

## 非小细胞肺癌淋巴结转移预测模型研究

卢孔尧<sup>1</sup>, 黄钢<sup>2</sup>, 左艳<sup>1</sup>

1. 上海理工大学医疗器械与食品学院, 上海 200093; 2. 上海健康医学院附属嘉定中心医院上海市分子影像学重点实验室, 上海 201318

**【摘要】**在非小细胞肺癌的治疗过程中,淋巴结转移状态是决定治疗方案的重要因素。为了辅助临床医生制定更精确的治疗方案,开发并验证了一种基于CT影像组学非小细胞肺癌淋巴结转移的预测模型。从TCIA数据库的NSCLC-Radiogenomics公共数据集中选取了134例符合试验要求的患者数据,然后从每例患者的CT影像数据中提取了1 648个特征,并用特征优化方法进行特征降维和选择,然后用朴素贝叶斯、线性判别分析、支持向量机和高斯过程5种机器学习方法建立预测模型,最后使用上海市胸科医院收集的44例患者数据进行外部验证。其中,最优淋巴结转移预测模型在训练集和测试集上准确率分别为0.802和0.795,AUC值分别为0.852和0.810。试验结果表明,所提出的预测模型分类性能良好,可以辅助医生更准确地评估淋巴结转移状态,从而制定出更精准的个性化治疗方案。

**【关键词】**非小细胞肺癌;影像组学;淋巴结转移;机器学习;预测模型

**【中图分类号】**R318;R734.2

**【文献标志码】**A

**【文章编号】**1005-202X(2022)02-0182-06

### Prediction model for lymph node metastasis in non-small cell lung cancer

LU Kongyao<sup>1</sup>, HUANG Gang<sup>2</sup>, ZUO Yan<sup>1</sup>

1. School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China;  
2. Shanghai Key Laboratory of Molecular Imaging, Jiading District Central Hospital Affiliated to Shanghai University of Medicine and Health Sciences, Shanghai 201318, China

**Abstract:** In the treatment of non-small cell lung cancer, the state of lymph node metastasis is an important factor when deciding the therapeutic schedule. In order to assist clinicians in formulating a more accurate therapeutic schedule, a CT radiomics-based prediction model for lymph node metastasis in non-small cell lung cancer is developed and validated. The patient data of 134 cases meeting the requirement of experiment are selected from the NSCLC-Radiogenomics public data set in TCIA database, and then 1 648 features are extracted from the CT image data of each patient. After feature reduction and selection by feature optimization method, prediction models are established by different machine learning methods, namely naïve Bayes, linear discriminant analysis, support vector machine and Gaussian process. Finally, patient data collected from 44 cases in Shanghai Chest Hospital are used for external validation. Among the established models, the optimal model for predicting lymph node metastasis has an accuracy of 0.802 and 0.795 in training set and test set, respectively, and the AUC is 0.852 and 0.810, respectively. The experimental results show that the proposed prediction model has good classification performances, and can assist doctors in assessing the state of lymph node metastasis more accurately, so as to develop a more precise and personalized therapeutic schedule.

**Keywords:** non-small cell lung cancer; radiomics; lymph node metastasis; machine learning; prediction model

**【收稿日期】**2021-11-05

**【基金项目】**国家自然科学基金重点项目(81830052);上海市分子影像学重点实验室建设项目(18DZ2260400)

**【作者简介】**卢孔尧,硕士,主要研究方向:医学图像处理,E-mail: 2734174992@qq.com

**【通信作者】**黄钢,博士,教授,主要研究方向:核医学分子影像,E-mail: huanggang@sumhs.edu.cn

### 前言

肺癌依然是全世界癌症相关死亡的主要原因<sup>[1]</sup>。肿瘤的早期发现和分期对肺癌的治疗非常重要<sup>[2]</sup>。而非小细胞肺癌(Non-Small Cell Lung Cancer, NSCLC)患者约占所有肺癌患者的80%~85%,其治疗方式主要根据TNM(Tumor, Lymph Node, Metastasis)期系统确定<sup>[3]</sup>。国家综合癌症网络指南(2018)指出,早期NSCLC患者应首选手术,晚期NSCLC患者应首选放疗和化

