

# 机器学习结合<sup>18</sup>F-FDG PET/CT 影像组学特征对肺腺癌 EGFR 突变亚型的预测价值

王子阳 尹国涛 李小凤 徐文贵

天津医科大学肿瘤医院分子影像及核医学诊疗科、国家肿瘤临床医学研究中心、天津市肿瘤防治重点实验室、天津市恶性肿瘤临床研究中心 300060

通信作者:徐文贵, Email: wenguixy@163.com

**【摘要】** 目的 探究<sup>18</sup>F-脱氧葡萄糖(FDG) PET/CT 影像结合机器学习算法对肺腺癌(LUAD)表皮生长因子受体(EGFR)突变亚型的预测价值。方法 回顾性收集 2016 年 4 月至 2020 年 5 月于天津医科大学肿瘤医院行<sup>18</sup>F-FDG PET/CT 检查的 238 例 LUAD 患者的影像及病理资料,其中 EGFR 突变型为 126 例(男 54 例,女 72 例,中位年龄 62 岁);EGFR 野生型 112 例(男 68 例,女 44 例,中位年龄 61 岁)。分别对 PET、CT 图上感兴趣体积(VOI)进行三维勾画,提取三维和二维影像组学特征。使用 K-最近邻(KNN)、支持向量机(SVM)和 Adaboost 分别对 CT、PET 和 PET/CT 融合组学特征进行学习,并对 EGFR 突变亚型进行预测。使用受试者工作特征(ROC)曲线对预测性能进行评估。结果 126 例 EGFR 突变型患者中 3 例为 18 号外显子突变,6 例为 20 号外显子突变,42 例为 19 号外显子突变,75 例为 21 号外显子突变。前 2 个亚型患者因数量较少,分类器难以进行充分的训练而被排除。PET/CT 平均融合特征模型预测 EGFR 突变亚型的效果[Adaboost;曲线下面积(AUC)=0.87, 95% CI: 0.75~0.99]优于 PET 特征模型(Adaboost;AUC=0.64, 95% CI:0.46~0.83;  $z=2.04, P<0.05$ )和 CT 特征模型(Adaboost;AUC=0.64, 95% CI:0.45~0.83;  $z=2.06, P<0.05$ )。PET/CT 平均融合特征模型(SVM;AUC=0.76, 95% CI:0.56~0.96)和 PET/CT 拼接融合特征模型(SVM;AUC=0.75, 95% CI:0.59~0.92)预测 EGFR 突变亚型的效果差异无统计学意义( $z=1.14, P>0.05$ )。结论 机器学习结合<sup>18</sup>F-FDG PET/CT 融合影像组学特征在 EGFR 突变亚型预测中有一定的价值。

**【关键词】** 肺肿瘤;腺癌;受体,表皮生长因子;突变;正电子发射断层显像术;体层摄影术,X 线计算机;脱氧葡萄糖;预测

DOI:10.3760/cma.j.cn321828-20201105-00401

## Value of machine learning and <sup>18</sup>F-FDG PET/CT radiomics features in lung adenocarcinoma EGFR mutation subtypes prediction

Wang Ziyang, Yin Guotao, Li Xiaofeng, Xu Wengui

Department of Molecular Imaging and Nuclear Medicine, Tianjin Medical University Cancer Institute and Hospital; National Clinical Research Center for Cancer; Tianjin Key Laboratory of Cancer Prevention and Therapy; Tianjin's Clinical Research Center for Cancer, Tianjin 300060, China

Corresponding author: Xu Wengui, Email: wenguixy@163.com

**【Abstract】** **Objective** To explore the predictive values for mutation subtypes of epidermal growth factor receptor (EGFR) in patients with lung adenocarcinoma (LUAD) based on machine learning and <sup>18</sup>F-fluorodeoxyglucose (FDG) PET/CT images. **Methods** <sup>18</sup>F-FDG PET/CT images and pathological data of 238 patients with LUAD (126 patients (54 males, 72 females, median age 62 years) with EGFR mutation; 112 patients (68 males, 44 females, median age 61 years) with wild-type EGFR) were retrospectively collected at Tianjin Medical University Cancer Institute and Hospital between April 2016 and May 2020. Volumes of interest (VOI) of PET and CT images were delineated respectively and three-dimensional-based and two-dimensional-based radiomics features were extracted from VOIs. Three machine learning classifiers of K-nearest neighbor (KNN), support vector machine (SVM) and Adaboost were trained in training set with CT, PET and fusion PET/CT radiomics features respectively. Well trained classifiers were tested in test set. Each predictive model was evaluated by using the receiver operating characteristic (ROC) curve. **Results** A total of 126 patients were EGFR mutation including 3 patients with 18 exon mutation, 6 patients with 20 exon mutation, 42 patients with 19 exon mutation, and 75 patients with 21 exon mutation. Finally, patients with 18 exon mutation and 20 exon mutation were removed due to the scale was too small to be trained adequately by machine learning classifiers. Predictive performance of mean PET/CT feature-based model (Adaboost; area under curve (AUC)=0.87, 95% CI: 0.75-0.99) in EGFR mutation subtypes was

better than PET feature-based model (Adaboost; AUC=0.64, 95% CI: 0.46–0.83;  $z=2.04$ ,  $P<0.05$ ) and CT feature-based model (Adaboost; AUC=0.64, 95% CI: 0.45–0.83;  $z=2.06$ ,  $P<0.05$ ). There was no statistical difference between predictive performance of mean PET/CT feature-based model (SVM; AUC=0.76, 95% CI: 0.56–0.96) and PET/CT concatenation feature-based model (SVM; AUC=0.75, 95% CI: 0.59–0.92;  $z=1.14$ ,  $P>0.05$ ). **Conclusion** Machine learning and  $^{18}\text{F}$ -FDG PET/CT radiomics features can provide predictive value for EGFR mutation subtypes in patients with LUAD.

