

## 基于T2WI与RS-EPI DWI影像组学特征的自动化机器学习模型预测直肠癌术前T分期的价值\*

文大光<sup>1</sup>, 胡斯娴<sup>1</sup>, 李真林<sup>1</sup>, 邓祥兵<sup>2</sup>, 田川<sup>1</sup>, 李昕<sup>3</sup>, 汪昕蓉<sup>3</sup>, 冷琦<sup>1</sup>, 夏春潮<sup>1△</sup>

1. 四川大学华西医院放射科(成都 610041); 2. 四川大学华西医院胃肠外科(成都 610041); 3. GE医疗(上海 200000)

**【摘要】目的** 探讨基于磁共振T2加权成像(T2 weighted image, T2WI)和分段读出平面回波成像(readout-segmented EPI, RS-EPI)与扩散加权成像(diffusion weighted image, DWI)的影像组学特征,通过开发和验证自动化机器学习模型,预测直肠癌术前病理T分期的价值。**方法** 回顾性分析2016年10月-2018年12月经手术病理结果证实为直肠癌且在术前直肠癌磁共振的患者131例。采用ITK-SNAP软件从T2WI和RS-EPI DWI图像中手动分割出肿瘤区域。使用PyRadiomics包提取出200个特征[100个特征来自于T2WI, 100个特征来自RS-EPI DWI的表观弥散系数(apparent diffusion coefficient, ADC)图]。使用mwmote与neater重采样均衡数据,加入13例T<sub>1-2</sub>期模拟数据。根据3:1的比例将总体数据分割成训练集111例和测试集37例。在训练集上使用Tree-based Pipeline Optimization Tool(TPOT)最优化模型参数并选取最重要的组学特征建模,得到5个互相独立的T分期模型。使用准确率和受试者工作特征(ROC)曲线下面积(area under the curve, AUC)筛选出最优模型。在测试集和原数据集上预测直肠癌T分期。**结果** 自动化机器学习推荐的5个T分期模型,在训练集上的准确率为0.802~0.838,敏感度为0.762~0.825,特异度为0.833~0.896, AUC范围为0.841~0.893, average precision(AP)范围为0.870~0.901。经过对比后,最终选择的模型的敏感度、特异度、AUC,在训练集上为0.810、0.875、0.893,在测试集上为0.810、0.813、0.810,在原始数据集上为0.810、0.830、0.860。**结论** 基于T2WI和RS-EPI DWI的影像组学数据,通过自动化机器学习建立的模型在预测直肠癌T分期上有较高的准确率。

**【关键词】** 直肠癌 影像组学 自动化机器学习 T分期

**Application of Automated Machine Learning Based on Radiomics Features of T2WI and RS-EPI DWI to Predict Preoperative T Staging of Rectal Cancer** WEN Da-guang<sup>1</sup>, HU Si-xian<sup>1</sup>, LI Zhen-lin<sup>1</sup>, DENG Xiang-bing<sup>2</sup>, TIAN Chuan<sup>1</sup>, LI Xin<sup>3</sup>, WANG Xin-rong<sup>3</sup>, LENG Qi<sup>1</sup>, XIA Chun-chao<sup>1△</sup>. 1. Department of Radiology, West China Hospital, Sichuan University, Chengdu 610041, China; 2. Department of Gastrointestinal Surgery, West China Hospital, Sichuan University, Chengdu 610041, China; 3. GE Healthcare, Shanghai 200000, China

