126(90.6)	71(1.0)			54(91.5)	28(90.3)		
有	13(9.4)	0(0.0)			5(8.5)	3(9.7)		
空洞[例(%)]			0.242	0.623			0.171	0.679
无	120(86.3)	63(88.7)			50(84.7)	28(90.3)		
有	19(13.7)	8(11.3)			9(15.3)	3(9.7)		
空泡征[例(%)]			0.320	0.572			0.569	0.450
无	126(90.6)	66(93.0)			53(89.8)	30(96.8)		
有	13(9.4)	5(7.0)			6(10.2)	1(3.2)		
钙化[例(%)]			4.977	0.026			0.958	0.328
无	119(85.6)	68(95.8)			52(88.1)	30(96.8)		
有	20(14.4)	3(4.2)			7(11.9)	1(3.2)		
肺门淋巴结肿大[例(%)]			1.532	0.216			5.062	0.024
无	58(41.7)	36(50.7)			18(30.5)	17(54.8)		
有	81(58.3)	35(49.3)			41(69.5)	14(45.2)		
纵隔淋巴结肿大[例(%)]			4.440	0.035			5.292	0.021
无	77(55.4)	50(70.4)			31(52.5)	24(77.4)		
有	62(44.6)	21(29.6)			28(47.5)	7(22.6)		

中的患者根据预后生存期分为两组,并利用CT扫描图像观察肿瘤的影像征象,通过影像组学方法定量分析肿瘤的异质性,尝试构建CT影像组学和形态学特征模型,以获得比肉眼阅片更多的辅助信息,预测NSCLC患者预后生存期,并对比模型的预测效能。

肺癌的CT形态学征象在其与其它肺内病变的鉴别诊断中可提供重要依据,且与肿瘤的病理基础关系密切。本研究利用患者的临床资料及CT形态学特征对NSCLC患者预后生存期进行评估,结果显示,在训练集1组与2组比较,肿瘤大小、胸膜凹陷

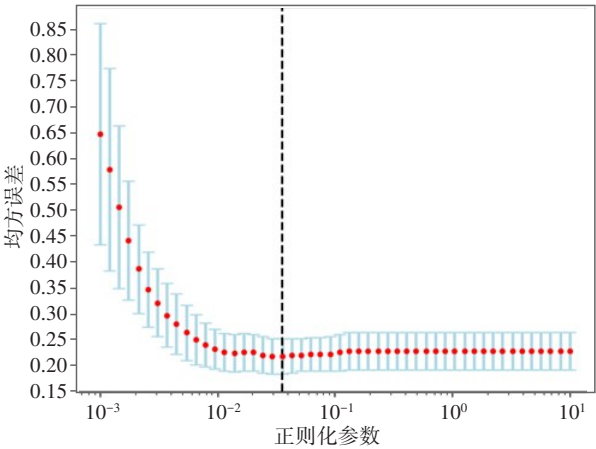


图4 LASSO 模型中正则化参数( $\lambda$ )选值图

Figure 4 Selecting the optimal regularization parameters ( $\lambda$ ) in LASSO model

LASSO 模型通过交叉验证,选取交叉验证误差最小的 $\lambda$ 值。  
虚线代表最佳 $\lambda$ 取值对应的 $\lambda$ 对数值

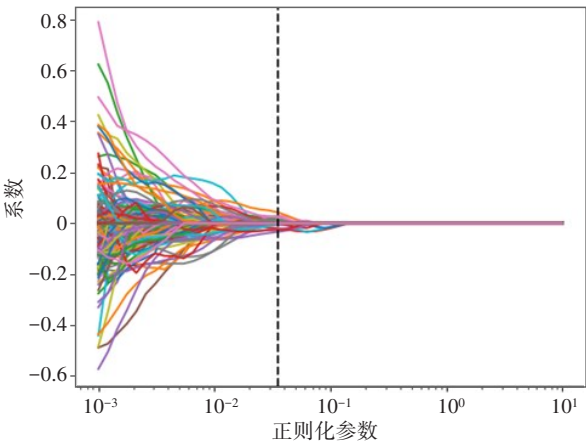


图5 使用10折交叉验证方法筛选特征的特征系数收敛图

Figure 5 Convergence of feature coefficients for feature screening using ten-fold cross-validation