**【Key words】** Lung neoplasms; Adenocarcinoma; Receptor, epidermal growth factor; Mutation; Positron-emission tomography; Tomography, X-ray computed; Deoxyglucose; Forecasting

DOI:10.3760/cma.j.cn321828-20201105-00401

近年来,由于表皮生长因子受体(epidermal growth factor receptor, EGFR)酪氨酸激酶抑制剂(tyrosine kinase inhibitors, TKIs)的推广与应用<sup>[1]</sup>,晚期肺腺癌(lung adenocarcinoma, LUAD)患者的预后得到较大的改善<sup>[2]</sup>。不同EGFR突变亚型肿瘤对EGFR-TKIs治疗的敏感性不同<sup>[3-4]</sup>,因此在治疗前准确诊断EGFR突变亚型对改善晚期LUAD患者预后具有重要的临床意义。

目前,EGFR突变亚型的诊断主要依靠对活组织检查(简称活检)或手术切除组织样本进行基因检测。但是,用于测序的少部分病理组织无法全面反映肿瘤的整体信息,并且部分病灶位置相对隐匿以及肿瘤内部广泛存在异质性等因素,在一定程度上限制了基因检测的应用<sup>[5]</sup>,另外,穿刺活检会在一定程度上增加肿瘤转移的风险。既往研究显示, $^{18}\text{F}$ -脱氧葡萄糖(flurodeoxyglucose, FDG) PET/CT参数在预测EGFR的突变状态方面具有一定的潜力<sup>[6-8]</sup>。目前,通过机器学习结合 $^{18}\text{F}$ -FDG PET/CT影像组学特征预测EGFR突变亚型的研究鲜有报道。因此,笔者开展了这方面的研究。

## 资料与方法

1.研究对象。回顾性收集2016年4月至2020年5月于天津医科大学肿瘤医院诊断为LUAD患者的 $^{18}\text{F}$ -FDG PET/CT影像资料与临床信息。纳入标准:(1)原发病灶经手术或活检病理证实为LUAD且进行EGFR基因检测的患者;(2)进行 $^{18}\text{F}$ -FDG PET/CT检查和EGFR基因检测的时间间隔 $<2$ 个月;(3)病灶边界清晰且有明显的 $^{18}\text{F}$ -FDG浓聚。排除标准:(1)既往有肿瘤病史;(2)进行 $^{18}\text{F}$ -FDG PET/CT检查前已接受抗肿瘤治疗。本研究符合《赫尔辛基宣言》原则。

2.图像采集。使用美国GE Discovery Elite PET/CT仪进行图像采集。显像前患者禁食约6 h,控制血糖水平 $<7.0$  mmol/L。 $^{18}\text{F}$ -FDG(美国GE MINI trace 医用回旋加速器生产,放化纯 $>95\%$ )注射剂量为按患者体质量4.2 MBq/kg,注射后50~60 min开始

图像采集。采用螺旋CT(电流180 mA、电压130 kV、层厚1.25 mm)进行精确解剖定位和胸部薄层扫描;采用低剂量CT(电流50~80mA、电压120 kV)进行衰减校正;PET扫描范围从股骨远端至颅骨顶部,扫描时间为2 min/床位,共扫描8个床位;最终采用迭代法进行图像重建。重建完成后,以标准摄取值(standardized uptake value, SUV)2.5为阈值,使用GE Advantage Workstation 4.6对患者原发病灶的最大SUV(maximum SUV,  $\text{SUV}_{\max}$ )、平均SUV(mean SUV,  $\text{SUV}_{\text{mean}}$ )、SUV峰值(peak of SUV,  $\text{SUV}_{\text{peak}}$ )、肿瘤代谢体积(metabolic tumor volume, MTV)、病灶糖酵解总量(total lesion glycolysis, TLG)进行测量。

3.基因突变检测。用于基因检测的组织标本通过穿刺活检或手术切除获得。EGFR突变亚型由天津医科大学肿瘤医院病理科进行分析。采用实时定量PCR扩增EGFR激酶结构域(18~24号外显子)的核苷酸序列,通过在琼脂糖凝胶(2 g/100 ml)上解析PCR产物来确认适当PCR产物的存在。产物经纯化后,使用ABI prismr9700和ABI prismr310基因分析仪(Applied Biosystems,美国)对凝胶上相应的片段进行正义和反义测序。利用SeqScape(Applied Biosystems,美国)对序列数据进行分析,鉴定突变状态和突变亚型。

4. PET/CT数据预处理。由2位有3年以上PET/CT诊断经验的医师使用三维 slicer 4.10.2<sup>[9]</sup>对纳入患者PET和CT图像的感兴趣体积(volume of interest, VOI)进行手动勾画,并由另一位有10年以上PET/CT诊断经验的医师对VOI的勾画结果进行查验与校对。

使用Python 3.7.1中Pyradiomics模块<sup>[10]</sup>分别对PET和CT图像进行相应的滤波预处理。采用指数(Exponential)、梯度(Gradient)、高斯拉普拉斯算子(Laplacian of Gaussian, LoG)、对数(Logarithm)、平方(Square)、平方根(Square root)和小波(Wavelet)滤波对图像进行预处理。其中,小波滤波由基于PET/CT图像进行各维度的高通(H)和低通(L)