△ Corresponding author, E-mail: xiachunchao@wchscu.cn

**【Abstract】Objective** To explore the radiomics features of T2 weighted image (T2WI) and readout-segmented echo-planar imaging (RS-EPI) plus diffusion-weighted imaging (DWI), to develop an automated machine-learning model based on the said radiomics features, and to test the value of this model in predicting preoperative T staging of rectal cancer. **Methods** The study retrospectively reviewed 131 patients who were diagnosed with rectal cancer confirmed by the pathology results of their surgical specimens at West China Hospital of Sichuan University between October, 2017 and December, 2018. In addition, these patients had preoperative rectal MRI. Tumor regions from preoperative MRI were manually segmented by radiologists with the ITK-SNAP software from T2WI and RS-EPI DWI images. PyRadiomics was used to extract 200 features—100 from T2WI and 100 from the apparent diffusion coefficient (ADC) calculated from the RS-EPI DWI. MWMOTE and NEATER were used to resample and balance the dataset, and 13 cases of T<sub>1-2</sub> stage simulation cases were added. The overall dataset was divided into a training set (111 cases) and a test set (37 cases) by a ratio of 3 : 1. Tree-based Pipeline Optimization Tool (TPOT) was applied on the training set to optimize model parameters and to select the most important radiomics features for modeling. Five independent T stage models were developed accordingly. Accuracy and the area under the curve (AUC) of receiver operating characteristic (ROC) were used to pick out the optimal model, which was then applied on the training set and the original dataset to predict the T stage of rectal cancer. **Results** The performance of the the five T staging models recommended by automated machine learning were as follows: The accuracy for the training set ranged from 0.802 to 0.838, sensitivity, from 0.762 to 0.825, specificity, from 0.833 to 0.896, AUC, from 0.841 to 0.893, and average precision (AP) from 0.870 to 0.901. After comparison, an optimal model was picked out, with sensitivity, specificity and AUC for the training set reaching 0.810, 0.875, and 0.893, respectively. The sensitivity, specificity and AUC for the test set were 0.810, 0.813, and 0.810, respectively. The sensitivity, specificity and AUC for the original dataset were 0.810, 0.830, and 0.860, respectively.

\* 四川省科技计划项目(No. 2019YFS0431)和四川大学华西医院学科卓越发展1.3.5工程项目(No. ZYGD18019)资助

△ 通信作者, E-mail: xiachunchao@wchscu.cn

**Conclusion** Based on the radiomics data of T2WI and RS-EPI DWI, the model established by automated machine learning showed a fairly high accuracy in predicting rectal cancer T stage.

**【Key words】** Rectal cancer Radiomics Automated machine learning T stage

结直肠癌是全世界发病率和死亡率较高的癌症之一<sup>[1]</sup>,其死亡率居所有癌症第四位<sup>[2-3]</sup>,其中,近1/3的结直肠癌发生于直肠<sup>[3]</sup>。直肠癌的临床治疗方式主要有手术切除和放化疗治疗<sup>[4]</sup>,而治疗方案的选择主要取决于病变的TNM分级。对于局部进展期直肠癌(T<sub>3,4</sub>期和/或N<sub>1,2</sub>期),与单纯手术切除相比,经新辅助放化疗(NAT)后手术切除,可使局部复发风险降低50%~61%<sup>[5-6]</sup>。然而,药物毒性、尿失禁和性功能障碍也是NAT的严重副作用<sup>[7-9]</sup>。因此,治疗前准确分期并判断进展期患者,能使患者从NAT获益,同时有助于避免早期直肠癌患者因过度放化疗而产生相应的副作用。

磁共振图像(MRI)的诊断是直肠癌的术前临床分期的重要依据<sup>[10]</sup>。然而,有Meta分析指出直肠癌分期诊断的准确率在不同的医生群体中有较大差异<sup>[11]</sup>。影像组学量化分析通过图像背后的规律来预测疾病,可以避免诊断医生因经验不足而导致的错误评估。目前机器学习方法众多,不同的建模方式得到的结果可能完全不同。自动机器学习框架可以计算几百种甚至上千种数学模型,最终筛选出最优模型。本研究探讨基于磁共振T2加权成像(T2 weighted image, T2WI)和分段读出平面回波成像(readout-segmented echo-planar imaging, RS-EPI)与扩散加权成像(diffusion-weighted image, DWI)的表观弥散系数(apparent diffusion coefficient, ADC)影像组学特征,使用自动化机器学习框架对直肠癌的分期进行预测,以期临床术前评估分期提供有效方法。