每条曲线代表自变量(共1037个特征)系数随 $\lambda$ 参数的变化

征、毛刺征、支气管气相、钙化、纵隔淋巴结肿大差异有统计学意义;而测试集仅肺门、纵隔淋巴结肿大差异有统计学意义。在训练集与测试集患者中,纵隔淋巴结肿大差异均有统计学意义,在生存时期 $\leq 3$ 年的患者中,纵隔淋巴结肿大的比例较高。本研究根据训练集的结果通过 Logistic 回归建立 NSCLC 患者预后生存期的形态学模型,在训练集的 AUC 为 0.734,测试集的 AUC 为 0.665,预测能力一般。多个 CT 形态特征显示结果欠佳,说明传统影像表现的评估存在主观性大的特点,结果可靠性受影响明显。纵隔淋巴结肿大的判断标准为短径 $\geq 1$  cm,尽管此标准的敏感度和特异度不高<sup>[18-19]</sup>,但纵隔淋巴结肿大属

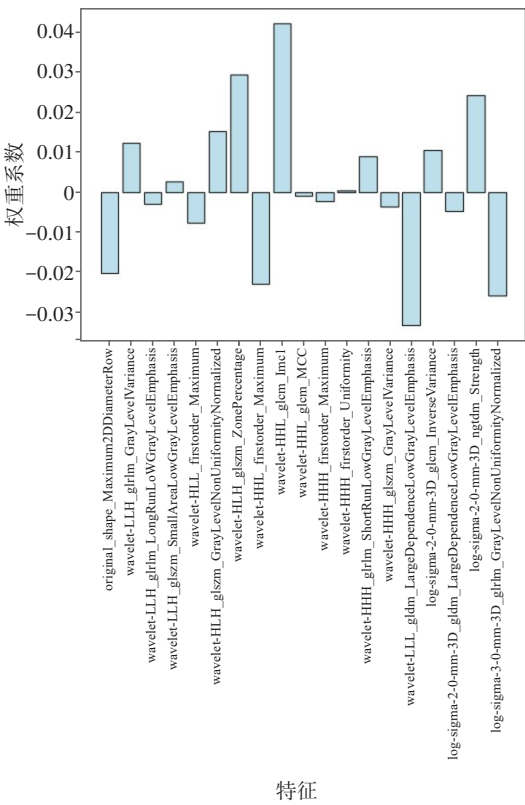


图6  $t$ 检验和LASSO筛选保留的19个特征参数

Figure 6 Nineteen feature parameters retained by  $t$ -test and LASSO screening

于测量性定量数据,稳定性好。同时纵隔淋巴结转移情况对患者的预后有至关重要的作用,无淋巴结转移的患者5年生存率约54.0%,出现纵隔淋巴结转移生存率降低至26.5%<sup>[20]</sup>。因此,本研究结果显示纵隔淋巴结肿大这项指标相较肿瘤的其它形态学征象,对于患者预后生存期表现出较好的提示作用。病灶最大径在预测肺腺癌浸润程度、肿瘤细胞的增殖能力上有一定价值,3.0 cm是肿瘤生物学行为发生转变的关键时期,但本研究病灶较大,在各组中平均值均大于3.0 cm。训练集形态特征结果显示中央型肺癌与周围型肺癌在预后生存时间上差异有统计学意义,提示肺癌的分型对研究的结果产生一定影响。本研究肺癌患者包括4种病理分类,病理是肿瘤形态学表现的基础,对病灶的CT形态学特征产生一定影响。