疗<sup>[4]</sup>。而淋巴结转移是肺癌最常见的转移途径。决定肺癌分期、治疗方案和预后的关键因素之一是淋巴结转移(Lymph Node Metastasis, LNM)。术前对肿瘤转移的了解可以为选择辅助治疗或手术切除方案提供有价值的信息,从而帮助临床医生做出正确的决定<sup>[5]</sup>。为了保护肺功能,对于没有淋巴结转移的患者,应行有限切除,如楔形切除或节段切除<sup>[6]</sup>。已有研究表明,肺癌患者的5年生存率与淋巴结转移相关,无淋巴结转移患者的5年生存率约为56%,而有淋巴结转移仅为38%<sup>[7]</sup>。因此,术前确定NSCLC患者的淋巴结转移情况非常重要。目前,评价NSCLC淋巴结是否有转移的方法有很多,这些方法包括计算机断层扫描(Computed Tomography, CT)、正电子发射断层扫描(Positron Emission Tomography-CT, PET-CT)、超声引导活检、胸腔镜等<sup>[8-9]</sup>。PET-CT检查可以较好地评估淋巴结转移状态,对肺癌患者的淋巴结分期具有较高的特异性。然而,PET-CT检查方式存在的误诊<sup>[10]</sup>和假阴性率<sup>[11]</sup>仍然是一个值得关注的问题。此外,发展中国家的许多患者负担不起PET-CT扫描的较高费用<sup>[12]</sup>。而胸腔镜检查 and 活检等病理检查方式也能对NSCLC的淋巴结状态进行评估,但是其有创性的操作会给无淋巴结转移的患者带来负面影响。因此,找到非侵入性的方法来预测患者淋巴结转移情况对后续的临床治疗尤为重要<sup>[13-15]</sup>。

近年来,影像组学作为一种新兴的方法,受到了越来越多学者和专家的关注。影像组学在癌症患者的个体化治疗中显得越来越重要。有研究表明,影像组学方法可以从数字医学图像中提取定量特征,使在图像上挖掘出的高维数据能够应用于临床决策支持,从而提高诊断、预后和预测的准确性<sup>[16-17]</sup>。本研究的目标是开发和验证一个基于影像组学的模型,用于术前预测NSCLC患者的淋巴结转移情况。

1 材料和方法

1.1 试验材料

本研究的病人数据一部分来自癌症影像档案(The Cancer Imaging Archive, TCIA)中的NSCLC Radiogenomics公共数据集<sup>[18]</sup>,使用数据库中的134例NSCLC患者的临床和CT影像数据。另一部分数据来自上海市胸科医院经病理证实的44例NSCLC患者的临床和CT影像数据。患者的影像数据包括178组CT图像序列,每幅图像的尺寸大小为512像素×512像素,所选病例的纳入标准为:(1)年龄均大于18周岁,肺已发育完全,可以避免其他因素干扰;(2)病理诊断为NSCLC且已确定其LNM情况;(3)有完整的CT图像和个人的基本信息。排除标准为:

(1)术前接受过放疗或化疗等治疗;(2)没有明确的病理分期。为了保证试验本身的客观和可靠性,并且能更有价值地预测淋巴结转移状况,在数据集的分配上,以公共数据集作为训练集,总共134例样本,其中109例无淋巴结转移,25例有淋巴结转移。从上海胸科医院采集的数据作为测试集,总共44例样本,其中17例无淋巴结转移,27例有淋巴结转移。两部分患者的临床信息如表1所示。

表1 两中心纳入病例的临床信息(n=178)  
Table 1 Clinical information of the enrolled cases from two centers (n=178)