## 1 资料与方法

### 1.1 研究对象

连续回顾性纳入2016年10月-2018年12月在本院经手术病理证实为直肠癌并且行MRI直肠检查的患者131例。纳入标准:①外科手术前2周内进行磁共振扫描;②肠镜确认病变位于肛门上方15 cm以内;③所有切除的肿瘤都有病理结果;④一致的磁共振机型与标准的参数(如后文1.2章节中所述)。排除标准:①已开始接受NAT;②病理证实非直肠癌;③磁共振序列、参数不一致;④磁共振图像存在伪影;⑤检查前肠道准备不充分,影响观察病变;⑥家族性息肉病。本研究经四川大学华西医院伦理委员会审查通过(2018年审140号)。收集患者临床特征和检查数据,包括:年龄、性别,实验室检测的血清癌胚抗原(CEA)、癌抗原199(CA199)以及术后病理类型。

### 1.2 MRI检查方法

使用西门子Magnetom Skyra 3.0T MR扫描仪,患者仰卧位,头先进,采用18通道相控阵体部线圈及其相匹配的脊柱线圈单元进行成像。T2WI的参数如下:重复时间(TR)/回波时间(TE)2 980 ms/88 ms;采集视野(field of view, FOV):216×216,层厚:3 mm;矩阵大小:160×160。RS-EPI DWI参数如下:TR/TE 5 000 ms/88 ms,层厚4.5 mm,FOV:216×216,矩阵大小:128×128,弥散b值:0和1 000 s/mm<sup>2</sup>,RS次数:5。ADC值为 $(SI_{b=1\ 000}-SI_{b=0})/(1\ 000-0)$ ,即通过b=1 000的弥散图像和b=0的弥散图像计算ADC值,1 000-0代表这两个b值之差,SI(signal intensity)为信号强度。

### 1.3 三维感兴趣区(volume of interest, VOI)勾画和诊断

所有数据随机分成两组。两位经验丰富的放射科医生在双盲法下,使用ITK-SNAP进行勾画。第一组由一人勾画,另一人审核。如果有异议,两人商讨之后确认最终VOI。第二组角色交换。采用逐层勾画模式,勾画的图像经融合得到三维VOI。两位放射科医生共同诊断,参照美国肿瘤协会TNM指南的直肠癌部分的分期原则<sup>[10]</sup>,确定肿瘤的MRI分期。

### 1.4 特征提取与数据分割

利用开源软件PyRadiomics,分别从T2WI和RS-EPI DWI图像上VOI中提取影像组学特征各100个。萃取的特征可分为以下6类:① 18个一阶统计量(first-order features);② 14个基于形状的特征(shape features);③ 22个灰度共生矩阵特征[gray-level co-occurrence matrix (GLCM) features];④ 16个灰度游程矩阵特征[gray level run length matrix (GLRLM) features];⑤ 16个灰度大小区域矩阵特征[gray level size zone matrix (GLSZM) features];⑥ 14个灰度依赖矩阵特征[gray level dependence matrix (GLDM) features]。

归一化将所有数据的特征值控制在(0, 1)区间。数据分布不均(47:84),使用mwmote重采样T<sub>1,2</sub>期数据100例,neater剔除83例,最终数据比(64:84)。将数据随机分成75%的训练集和25%的测试集。

### 1.5 建模和统计学方法

为使特征数量与训练集样本量相当,从而减少过拟合出现的概率,仅在训练集上使用sklearn.feature\_selection的SelectKBest和chi2函数(卡方检验)进行特征筛选。筛

选出前100个差异性较大的特征,作为输入pipeline的特征,纳入自动化机器学习进行建模。

本研究采用的自动机器学习框架Tree-based Pipeline Optimization Tool(TPOT)由宾夕法尼亚大学遗传算法实验室开发<sup>[12]</sup>(<http://epistasislab.github.io/tpot>)。TPOT具体步骤包括特征选择、模型选择、参数优化,我们将流程绘制为图1。在TPOT中,使用Python包DEAP<sup>[13]</sup>来实现遗传学习,生成树形的迭代pipeline,最大限度提高最终的分类精度。在每次迭代过程中,将删除所有pipeline中性能最差的,然后进行下一次迭代。本研究的迭代次数为100次,初始pipeline规模为100(population size)。训练集采用10折交叉验证,得到平均准确率。最终,筛选出平均准确率最高的pipeline并对每个特征的重要性进行排序。