影像组学是影像学近年来快速发展的研究方向<sup>[21]</sup>,通过高通量方法对CT图像特征进行提取,从而对病变进行更深层次的研究<sup>[22-23]</sup>。前期的研究显示,放射组学特征与肿瘤的基因表型、蛋白表达等存在着一定的相关性,对临床的结果有一定的预测性<sup>[24-25]</sup>。如用CT为基础的放射组学方法对肺癌进行诊断及管理<sup>[26]</sup>,肺纹理特征鉴别肺肿瘤的良恶性,对

表 3 影像组学特征筛选结果

Table 3 Radiomics feature screening results

影像组学特征	权重系数
original_shape_Maximum2DDiameterRow	-0.020 461
wavelet-LLH_glrlm_GrayLevelVariance	0.012 205
wavelet-LLH_glrlm_LongRunLowGrayLevelEmphasis	-0.002 982
wavelet-LHL_glszm_SmallAreaLowGrayLevelEmphasis	0.002 563
wavelet-HLL_firstorder_Maximum	-0.007 863
wavelet-HLH_glszm_GrayLevelNonUniformityNormalized	0.015 134
wavelet-HLH_glszm_ZonePercentage	0.029 432
wavelet-HHL_firstorder_Maximum	-0.023 130
wavelet-HHL_glcm_Imc1	0.042 229
wavelet-HHL_glcm_MCC	-0.000 914
wavelet-HHH_firstorder_Maximum	-0.002 408
wavelet-HHH_firstorder_Uniformity	0.000 391
wavelet-HHH_glrlm_ShortRunLowGrayLevelEmphasis	0.008 948
wavelet-HHH_glszm_GrayLevelVariance	-0.003 664
wavelet-LLL_gldm_LargeDependenceLowGrayLevelEmphasis	-0.033 436
log-sigma-2-0-mm-3D_glcm_InverseVariance	0.010 456
log-sigma-2-0-mm-3D_gldm_LargeDependenceLowGrayLevelEmphasis	-0.004 754
log-sigma-2-0-mm-3D_ngtdm_Strength	0.024 111
log-sigma-3-0-mm-3D_glrlm_GrayLevelNonUniformityNormalized	-0.025 966

表 4 3种预测模型在训练集与测试集的 ROC 曲线结果对照

Table 4 Comparison of ROC curve results among 3 prediction models in the training set and the test set

组别	AUC	95%CI	敏感度	特异性	阳性预测值	阴性预测值
训练集						
影像组学模型	0.784	0.722~0.847	0.831	0.655	0.627	0.755
形态学模型	0.734	0.664~0.804	0.704	0.640	0.660	0.755
联合模型	0.748	0.680~0.815	0.592	0.820	0.571	0.733
测试集						
影像组学模型	0.737	0.630~0.844	0.871	0.492	0.625	0.758
形态学模型	0.665	0.554~0.777	0.871	0.542	0.462	0.675
联合模型	0.687	0.578~0.797	0.677	0.695	0.353	0.658

肺癌进行疗效评估及病理预测<sup>[27]</sup>,分析肺癌的放疗反应及放射性肺损害<sup>[28]</sup>,通过放射组学预测肺腺癌的远处转移等<sup>[29]</sup>。本研究通过 3D Slicer 软件提取影像组学特征 1 037 个,经过 *t* 检验和 LASSO 等降维处理后,筛选出 19 个具有较高诊断效能的影像组学特征用于建模预测 NSCLC 患者预后生存时间。这些特征包括:1 个原始形状特征,14 个小波纹理特征,4 个拉普拉斯纹理特征。上述特征中系数最大的是纹理

特征 wavelet-HHL\_glcm\_Imc1,属于灰度共生矩阵 (GLCM),与 Wang 等<sup>[30]</sup>,Hawkins 等<sup>[31]</sup>研究结果大致一致。GLCM 是一种通过研究灰度的空间相关特性来描述纹理的常用方法,反映在给定的方向和距离上体素值出现的概率<sup>[32]</sup>。Wang 等<sup>[30]</sup>的研究发现 GLCM 逆差距及角二阶矩在所有筛选出的特征中权重系数最高,提出这些影像组学纹理特征可以体现肿瘤表面的粗糙程度,而肿瘤表面越粗糙,其恶性程