类别	汇总	训练集		独立测试集	
		N0	N1~N3	N0	N1~N3
数量	178	109	25	17	27
性别					
男	124	82	18	8	16
女	54	27	7	9	11
年龄					
中位数	68	69	70	63	62
最小值	37	46	43	37	45
最大值	87	87	86	74	77
吸烟史					
吸烟	135	96	19	7	13
不吸烟	43	13	6	10	14
EGFR突变状况					
突变	47	16	4	14	13
不突变	105	70	19	3	13
未知	26	23	2	0	1
病理类型					
腺癌	143	84	21	16	22
鳞癌	30	23	3	0	4
其它	5	2	1	1	1

N0:未发生淋巴结转移的NSCLC患者;N1~N3:发生淋巴结转移的NSCLC患者

1.2 方法

首先,由经验丰富的放射科医生通过开源软件ITK-SNAP<sup>[19]</sup>(3.6.0版本, www.itksnap.org)手工勾画感兴趣区域(Region of Interest, ROI)。两个病例样本最大直径处勾画的肿瘤标记CT切片图和最终经三维重建得到的三维肿瘤图见图1。然后,再从中提取形态学特征、一阶统计学特征和纹理特征。接着采用综合采样人工合成数据算法(Synthetic Minority Over-sampling Technique, SMOTE)对训练集数据做平衡化处理,再用主成分分析(Principal Component

Analysis, PCA)方法进行特征降维,而后用Relief、方差分析(Analysis of Variance, ANOVA)和递归特征消除(Recursive Feature Elimination, RFE)算法进行特征选择,最后将预处理后的数据对5种分类器进行训练,并经过5折交叉验证,建立NSCLC淋巴结转移预测模型,然后进行外部验证。工作流程如图2所示。

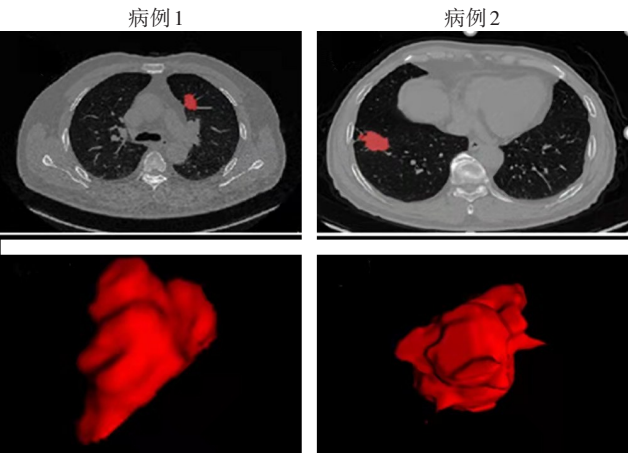


图1 两个病例肿瘤切片标记图(上)和三维重建图(下)  
Figure 1 Tumor slice labeled images (up) and three-dimensional reconstruction images (down) of 2 cases

**1.2.1 特征提取** 特征提取是影像组学中非常关键的一步,通过此步骤可以从医学影像中获取有用的信息和数据,以便于后续对肿瘤异质性的定量分析和描述。为了使CT影像组学特征更准确地分析并预测NSCLC淋巴结转移状况,应该多角度、多种类地提取相关的定量指标特征。本研究使用python中的开源软件包Pyradiomics<sup>[20]</sup> (<https://github.com/Radiomics/pyradiomics>)从分割的肿瘤区域中自动提取出1 648个影像组学特征,其中包括形态学、纹理、一阶统计学特征,具体的特征类型及名称如表2所示。

**1.2.2 数据平衡** 在试验中,由于训练集的样本中有109例无淋巴结转移,25例淋巴结转移,两者在数量上的比例不均衡。而这种样本类别之间的比例失调,往往会导致预测得出的结果出现很大的偏倚,即预测分类结果会更偏向于多数类样本,致使分类结果不准确。因此,为了解决数据不平衡问题,采用SMOTE算法对训练集数据做平衡化处理。SMOTE算法是2002年Chawla等<sup>[21]</sup>提出的一种技术,这种技术可以用过采样方法合成少数类样本的数量,以达到类别数量之间趋于均衡的效果。算法的具体流程如下所示。