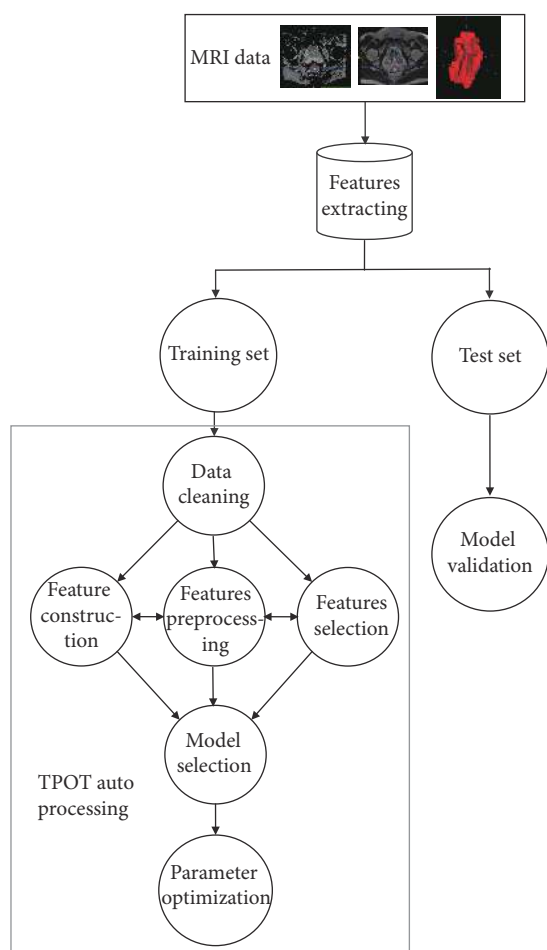


图1 TPOT框架算法流程

Fig 1 TPOT framework algorithm flowchart

本研究使用TPOT自动化机器学习对T分期建模,通过固定训练集和测试集分组来减少实验分组和参数变化带来的随机性,在TPOT参数中输入不同的随机数,最终产生5个模型。在训练集上使用敏感度、特异度、准确率、Hamming loss、F1score、Kappa值、操作者工作特征(ROC)

曲线下面积(AUC)、average precision(AP)来评估不同模型的最终表现。Hamming loss相关标记未出现在预测的标记集合中或无关标记出现在预测的标记集合中(预测结果中,错误数/总数)。该指标取值越小则系统性能越优。在二分类上Hamming loss=1-accuracy。准确率=(真阳性+真阴性)/总样本,而精准率(precision,阳性预测值)=真阳性/(真阳性+假阳性),召回率(recall,敏感度)=真阳性/(真阳性+假阴性),二者计算可得到F1-score,F1-score表示precision与recall的权衡,防止这两个参数差异较大。越接近1,说明precision越接近recall,比较均衡,但是并不能反映模型最佳。Kappa值反映的是机器学习与真实结果的一致性,值越大,机器学习的结果越接近真实结果。Kappa值0.0~0.20反映极低的一致性;0.21~0.40反映一般的一致性;0.41~0.60反映中等的一致性;0.61~0.80反映高度的一致性;0.81~1反映几乎完全一致。这些都可以用来反映多分类机器学习效能的参数,同样也可以反映二分类模型结果的好坏。AP类似AUC,是计算机领域某些评价模型性能的办法,本文以AUC及准确率作为主指标评价模型。用AUC和准确率筛选出最优的T分期模型,得到6个最重要的特征及权重值并分析特征类型。在测试集上使用敏感度、特异度、准确率、AUC、AP来评估模型的诊断效能。计算出最优模型在原数据集上的敏感度、特异度、AUC值。

满足正态分布的计量资料以 $\bar{x} \pm s$ 表示,采用Mann-Whitney检验训练集、测试集上不同分期患者的特征值的差异,使用AUC评价单独使用特征预测T分期的诊断效能, $P < 0.05$ 为差异有统计学意义。使用散点图显示最优3个特征的数据集在采样前后分布情况。