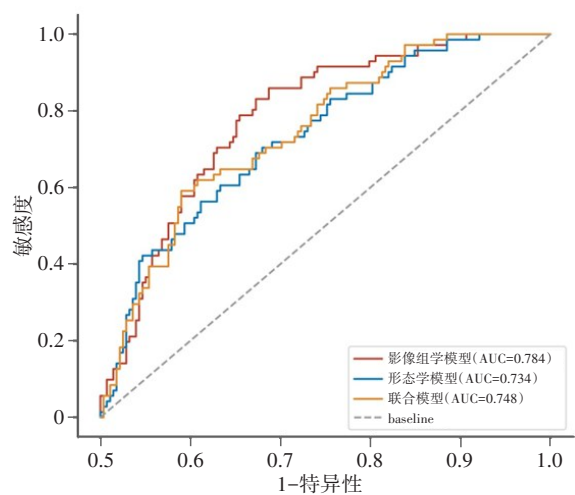


图7 ROC曲线评估训练集的影像组学模型、形态学模型和联合模型的预测效能

Figure 7 ROC curve evaluates the predictive performance of radiomics model, morphological model and combined model in the training set

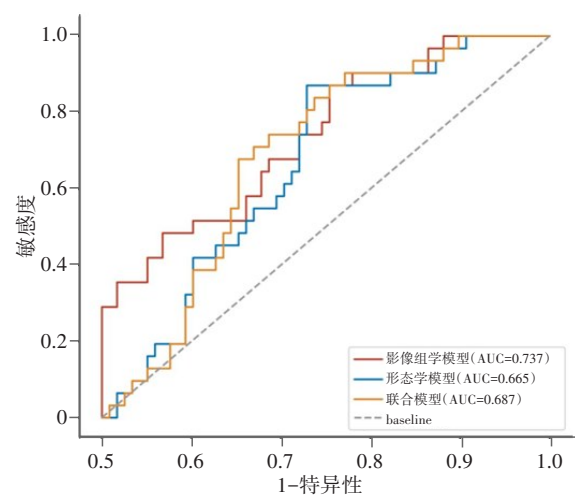


图8 ROC曲线评估测试集的影像组学模型、形态学模型和联合模型的预测效能

Figure 8 ROC curve evaluates the predictive performance of radiomics model, morphological model and combined model in the test set

度往往越高,患者预后越差。本研究权重系数最高的影像组学特征 wavelet-HHL\_glc<sub>m</sub>\_Imc1,表示信息相关度,它把体素分布的丰富程度(熵)纳入系数里计算的相关度,可量化纹理的复杂性。NSCLC患者预后生存时间≤3年组患者 wavelet-HHL\_glc<sub>m</sub>\_Imc1的平均值绝对值大于预后生存时间>3年组,分析其原因:可能是由于预后差的肺癌瘤体内部异质性更明显,CT图像体素的分布更丰富,纹理的复杂性越大。本研究通过 Logistic 回归建立 NSCLC 预后生存期影像组学模型,此模型在训练集的 AUC 为 0.784,测试集的 AUC 为 0.737,具有较高的预测能力。

本研究根据这 19 个影像组学特征计算出影像组

学标签值,在训练集 1 组和 2 组患者中差异有统计学意义,在测试集 1 组与 2 组中差异无统计学意义( $P=0.072$ ),可能是由于测试集样本量较少所致。随后通过多因素 Logistic 回归分析,将 NSCLC 患者病灶的形态学特征(病灶最大径、胸膜凹陷征、毛刺征、支气管气相、钙化、纵隔淋巴结肿大)和影像组学标签值相结合构建联合诊断模型,结果显示联合诊断模型在训练集的 AUC 为 0.748,测试集 AUC 为 0.687,较影像组学模型低,但高于形态学模型。提示 NSCLC 患者预后生存时间的长短,相较于肉眼可见的形态学特征,影像组学特征更能体现瘤体内部的异质性,对预测患者的生存有更好的作用。