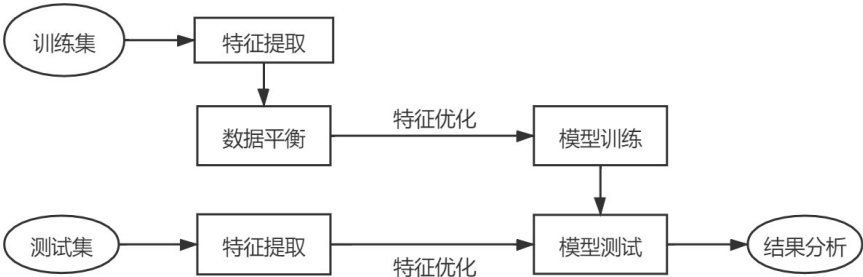


图2 本文工作流程  
Figure 2 Workflow of the proposed method

(1)在特征数据空间中,取每一个少数类的样本 $m$ ,找到该样本的 $k$ 个最近的相邻样本,一般 $k$ 值取5。(2)根据自身试验所需样本的过采样量,确定过采样倍率,然后从 $k$ 个最近的相邻样本中随机选择 $n$ 个样本( $n < k$ )。(3)先算出某一少数样本与其最近邻样本的差值,然后将该差值乘以0和1之间的一个随机数,最后将其加入原来的少数样本中,从而合成新的样本。(4)本文采用SMOTE算法,使训练集数据的无淋巴结转移样本扩充到109例。由于使用该算法得到的样本并不是真实样本数据,所以该部分样本只参与分类器模型的训练过程,而不参与模型的测试过程,测试过程选用的是44例外部数据集作为测试集进行模型验证。

**1.2.3 特征优化** 通过影像组学的方法可以从肿瘤区域中提取出大量的定量特征,但是由于提取的特征

中包含一些噪声和冗余的特征,会使后续的计算变得更加复杂,甚至会导致预测模型出现过拟合现象。因此,特征优化是一种非常有必要的方法,借助此方法可以选到一些对提高模型预测性能和数据处理速度真正重要的特征子集。试验中首先通过PCA方法压缩和简化提取到的高维特征数据,去除无关和冗余信息对预测模型的干扰。然后再使用Relief、ANOVA和RFE算法进行特征选择,最终达到特征优化的目的。

本文采用PCA算法进行特征降维,PCA可以将多个变量转化为重要的几个综合主成分,而忽略掉其中不重要的成分,这些主成分是通过原始数据的变量特征经过线性变换转化而来的组合,组合之间是互不相关的,但是能反映原始数据的大量信息,从而实现高维特征向量向低维特征向量的转换。其具

表2 所提取的特征  
Table 2 Extracted features

图像类型	特征类型	子类型	特征统计量
原图	形态特征	2D或3D	伸长度、平面度、最小轴长、主轴长、二维最长直径(行)、二维最长直径(列)、二维最长直径(切片)等
原图、小波滤波图、局部二值模式图(3D)	一阶统计特征		能量、总能量、熵、第10%的特征值、最大特征值、强度值范围、平均绝对偏差、鲁棒平均绝对偏差、均方根、偏度、峰度、方差、均匀度等
		纹理特征	和熵、差熵、联合熵、最大概率、联合能量、自相关、逆方差、差均值、集群趋势、集群突出、对比度、相关性、逆差矩、逆差等
		灰度共生矩阵	强调短游程、强调长游程、灰度不均匀性、游程长度不均匀性、游程长度分布、游程比例、均值、均方差、能量、熵等
		灰度游程矩阵	稀疏度、繁忙度、复杂度、对比度、纹理力度
		灰度区域大小矩阵	灰度不均匀性、归一化灰度不均匀性、区域大小不均匀性、区域百分比、灰度方差、区域方差、区域熵等
		灰度相关矩阵	灰度不均匀性、灰度方差、依赖方差等