滤波组合组成,包括 Wavelet-LLH、Wavelet-LHL、Wavelet-LHH、Wavelet-HLL、Wavelet-HLH、Wavelet-HHL、Wavelet-LLL 和 Wavelet-HHH。

5.影像组学特征提取。使用 Python 3.7.1 中 Pyradiomics 模块<sup>[10]</sup>提取 PET/CT 影像组学特征。从 PET/CT 原始图像中提取三维和二维形状特征;从 PET/CT 预处理图像和原始图像中提取一阶(first order, FO)、灰度共生矩阵(gray level cooccurrence matrix, GLCM)、灰度游程长度矩阵(gray level run length matrix, GLRLM)、灰度区域大小矩阵(gray level size zone matrix, GLSZM)、灰度相关矩阵(gray level dependence matrix, GLDM)特征。

6.影像组学特征筛选。使用 Python 3.7.1 中 sklearn 和 statsmodels<sup>[11]</sup>模块对影像组学特征进行筛选。分别通过 PET、CT 和 PET/CT 融合影像组学特征对 EGFR 基因突变亚型进行预测。PET/CT 融合特征通过用特征拼接和特征平均的方法对 PET 和 CT 影像学特征进行融合获得;同时纳入同一患者的 CT 特征和 PET 特征为 PET/CT 拼接特征(PET/CTconcat)融合方式;将同一患者对应的 PET 和 CT 特征数值相加取平均值为 PET/CT 平均特征(PET/CTmean)融合方式。筛除方差小于 0.5 的特征。使用 Mann-Whitney *U* 检验对保留的特征进行差异比较,保留不同外显子突变之间差异有统计学意义的特征。使用决策树(decision tree)依据基尼系数(Gini)对保留的特征进行筛选。最后,通过计算方差膨胀因子(variance inflation factor, VIF)来评价放射学特征之间的共线性。循环计算所有保留特征的 VIF,筛除 VIF 最高的影像组学特征,直到 VIF 最大值<4,停止循环,输出筛选后的特征。

7.构建机器学习模型。使用 Python 3.7.1 中 sklearn 模块进行机器学习模型的搭建。共构建 3 种经典的机器学习分类器,分别使用 K-最近邻(K-nearest neighbor, KNN)、支持向量机(support vector machine, SVM)和 Adaboost 进行 EGFR 突变亚型的测试。随机抽选 70% 的患者用于模型训练,剩余 30% 的患者用于模型测试,使用十折交叉验证减少过拟合,使用网格搜索法进行参数调整,优化模型预测性能。

8.统计学处理。使用 R 3.6.0 和 Python 3.7.1 软件进行统计学分析。符合正态分布的定量资料以  $\bar{x}\pm s$  表示,2 组间比较使用两独立样本 *t* 检验,多组间比较使用单因素方差分析;不符合正态分布的定量资料以  $M(P_{25}, P_{75})$  表示,多组间比较使用 Kruskal-

Wallis 秩和检验,进一步的两两比较用 Nemenyi 检验。定性资料的比较使用  $\chi^2$  检验。使用 Spearman 相关分析定量资料之间的相关性。对各预测模型的预测结果绘制受试者工作特征(receiver operating characteristic, ROC)曲线,计算曲线下面积(area under curve, AUC)进行预测性能评估。两分类器之间 ROC 曲线的比较使用 Delong 检验。 $P<0.05$  为差异或相关性有统计学意义。

## 结 果

1.患者临床信息。共收集 238 例 LUAD 患者,其中 126 例(男 54 例,女 72 例,中位年龄 62 岁)为 EGFR 突变型,112 例(男 68 例,女 44 例,中位年龄 61 岁)为 EGFR 野生型。在 126 例 EGFR 为突变型的患者中,3 例为 18 号外显子突变,42 例为 19 号外显子突变,6 例为 20 号外显子突变,75 例为 21 号外显子突变。EGFR 各个突变亚型的患者在年龄( $H=1.82, P>0.05$ )和性别( $\chi^2=2.57, P>0.05$ )上差异无统计学意义。由于 EGFR 18 号外显子突变和 20 号外显子突变的患者数量较少,分类器难以进行充分的训练,故将其排除在外。最终纳入 229 例 LUAD 患者,其中 EGFR 19 号外显子和 21 号外显子突变的患者共 117 例,用于本研究中 EGFR 突变亚型预测。纳入患者信息见表 1。

2.影像组学特征选择。共提取 1 218 个 PET 特征和 1 218 个 CT 特征。经过特征筛选后,保留了 11 个 CT 影像组学特征,包括 1 个梯度特征、1 个对数特征和 9 个小波特征;5 个 PET 影像组学特征,包括 1 个原始形状特征、1 个对数特征、1 个梯度特征和 2 个小波特征;10 个 PET/CT 拼接融合特征,包括 1 个 CT 平方特征、8 个 CT 小波特征和 1 个 PET 小波特征;7 个 PET/CT 平均融合特征,包括 1 个指数特征、1 个梯度特征、2 个平方特征、1 个平方根特征和 2 个小波特征。各模式下影像组学特征间的 Spearman 相关性热图见图 1。

3. PET 传统代谢参数对 EGFR 突变亚型的预测价值。PET 传统代谢参数  $SUV_{max}$ 、 $SUV_{mean}$ 、 $SUV_{peak}$ 、MTV、TLG 对 117 例 EGFR 突变型患者突变亚型预测的 AUC 分别为 0.52 (95% CI: 0.42 ~ 0.63,  $P>0.05$ )、0.51 (95% CI: 0.41 ~ 0.61,  $P>0.05$ )、0.53 (95% CI: 0.43 ~ 0.64,  $P>0.05$ )、0.54 (95% CI: 0.43 ~ 0.65,  $P>0.05$ ) 和 0.54 (95% CI: 0.43 ~ 0.65,  $P>0.05$ )。