本研究不足之处:(1)本研究数据来自 TCIA 数据库,病例胸部 CT 扫描层厚较厚,为 3 mm,虽然采用了像素大小归一化以及重采样等预处理,但也可能对结果产生影响。(2)数据集仅包括肿瘤患者的 CT 图像、截止生存时间以及一些基本信息,缺少血清肿瘤标志物及肿瘤病理学分化程度等其它预后相关信息。因此,本研究的结论需要后续进行多中心、多参数的研究进一步验证。

综上所述,本研究通过筛选影像组学特征,建立了一个评估 NSCLC 患者预后生存期的 Logistic 回归模型,并与形态学特征模型对比,具有较高的预测价值,有望在辅助临床医生对 NSCLC 患者预后生存评估及治疗策略选择上提供帮助。

【参考文献】

[1] Alberg AJ, Brock MV, Ford JG, et al. Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines[J]. Chest, 2013, 143(5): e1S-e29S.

[2] Coroller TP, Agrawal V, Narayan V, et al. Radiomic phenotype features predict pathological response in non-small cell lung cancer[J]. Radiother Oncol, 2016, 119(3): 480-486.

[3] Huynh E, Coroller 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.

[4] Ma XL, Xia LM, Chen J, et al. Development and validation of a deep learning signature for predicting lymph node metastasis in lung adenocarcinoma: comparison with radiomics signature and clinical-semantic model[J]. Eur Radiol, 2023, 33(3): 1949-1962.

[5] Hattori A, Matsunaga T, Takamochi K, et al. Prognostic impact of a ground glass opacity component in the clinical T classification of non-small cell lung cancer[J]. J Thorac Cardiovasc Surg, 2021, 161(4): 1469-1480.

[6] Chen L, Yu LJ, Li XY, et al. Value of CT radiomics and clinical features in predicting bone metastases in patients with NSCLC[J]. Contrast Media Mol Imaging, 2023. Doi: 10.1155/2022/7642511.

[7] Jiang YL, Wang YX, Fu S, et al. A CT-based radiomics model to predict subsequent brain metastasis in patients with ALK-rearranged non-small cell lung cancer undergoing crizotinib treatment[J]. Thorac Cancer, 2022, 13(11): 1558-1569.

[8] Huang LY, Lin WH, Xie DP, et al. Development and validation of a preoperative CT-based radiomic nomogram to predict pathology invasiveness in patients with a solitary pulmonary nodule: a machine learning approach, multicenter, diagnostic study[J]. Eur Radiol, 2022,