体的算法流程如下。

假设有一个特征数据集,其样本数量为 $n$ ,特征数量为 $p$ ,将其组合成一个大小为 $n \times p$ 矩阵 $\mathbf{x}$ ,如公式(1)所示:

$$\mathbf{x} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = (x_1, x_2, \cdots, x_p)$$

(1)

由于每个特征的度量单位不一致的情况容易影响后续的特征降维,所以先对数据集中的每一个特征进行标准化处理,即按列计算出每个特征的均值和标准差,然后通过计算得到标准化数据,使标准化数据的每一个特征减去对应该列特征的平均值,得到新的矩阵 $\mathbf{X}$ ,如公式(2)所示:

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}$$

(2)

然后计算标准化样本的协方差矩阵 $\mathbf{R}$ ,如公式(3)所示:

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{np} \end{bmatrix}$$

(3)

其中  $r_{ij} = \frac{1}{n-1} \sum_{k=1}^n (X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j) = \frac{1}{n-1} \sum_{k=1}^n X_{ki}X_{kj}$ , 接下来算出矩阵 $\mathbf{R}$ 的特征值和特征向量,再根据其累计贡献率保留前 $k$ 个向量作为主成分,去除多余的向量,最后通过计算将原高维空间数据转换到 $k$ 个特征向量构建的低维空间中。

一般来说,贡献率越大,表明该成分综合的信息越多。而为了达到降维又尽可能地保证原数据信息的完整性,保留累计贡献率达到较高值的 $k$ 个特征向量作为主成分。本试验选取累计贡献率超过80%的前几个特征作为主成分,在保证原数据信息完整的条件下,去除噪声和冗余信息,达到特征降维并优化的效果。在试验中,将原来提取的1 648个影像组学特征通过PCA方法降维到60个更具代表性的主成分特征,既达到了特征压缩的目的,又能充分代表原数据的信息,为后续分类预测模型的建立奠定了良好的基础。

经过PCA特征降维之后,再使用Relief、ANOVA和RFE算法进行特征选择,最后得到特征冗余量小并与淋巴结转移更显著相关的特征子集。

**1.2.4 预测分类模型的建立和验证** 为了得到最优的预测分类模型,进行了大量的试验研究和测试,最终采用了5种泛化性强、适合小样本数据的代表性分类算法,包括朴素贝叶斯(Naïve Bayes, NB)、线性判别分析(Linear Discriminant Analysis, LDA)、逻辑回归(Logistic Regression, LR)、支持向量机(Support Vector Machine, SVM)和高斯过程(Gaussian Process, GP),再用优化后的训练集数据训练分类器,接着用测试集数据进行分类预测研究。此外,在试验中还引入了5折交叉验证的方法,该方法既能提升模型的可靠性,还能在很大程度上提高现有数据的利用率。

2 结果

使用基于5种不同的机器学习分类算法和5折交叉验证的方法进行分类试验和模型评估,并使用准确率(Accuracy, ACC)、受试者操作特征曲线(Receiver Operative Characteristic Curve, ROC)的曲线下面积(Area Under Curve, AUC)、敏感性(Sensitivity, SEN)及特异性(Specificity, SPE)这些指标评价分类结果的好坏。在训练集和测试集上的评价结果如表3所示,5种分类器模型最优子类型的ROC曲线如图3所示。