4.影像组学特征对 EGFR 突变状态和突变亚型的预测价值。KNN、SVM 和 Adaboost 在 229 例 LUAD

表 1 3 组肺腺癌患者的临床信息

组别	例数	性别(例)		年龄 [岁;M(P <sub>25</sub> ,P <sub>75</sub> )]	吸烟史(例)		病灶位置 <sup>c</sup> (例)					病灶直径 (cm; $\bar{x} \pm s$ )
		男	女		有	无	右肺上叶	右肺下叶	右肺中叶	左肺上叶	左肺下叶	
A	42	19	23	63(54,65)	13	29	12	12	8	9	3	4.05±1.35
B	75	27	48	62(56,67)	18	57	34	10	6	20	7	4.21±1.33
C	112	68	44	60(53,65)	47	65	35	26	6	29	17	4.41±1.79
检验值	$\chi^2$ 值:11.4 <sup>a</sup> 、0.61 <sup>b</sup> 、2.38 <sup>c</sup> 、10.01 <sup>d</sup>		H=2.73		$\chi^2$ 值:6.67 <sup>a</sup> 、0.35 <sup>b</sup> 、1.12 <sup>c</sup> 、5.62 <sup>d</sup>		$\chi^2=6.00$ (校正后)					F=0.99
P 值	<0.05 <sup>a,d</sup> 、>0.05 <sup>b,c</sup>		>0.05		<0.05 <sup>a,d</sup> 、>0.05 <sup>b,c</sup>		>0.05					>0.05

组别	例数	病理类型(例)						SUV <sub>max</sub> [M(P <sub>25</sub> ,P <sub>75</sub> )]
		附壁型	乳头型	实体型	微乳头型	腺泡型	数据缺失	
A	42	11	0	3	3	20	5	23.31(17.15,31.61)
B	75	18	6	5	4	30	12	26.95(15.75,34.53)
C	112	10	5	19	2	27	49	26.98(13.59,36.81)
检验值	$\chi^2$ 值:43.41 <sup>a</sup> 、4.24 <sup>b</sup> 、27.67 <sup>c</sup> 、27.56 <sup>d</sup>						H=0.60	
P 值	<0.05 <sup>a,c,d</sup> 、>0.05 <sup>b</sup>						>0.05	

组别	例数	SUV <sub>mean</sub> [M(P <sub>25</sub> ,P <sub>75</sub> )]	SUV <sub>peak</sub> [M(P <sub>25</sub> ,P <sub>75</sub> )]	MTV[cm <sup>3</sup> ;M(P <sub>25</sub> ,P <sub>75</sub> )]	TLG[g;M(P <sub>25</sub> ,P <sub>75</sub> )]	
A	42	5.28(4.31,6.71)	17.13(11.80,25.17)	36.53(14.76,65.29)	175.52(84.91,378.74)	
B	75	5.28(4.31,6.71)	19.99(10.52,26.08)	44.33(19.78,63.56)	220.34(97.69,375.57)	
C	112	6.60(4.59,8.60)	20.68(10.58,29.55)	41.66(15.98,91.30)	249.39(91.30,659.37)	
检验值	H=8.56 <sup>a</sup> 、D=0.99 <sup>b</sup> 、D=6.80 <sup>d</sup>		H=1.59		H=1.47	
P 值	<0.05 <sup>a,d</sup> 、>0.05 <sup>b,c</sup>		>0.05		>0.05	

注:A 组为表皮生长因子受体(EGFR)突变(19 号外显子)组,B 组为 EGFR 突变(21 号外显子)组,C 为 EGFR 野生型组;<sup>a</sup>3 组间比较,<sup>b</sup>A 组与 B 组比较,<sup>c</sup>A 组与 C 组比较,<sup>d</sup>B 组与 C 组比较;<sup>e</sup>对于跨肺叶病灶,病灶所跨各肺叶均计数;MTV 为肿瘤代谢体积,SUV<sub>max</sub> 为最大标准摄取值,SUV<sub>mean</sub> 为平均标准摄取值,SUV<sub>peak</sub> 为标准摄取峰值,TLG 为病灶糖酵解总量

患者 EGFR 突变状态的预测中,CT 特征预测的 AUC 分别为 0.69(95% CI:0.57~0.81, P<0.05)、0.72(95% CI:0.60~0.85, P<0.05)和 0.72(95% CI:0.60~0.84, P<0.05);PET 特征预测的 AUC 分别为 0.69(95% CI:0.57~0.81, P<0.05)、0.68(95% CI:0.55~0.81, P<0.05)和 0.67(95% CI:0.54~0.80, P<0.05);PET/CT 拼接融合特征预测的 AUC 分别为 0.72(95% CI:0.59~0.84, P<0.05)、0.73(95% CI:0.61~0.85, P<0.05)和 0.74(95% CI:0.62~0.85, P<0.05);PET/CT 平均融合特征预测的 AUC 分别为 0.70(95% CI:0.58~0.82, P<0.05)、0.67(95% CI:0.54~0.80, P<0.05)和 0.72(95% CI:0.60~0.84, P<0.05)。ROC 曲线见图 2。