- 32(3): 1983-1996.
- [9] Xiong ZQ, Jiang YN, Che SY, et al. Use of CT radiomics to differentiate minimally invasive adenocarcinomas and invasive adenocarcinomas presenting as pure ground-glass nodules larger than 10 mm[J]. *Eur J Radiol*, 2021, 141: 109772.
  - [10] Zheng XS, He B, Hu YH, et al. Diagnostic accuracy of deep learning and radiomics in lung cancer staging: a systematic review and meta-analysis[J]. *Front Public Health*, 2023, 10: 938113.
  - [11] Chen NB, Xiong M, Zhou R, et al. CT radiomics-based long-term survival prediction for locally advanced non-small cell lung cancer patients treated with concurrent chemoradiotherapy using features from tumor and tumor organismal environment[J]. *Radiat Oncol*, 2022, 17(1): 184.
  - [12] de Jong EE, Sanders KJ, Deist TM, et al. Can radiomics help to predict skeletal muscle response to chemotherapy in stage IV non-small cell lung cancer?[J]. *Eur J Cancer*, 2019, 120: 107-113.
  - [13] Kirienko M, Sollini M, Corbetta M, et al. Radiomics and gene expression profile to characterise the disease and predict outcome in patients with lung cancer[J]. *Eur J Nucl Med Mol Imaging*, 2021, 48(11): 3643-3655.
  - [14] Le NQ, Kha QH, Nguyen VH, et al. Machine learning-based radiomics signatures for EGFR and KRAS mutations prediction in non-small-cell lung cancer[J]. *Int J Mol Sci*, 2021, 22(17): 9254.
  - [15] Zhu HY, Song YQ, Huang Z, et al. Accurate prediction of epidermal growth factor receptor mutation status in early-stage lung adenocarcinoma, using radiomics and clinical features[J]. *Asia Pac J Clin Oncol*, 2022, 18(6): 586-594.
  - [16] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017[J]. *CA Cancer J Clin*, 2017, 67(1): 7-30.
  - [17] Kurland BF, Gerstner ER, Mountz JM, et al. Promise and pitfalls of quantitative imaging in oncology clinical trials[J]. *Magn Reson Imaging*, 2012, 30(9): 1301-1312.
  - [18] Silvestri GA, Gould MK, Margolis ML, et al. Noninvasive staging of non-small cell lung cancer: ACCP evidenced-based clinical practice guidelines(2nd edition)[J]. *Chest*, 2007, 132(3 Suppl): 178s-201s.
  - [19] Chopra A, Modi A, Chaudhry H, et al. Assessment of mediastinal lymph node size in pneumococcal pneumonia with bacteremia[J]. *Lung*, 2018, 196(1): 43-48.
  - [20] Pak K, Park S, Cheon GJ, et al. Update on nodal staging in non-small cell lung cancer with integrated positron emission tomography/computed tomography: a meta-analysis[J]. *Ann Nucl Med*, 2015, 29(5): 409-419.
  - [21] Gillies RJ, Kinahan PE, Hricak H, et al. Radiomics: images are more than pictures, they are data[J]. *Radiology*, 2016, 278(2): 563-577.
  - [22] 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.
  - [23] Kumar V, Gu YH, Basu S, et al. Radiomics: the process and the challenges[J]. *Magn Reson Imaging*, 2012, 30(9): 1234-1248.
  - [24] Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach[J]. *Nat Commun*, 2014, 5: 4006.
  - [25] Parmar C, Grossmann P, Rietveld D, et al. Radiomic machine learning classifiers for prognostic biomarkers of head & neck cancer. *Front Oncol*, 2015, 5: 272.
  - [26] Kim H, Park CM, Goo JM, et al. Quantitative computed tomography imaging biomarkers in the diagnosis and management of lung cancer[J]. *Invest Radiol*, 2015, 50(9): 571-583.
  - [27] Coroller TP, Agrawal V, Narayan V, et al. Radiomic phenotype features predict pathological response in non-small cell lung cancer[J]. *Radiother Oncol*, 2016, 119(3): 480-486.
  - [28] Defraene G, van Elmpt W, Crijns W, et al. CT characteristics allow identification of patient-specific susceptibility for radiation-induced lung damage[J]. *Radiother Oncol*, 2015, 117(1): 29-35.
  - [29] Coroller TP, Grossmann P, Hou Y, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma[J]. *Radiother Oncol*, 2015, 114(3): 345-350.
  - [30] Wang X, Duan H, Li X, et al. A prognostic analysis method for non-small cell lung cancer based on the computed tomography radiomics[J]. *Phys Med Biol*, 2020, 65(4): 045006.
  - [31] Hawkins SH, Korecki JN, Balagurunathan Y, et al. Predicting outcomes of nonsmall cell lung cancer using CT image features[J]. *IEEE Access*, 2014, 2: 1418-1426.
  - [32] Szczypinski PM, Strzelecki M, Materka A, et al. MaZda-A software package for image texture analysis[J]. *Comput Meth Prog Biomed*, 2009, 94(1): 66-76.

(编辑:薛泽玲)