表 3 5 种分类器的分类结果  
Table 3 Classification results of 5 classifiers

分类器	特征选择方法	训练集		测试集			
		准确率	AUC	准确率	AUC	敏感性	特异性
NB	ANOVA	0.741	0.784	0.692	0.686	0.750	0.631
	Relief	0.802	0.852	0.795	0.810	0.725	0.869
	RFE	0.857	0.921	0.731	0.791	0.525	0.947
LDA	ANOVA	0.764	0.847	0.705	0.755	0.675	0.737
	Relief	0.507	0.610	0.692	0.565	0.925	0.447
	RFE	0.775	0.845	0.692	0.676	0.600	0.790
LR	ANOVA	0.747	0.816	0.744	0.776	0.725	0.684
	Relief	0.660	0.667	0.590	0.555	0.950	0.211
	RFE	0.747	0.791	0.705	0.678	0.725	0.684
SVM	ANOVA	0.775	0.832	0.744	0.773	0.925	0.553
	Relief	0.539	0.348	0.513	0.322	0.975	0.026
	RFE	0.786	0.845	0.692	0.677	0.600	0.790
GP	ANOVA	0.753	0.822	0.731	0.772	0.875	0.579
	Relief	0.742	0.788	0.692	0.640	0.750	0.632
	RFE	0.764	0.832	0.692	0.676	0.600	0.790

NB:朴素贝叶斯;LDA:线性判别分析;LR:逻辑回归;SVM:支持向量机;GP:高斯过程;ANOVA:方差分析;RFE:递归特征消除;AUC:受试者工作特征曲线下的面积

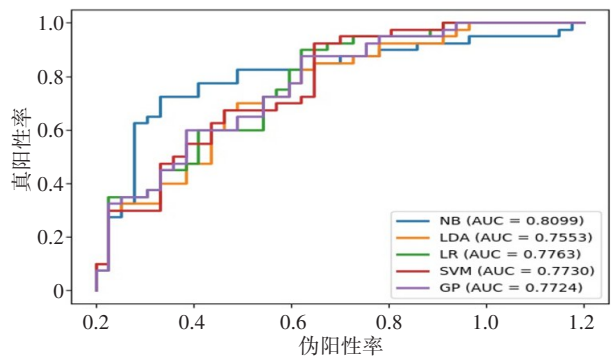


图 3 5 种模型的最优子类型 ROC 曲线图  
Figure 3 ROC curves of the optimal subtypes of 5 models

为了得到一个稳定有效的模型,并且保证该模型在训练时不会出现过拟合现象。于是对多个分类器进行训练,在保证较高的准确率和 AUC 值的情况下,还要使模型的敏感性和特异性保持在一种较为均衡的状态。如上面试验结果所示,用不同模型进行训练和预测时,采用 Relief 特征选择方法下的 NB 预测模型为本试验中最好的模型,在所有模型中,此模型下测试集的准确率和 AUC 的值均为最高,分别为 0.795 和 0.810,敏感性和特异性也处于较高水平,分别为 0.725 和 0.869,表明通过该影像组学模型能较准确地对 NSCLC 患者的淋巴结转移情况进行预测。

3 讨论

本研究基于影像组学的方法,通过高通量计算,从 CT 图像中提取定量特征,将数字医学图像转换为可挖掘分析的高维特征数据,然后通过数据统计分析探究 CT 图像的影像组学特征与 NSCLC 患者淋巴结转移的相关联系,并以多中心实验数据开发并验证一个稳健有效的影像组学预测模型,进一步实现对 NSCLC 患者淋巴结转移状况无创准确的预测。其中所创建的最优影像组学模型在测试集上(AUC=0.810)具有良好的鉴别分类能力,并且其预测 NSCLC 患者 LNM 的有效性显著高于 CT 报告中淋巴结状态模型和形态学模型<sup>[22-23]</sup>。结果表明,影像组学特征可以为我们提供更丰富的信息,以其为基础建立的模型在很大程度上提高 NSCLC 患者 LNM 的预测准确率,从而辅助医生做出更精准的决策,因此在临床应用上具有重要的指导意义。本研究有如下创新点,第一,我们所开发的影像组学预测模型是基于多个中心获得的数据建立的,因此可以从外部数据验证预测模型的稳健性和可重复性。第二,从 CT 医学影像提取 1 648 个定量影像组学特征,并进行有效的统计分析,更全面准确地反映 CT 图像中的肿瘤信息,大大提高了影像组学模型的预测准确性。第三,采用了 SMOTE 算法,解决了模型训练过程中的数据不平衡问