KNN、SVM 和 Adaboost 在 117 例 LUAD 患者 EGFR 突变亚型预测中,CT 特征预测的 AUC 分别为 0.60(95% CI:0.42~0.79, P>0.05)、0.60(95% CI:0.40~0.80, P>0.05)和 0.64(95% CI:0.45~0.83, P>0.05);PET 特征预测的 AUC 分别为 0.63(95% CI:0.43~0.84, P>0.05)、0.59(95% CI:0.26~0.81, P>0.05)和 0.64(95% CI:0.46~0.83, P>0.05);PET/CT 拼接融合特征预测的 AUC 分别为 0.73(95% CI:

0.54~0.92, P<0.05)、0.75(95% CI:0.59~0.92, P<0.05)和 0.70(95% CI:0.48~0.91, P<0.05);PET/CT 平均融合特征预测的 AUC 分别为 0.72(95% CI:0.55~0.89, P<0.05)、0.76(95% CI:0.56~0.96, P<0.05)和 0.87(95% CI:0.75~0.99, P<0.05)。ROC 曲线见图 3。2 种 PET/CT 融合特征的预测表现之间差异无统计学意义(z=1.14, P>0.05)。PET/CT 平均融合特征的预测表现要优于 CT 特征(z=2.06, P<0.05)和 PET 特征的预测表现(z=2.04, P<0.05)。

### 讨论

LUAD 是非小细胞肺癌中最常见的病理亚型<sup>[12]</sup>。EGFR-TKIs 的出现明显改善了晚期 LUAD 患者的预后。然而,不同 EGFR 突变亚型的患者对 EGFR-TKIs 治疗的敏感性不同。既往研究报道,与 21 号外显子突变相比,19 号外显子突变的晚期 LUAD 患者有更高的 EGFR-TKIs(如吉非替尼和厄罗替尼)治疗敏感性,进而有更好的预后<sup>[13-14]</sup>。在治疗前准确诊断 EGFR 的突变亚型有助于筛选对 EGFR-TKIs 治疗更敏感的 LUAD 患者。