题,使试验得到的预测模型更加稳定和准确。第四,采用PCA特征降维和Relief、ANOVA和RFE特征选择方法做特征优化处理,最后获得相关性最高的最优特征,为建立更准确的预测模型奠定良好的基础。最后,通过5折交叉验证的方法对NB, LDA, LR, SVM和GP 5种泛化性强、适合小样本数据的代表性分类算法进行分类训练,再用外部数据评估验证模型的准确性和稳定性,这些方法既能降低数据集产生的偶然性,提高模型的泛化能力,还能高效地利用现有数据,并通过试验结果对比得到性能最优的预测模型。

目前已有学者开发相似的基于影像组学的预测模型用于预测NSCLC患者的LNM状态,Cong等<sup>[24]</sup>从411例NSCLC患者的CT图像中提取影像组学特征,然后使用ANOVA和最小绝对收缩和选择运算(LASSO)方法选择最优的10个影像组学特征,并使用LR算法建立预测模型。朱静等<sup>[25]</sup>从200例样本中提取影像组学特征,使用LASSO和LR方法建立预测模型。然后将文献[24-25]中用到的方法建立的预测模型结果与本文相比,以预测模型的AUC值作为评估指标进行对比,其对比结果如表4所示。从结果上看,本文采用的算法所开发的预测模型的AUC值比其他两者都高,说明本文提出的预测模型分类性能更好,准确度更高。

表4 不同方法的预测模型结果比较

Table 4 Comparison of results of prediction models with different methods

方法	分类算法	特征选择方法	AUC
Cong等 <sup>[24]</sup>	LR	ANOVA+LASSO	0.730
朱静等 <sup>[25]</sup>	LR	LASSO	0.776
本文方法	NB	PCA+Relief	0.810

LASSO:最小绝对收缩和选择运算

虽然本文开发的预测模型分类鉴别效果良好,但是还是存在一些不足之处。一、本试验研究所用到的数据样本量比较少,今后需要获取更多的试验数据进行下一步测试和验证。二、仅基于影像组学特征建立预测模型,没有将患者的临床和基因组学特征纳入试验研究中。总之,我们开发并验证了一种基于影像组学的淋巴结预测模型,可以对NACLC患者的淋巴结状态进行无创并有效地评估,并辅助临床决策。

【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] PALMA J F, DAS P, LIESENFELD O. Lung cancer screening: utility of molecular applications in conjunction with low-dose computed

tomography guidelines[J]. Expert Rev Mol Diagn, 2016, 16(4): 435-447.

[3] ETTINGER D S, WOOD D E, AISNER D L, et al. Non-small cell lung cancer, version 5.2017, NCCN clinical practice guidelines in oncology[J]. J Natl Compr Canc Netw, 2017, 15(4): 504-535.

[4] ETTINGER D S, AKERLEY W, BEPLER G, et al. Non-small cell lung cancer[J]. J Natl Compr Canc Netw, 2010, 8(7): 740-801.

[5] SILVESTRI G A, GOULD M K, MARGOLIS M L, et al. Noninvasive staging of non-small cell lung cancer: ACCP evidenced-based clinical practice guidelines[J]. Chest, 2007, 132(3): 178S-201S.

[6] OKADA M, TSUTANI Y, IKEDA T, et al. Radical hybrid video-assisted thoracic segmentectomy: long-term results of minimally invasive anatomical sublobar resection for treating lung cancer[J]. Interact Cardiovasc Thorac Surg, 2012, 14(1): 5-11.

[7] RAMI-PORTA R, BALL D, CROWLEY J, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the T descriptors in the forthcoming (seventh) edition of the TNM classification for lung cancer[J]. J Thorac Oncol, 2007, 2(7): 593-602.

[8] WALKER C M, CHUNG J H, ABBOTT G F, et al. Mediastinal lymph node staging: from noninvasive to surgical[J]. AJR Am J Roentgenol, 2012, 199(1): W54-W64.

[9] SZLUBOWSKI A, ZIELIŃSKI M, SOJA J, et al. Accurate and safe mediastinal restaging by combined endobronchial and endoscopic ultrasound-guided needle aspiration performed by single ultrasound bronchoscope[J]. Eur J Cardiothorac Surg, 2014, 46(2): 262-266.

[10] ROBERTS P F, FOLLETTE D M, VON HAAG D, et al. Factors associated with false-positive staging of lung cancer by positron emission tomography[J]. Ann Thorac Surg, 2000, 70(4): 1154-1159.

[11] KANZAKI R, HIGASHIYAMA M, FUJIWARA A, et al. Occult mediastinal lymph node metastasis in NSCLC patients diagnosed as clinical N0-1 by preoperative integrated FDG-PET/CT and CT: risk factors, pattern, and histopathological study[J]. Lung Cancer, 2011, 71(3): 333-337.

[12] VERDUZCO-AGUIRRE H C, LOPES G, SOTO-PEREZ-DE-CELI E. Implementation of diagnostic resources for cancer in developing countries: a focus on PET/CT[J]. Ecancermedicalscience, 2019, 13. DOI:10.3332/ecancer.2019.ed87.

[13] GLASGOW S C, BLEIER J I, BURGART L J, et al. Meta-analysis of histopathological features of primary colorectal cancers that predict lymph node metastases[J]. J Gastrointest Surg, 2012, 16(5): 1019-1028.

[14] SMITH A J, DRIMAN D K, SPITHOFF K, et al. Guideline for optimization of colorectal cancer surgery and pathology[J]. J Surg Oncol, 2010, 101(1): 5-12.

[15] TOIYAMA Y, INOUE Y, SHIMURA T, et al. Serum angiopoietin-like protein 2 improves preoperative detection of lymph node metastasis in colorectal cancer[J]. Anticancer Res, 2015, 35(5): 2849-2856.

[16] LAMBIN P, LEJENNAAR R T, DEIST T M, et al. Radiomics: the bridge between medical imaging and personalized medicine[J]. Nat Rev Clin Oncol, 2017, 14(12): 749-762.

[17] LIMKIN E J, SUN R, DERCLE L, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology[J]. Ann Oncol, 2017, 28(6): 1191-1206.

[18] PRIOR F W, CLARK K, COMMEAN P, et al. TCIA: an information resource to enable open science[C]//2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2013: 1282-1285.

[19] YUSHKEVICH P A, PIVEN J, HAZLETT H C, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability[J]. Neuroimage, 2006, 31(3): 1116-1128.

[20] 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.

[21] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique[J]. J Artif Intell Res, 2002, 16: 321-357.

[22] HUANG Y, LIANG C, HE L, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer[J]. J Clin Oncol, 2016, 34(18): 2157-2164.

[23] JI G W, ZHANG Y D, ZHANG H, et al. Biliary tract cancer at CT: a radiomics-based model to predict lymph node metastasis and survival outcomes[J]. Radiology, 2019, 290(1): 90-98.

[24] CONG M, YAO H, LIU H, et al. Development and evaluation of a venous computed tomography radiomics model to predict lymph node metastasis from non-small cell lung cancer[J]. Medicine, 2020, 99(18): e20074.

[25] 朱静,徐维国,肖欢,等.影像组学联合预测模型在预测非小细胞肺癌淋巴结转移中的建立和应用价值[J].四川大学学报(医学版), 2019, 50(3): 373-378.

ZHU J, XU W G, XIAO H, et al. Establishment and application value of combined imaging omics prediction model in predicting lymph node metastasis of non-small cell lung cancer[J]. Journal of Sichuan University (Medical Science Edition), 2019, 50(3): 373-378.

(编辑:薛泽玲)