既往文献报道,肺癌的 EGFR 突变常见于 LUAD、女性、亚洲人群体,并且超过 80% 的 EGFR 突变为

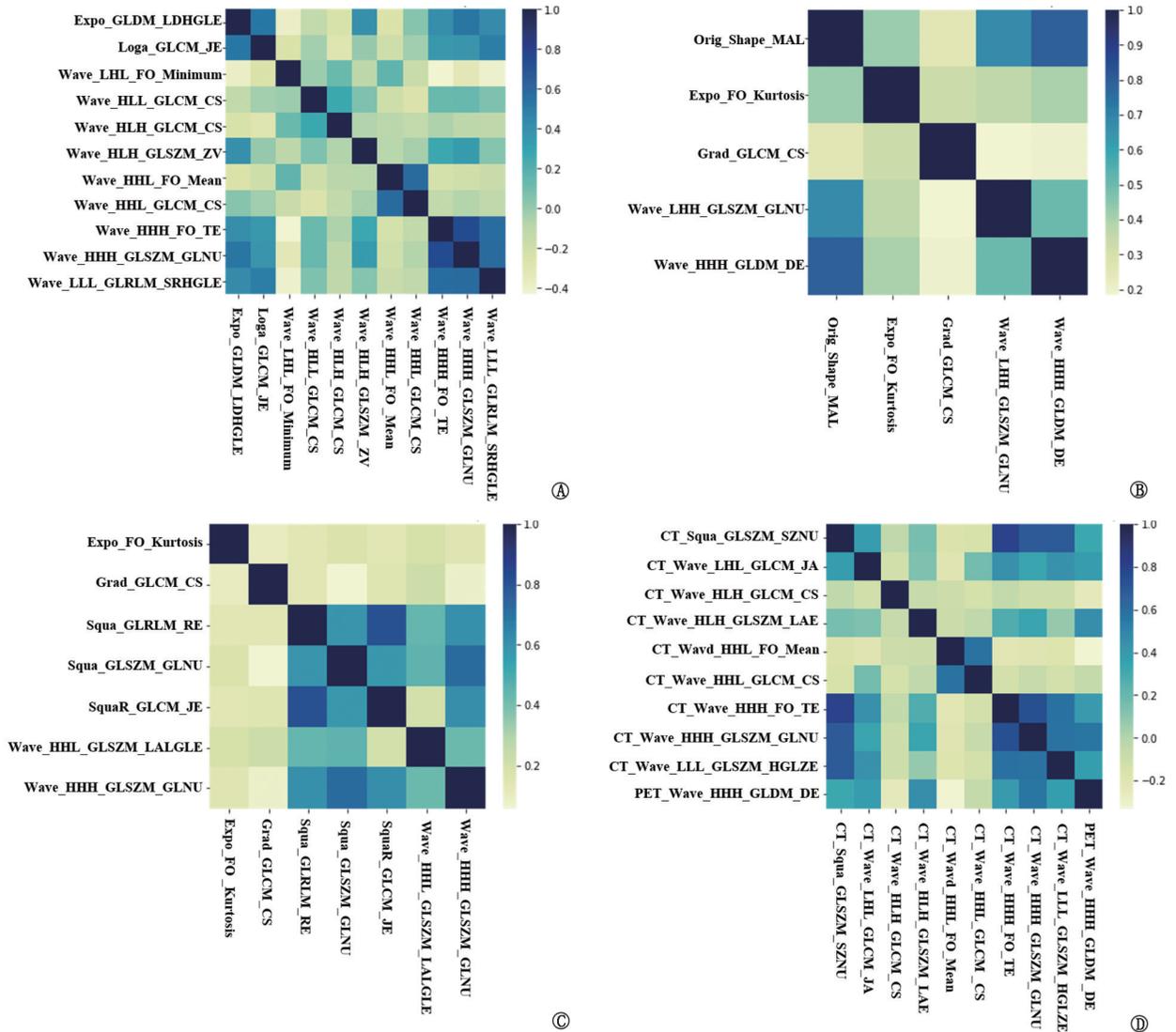


图 1 肺癌患者影像组学特征的相关性热图。A. CT 特征;B. PET 特征;C. PET/CT 平均融合特征;D. PET/CT 拼接融合特征。颜色表示相关性,深色表示正相关,浅色表示负相关,右侧颜色条带数值为相关系数;CS 代表集群阴影,DE 代表相关熵,Expo 代表指数滤波,FO 代表一阶,GLCM 代表灰度共生矩阵,GLDM 代表相关矩阵,GLNU 代表灰度不均匀度,GLRLM 代表灰度游程长度矩阵,GLSZM 代表灰度区域大小矩阵,Grad 代表梯度滤波,H 代表高通,HGLZE 代表高灰度区域优势,JA 代表联合平均,JE 代表联合熵,Kurtosis 代表峰度,L 代表低通,LAE 代表大区域优势,LALGLE 代表大区域低灰度优势,LDHGLE 代表大相关低灰度优势,Loga 代表对数滤波,MAL 代表长轴长度,Mean 代表平均值,Minimum 代表最小值,Orig 代表原始图像,RE 代表运行熵,Shape 代表形状特征,Squa 代表平方滤波,SquaR 代表平方根滤波,SRHGLE 代表短游程高灰度优势,SZNU 代表区域不均匀度,TE 代表总能量,Wave 代表小波滤波,ZV 代表区域方差

19 号外显子缺失突变和 21 号外显子 L858R 突变<sup>[3]</sup>。本研究中 126 例 EGFR 突变型的 LUAD 患者中女性 72 例(57.1%);21 号外显子突变和 19 号外显子突变的患者占突变患者的 92.9%(117/126),与先前文献报道一致。

研究表明,影像组学特征对 EGFR 突变亚型有一定预测价值<sup>[15-17]</sup>。Liu 等<sup>[16]</sup>仅使用了 CT 影像组学特征对 EGFR 突变亚型进行预测,预测效能较有限,本研究使用了<sup>18</sup>F-FDG PET/CT 影像数据。虽然 Yang 等<sup>[15]</sup>和 Liu 等<sup>[17]</sup>使用了<sup>18</sup>F-FDG PET/CT 影

像组学特征,但是本研究纳入了更多的病例与影像组学特征。在影像组学特征的筛选部分,Yang 等<sup>[15]</sup>使用了最小绝对收缩和选择算子(least absolute shrinkage and selection operator, LASSO)回归实现高维数据的降维。单纯使用 LASSO 回归容易造成重要信息的丢失,从而导致训练模型的欠拟合。因此,本研究从多方面多角度对高维组学特征进行分布降维,在一定程度上减少了信息的丢失。另外,使用了 3 种不同的分类模型对 CT、PET 和 2 种融合方式的 PET/CT 影像组学特征进行学习,并根据不

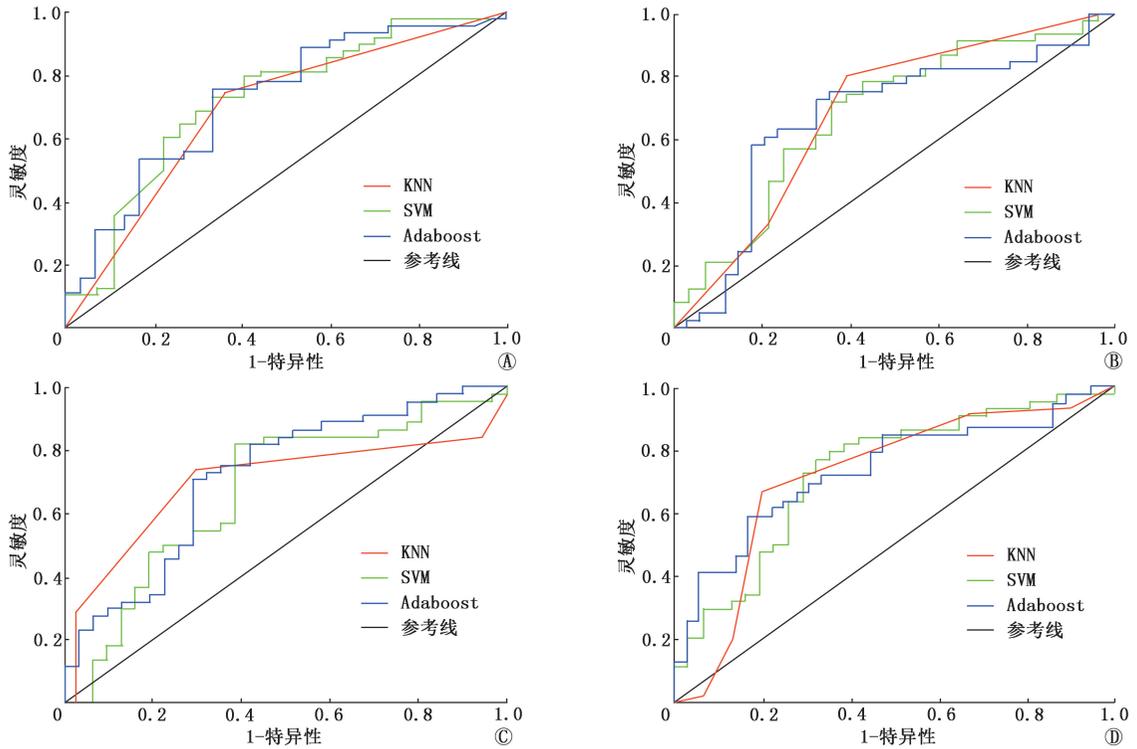


图 2 4 种影像组学特征下各机器学习分类器对肺癌患者 ( $n=229$ ) 表皮生长因子 (EGFR) 突变型和 EGFR 野生型预测的受试者工作特征 (ROC) 曲线。A. CT 特征; B. PET 特征; C. PET/CT 平均融合特征; D. PET/CT 拼接融合特征。KNN 为 K-最近邻, SVM 为支持向量机

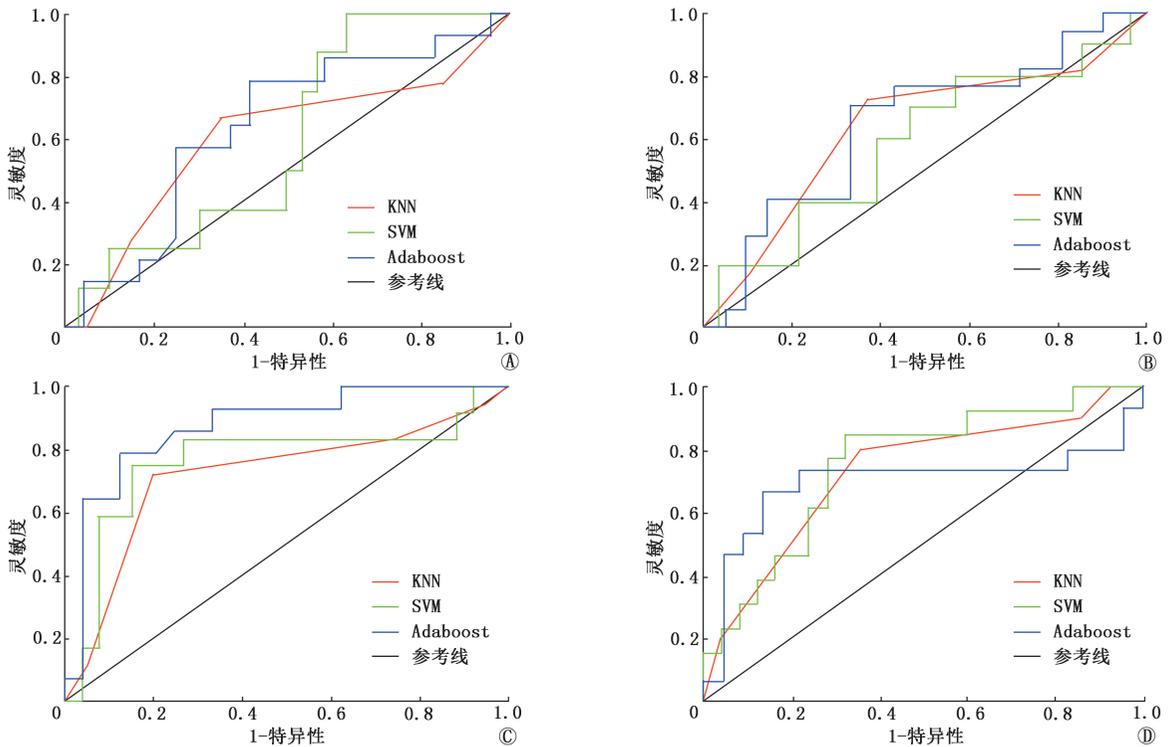


图 3 4 种影像组学特征下各机器学习分类器对 EGFR 突变的肺癌患者 ( $n=117$ ) 亚型预测的 ROC 曲线。A. CT 特征; B. PET 特征; C. PET/CT 平均融合特征; D. PET/CT 拼接融合特征

同模型的预测表现选择了最优的分类模型。

本研究结果显示, KNN、SVM 和 Adaboost 在 229 例 LUAD 患者 EGFR 突变状态的预测中, CT 特征、PET

特征、PET/CT 拼接特征和 PET/CT 平均特征预测的 AUC 范围为 0.67~0.74; KNN、SVM 和 Adaboost 在 117 例 LUAD 患者 EGFR 突变亚型预测中, PET/

CT 平均特征预测的 AUC 分别为 0.72、0.76 和 0.87, PET/CT 平均特征的预测表现要优于 CT 特征和 PET 特征的预测表现(均  $P < 0.05$ ), 但与 PET/CT 拼接特征的预测表现差异无统计学意义 ( $P > 0.05$ )。本研究表明, 机器学习结合  $^{18}\text{F}$ -FDG PET/CT 影像组学特征在 EGFR 突变亚型的预测中有一定价值, 而单独 CT 或 PET 影像组学特征所提供的预测信息相对有限, PET/CT 融合特征可以在单模态的基础上提高预测效能。

本研究尚存在一些不足之处: 首先, 研究为回顾性分析, 病例收集的时间跨度较大, 不同病例之间可能存在批次效应; 其次, PET/CT 特征的融合方式较为简单, 应当选择更多的融合方式进行预测; 最后, 未能纳入更多的病例对研究结果进行多层次和多中心的论证。

**利益冲突** 所有作者均声明不存在利益冲突

### 参 考 文 献

- [1] 李晓琳, 谢鹏, 孟雪, 等. 表皮生长因子受体分子显像与基因突变在非小细胞肺癌中的研究进展[J]. 中华核医学与分子影像杂志, 2012, 32(5): 393-396. DOI: 10.3760/cma.j.issn.2095-2848.2012.05.020.  
Li XL, Xie P, Meng X, et al. The development of EGFR molecular imaging and gene mutation in non-small cell lung cancer [J]. Chin J Nucl Med Mol Imaging, 2012, 32(5): 393-396. DOI: 10.3760/cma.j.issn.2095-2848.2012.05.020.
- [2] Recondo G, Fachinetti F, Olausson KA, et al. Making the first move in EGFR-driven or ALK-driven NSCLC: first-generation or next-generation TKI? [J]. Nat Rev Clin Oncol, 2018, 15(11): 694-708. DOI: 10.1038/s41571-018-0081-4.
- [3] Renaud S, Seitlinger J, Guerrero F, et al. Prognostic value of exon 19 versus 21 EGFR mutations varies according to disease stage in surgically resected non-small cell lung cancer adenocarcinoma [J]. Ann Surg Oncol, 2018, 25(4): 1069-1078. DOI: 10.1245/s10434-018-6347-3.
- [4] Truini A, Starrett JH, Stewart T, et al. The EGFR exon 19 mutant L747-A750>P exhibits distinct sensitivity to tyrosine kinase inhibitors in lung adenocarcinoma [J]. Clin Cancer Res, 2019, 25(21): 6382-6391. DOI: 10.1158/1078-0432.CCR-19-0780.
- [5] 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. DOI: 10.1016/j.ejca.2011.11.036.
- [6] Li X, Yin G, Zhang Y, et al. Predictive power of a radiomic signature based on  $^{18}\text{F}$ -FDG PET/CT images for EGFR mutational status in NSCLC [J]. Front Oncol, 2019, 9: 1062. DOI: 10.3389/fonc.2019.01062.
- [7] 郭虹霞, 任筱璐, 张俊萍.  $^{18}\text{F}$ -FDG PET/CT 对肺腺癌患者 EGFR 突

变的预测价值 [J]. 中华核医学与分子影像杂志, 2020, 40(8): 475-479. DOI: 10.3760/cma.j.cn321828-20200113-00017.

- Guo HX, Ren XL, Zhang JP. Predictive value of  $^{18}\text{F}$ -FDG PET/CT for EGFR mutations in patients with lung adenocarcinoma [J]. Chin J Nucl Med Mol Imaging, 2020, 40(8): 475-479. DOI: 10.3760/cma.j.cn321828-20200113-00017.
- [8] 尹国涛, 徐文贵, 李小凤.  $^{18}\text{F}$ -FDG PET/CT 在非小细胞肺癌 EGFR 突变状态预测中的应用价值 [J]. 中华核医学与分子影像杂志, 2019, 39(10): 623-626. DOI: 10.3760/cma.j.issn.2095-2848.2019.10.012.  
Yin GT, Xu WG, Li XF. Application value of  $^{18}\text{F}$ -FDG PET/CT in predicting EGFR mutation status in non-small cell lung cancer [J]. Chin J Nucl Med Mol Imaging, 2019, 39(10): 623-626. DOI: 10.3760/cma.j.issn.2095-2848.2019.10.012.
- [9] Kikinis R, Pieper SD, Vosburgh KG. 3D slicer: a platform for subject-specific image analysis, visualization, and clinical support [M]// Jolesz F. Intraoperative image-guided therapy. New York: Springer, 2014: 277-289. DOI: 10.1007/978-1-4614-7657-3\_19.
- [10] van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype [J]. Cancer Res, 2017, 77(21): e104-e107. DOI: 10.1158/0008-5472.CAN-17-0339.
- [11] Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with python [C]. Proceedings of the 9th Python in Science Conference, 2010.
- [12] Chen Z, Fillmore CM, Hammerman PS, et al. Non-small-cell lung cancers: a heterogeneous set of diseases [J]. Nat Rev Cancer, 2014, 14(8): 535-546. DOI: 10.1038/nrc3775.
- [13] Zhu JQ, Zhong WZ, Zhang GC, et al. Better survival with EGFR exon 19 than exon 21 mutations in gefitinib-treated non-small cell lung cancer patients is due to differential inhibition of downstream signals [J]. Cancer Lett, 2008, 265(2): 307-317. DOI: 10.1016/j.canlet.2008.02.064.
- [14] Choi YW, Jeon SY, Jeong GS, et al. EGFR exon 19 deletion is associated with favorable overall survival after first-line gefitinib therapy in advanced non-small cell lung cancer patients [J]. Am J Clin Oncol, 2018, 41(4): 385-390. DOI: 10.1097/COC.0000000000000282.
- [15] Yang B, Ji HS, Zhou CS, et al.  $^{18}\text{F}$ -fluorodeoxyglucose positron emission tomography/computed tomography-based radiomic features for prediction of epidermal growth factor receptor mutation status and prognosis in patients with lung adenocarcinoma [J]. Transl Lung Cancer Res, 2020, 9(3): 563-574. DOI: 10.21037/tlcr-19-592.
- [16] Liu G, Xu Z, Ge Y, et al. 3D radiomics predicts EGFR mutation, exon-19 deletion and exon-21 L858R mutation in lung adenocarcinoma [J]. Transl Lung Cancer Res, 2020, 9(4): 1212-1224. DOI: 10.21037/tlcr-20-122.
- [17] Liu Q, Sun D, Li N, et al. Predicting EGFR mutation subtypes in lung adenocarcinoma using  $^{18}\text{F}$ -FDG PET/CT radiomic features [J]. Transl Lung Cancer Res, 2020, 9(3): 549-562. DOI: 10.21037/tlcr.2020.04.17.

(收稿日期: 2020-11-05)