## 2 结果

### 2.1 临床资料

本研究纳入131例患者,其中男性90例,女性41例,年龄中位数为62岁,分化程度:G<sub>1</sub>期5例,G<sub>2</sub>期109例,G<sub>3</sub>期17例。环周切缘(circumferential resection margin,CRM)阳性5例,阴性126例。对于T分期过采样后,训练集中,T<sub>1,2</sub>期48例,T<sub>3,4</sub>期63例;测试集中,T<sub>1,2</sub>期16例,T<sub>3,4</sub>期21例。原始数据T<sub>1</sub>期10例,T<sub>2</sub>期37例,T<sub>3</sub>期70例,T<sub>4</sub>期14例。N<sub>0</sub>期86例,N<sub>1</sub>期45例。脉管浸润阳性22例,阴性109例。神经侵犯阳性47例,阴性84例。CEA:(6.86±10.18) ng/mL,CA-199:(17.37±23.37) ng/mL。

### 2.2 模型参数

在将筛选后的特征通过自动机器学习建模后,最终产生5个独立的TPOT模型来预测T分期。各个模型的建模参数如表1所示。

表 1 T分期模型参数  
Table 1 T staging model parameters

Model	Modeling classifier	Parameters
1	DecisionTreeClassifier, MultinomialNB	Make_union (make_pipeline (StackingEstimator (estimator= DecisionTreeClassifier (criterion= "entropy", max_depth= 5, min_samples_leaf= 8, min_samples_split= 8)), Binarizer (threshold= 0.6000000000000001)), make_union (FunctionTransformer (copy), FunctionTransformer (copy))), MultinomialNB (alpha= 1.0, fit_prior= True)
2	MultinomialNB	MinMaxScaler (), StackingEstimator (estimator= MultinomialNB (alpha= 10.0, fit_prior= True)), Binarizer (threshold= 0.6000000000000001), StackingEstimator (estimator= MultinomialNB (alpha= 10.0, fit_prior= False)), MultinomialNB (alpha= 1.0, fit_prior= True)
3	MultinomialNB, DecisionTreeClassifier	StackingEstimator (estimator= MultinomialNB (alpha= 10.0, fit_prior= True)), Binarizer (threshold= 0.6000000000000001), StackingEstimator (estimator= DecisionTreeClassifier (criterion= "entropy", max_depth= 5, min_samples_leaf= 17, min_samples_split= 3)), MultinomialNB (alpha= 1.0, fit_prior= True)
4	MultinomialNB, DecisionTreeClassifier	StackingEstimator (estimator= MultinomialNB (alpha= 100.0, fit_prior= False)), Normalizer (norm= "max"), StackingEstimator (estimator= DecisionTreeClassifier (criterion= "entropy", max_depth= 2, min_samples_leaf= 16, min_samples_split= 20)), MultinomialNB (alpha= 1.0, fit_prior= True)
5	MultinomialNB	Binarizer (threshold= 0.6000000000000001), RobustScaler (), Binarizer (threshold= 0.65), MultinomialNB (alpha= 1.0, fit_prior= True)

2.3 模型预测结果

见表2、图2。5个模型在训练集的概率阈值均为0.5, 准确率为0.802 ~ 0.838, 敏感度为0.762 ~ 0.825, 特异度为0.833 ~ 0.896, AUC为0.840 ~ 0.893。见图3、表3。对比训练集所有模型(表2)的数据, model1为最优模型, 准确率最高, AUC值最高, 分别为0.838、0.893, 其Hamming loss最低, 为0.162, F1 score最大, 为0.850, kappa score最

大, 为0.674, AP值最大, 为0.901。model1在训练集上的错误率更低, 预测的标签和真实的病例结果一致性(kappa score)较高, 而其F1 score值大表示精准率(precision)和召回率(recall)非常接近。在测试集模型数据(表3)中, model1的敏感度最高, 特异度低于model3、4、5, 准确率最高, AUC排第二, 低于model4, AP值也低于model3、4、5。在测试集上的AP如图4所示, 所有曲线50%以上的点

表 2 T分期模型训练集上的评估指标  
Table 2 T staging model evaluation in the training set

Model	Threshold	Sensitivity	Specificity	Accuracy	Hamming loss	F1 score	Kappa score	AUC	AP
1*	0.5	0.809	0.875	0.838	0.162	0.850	0.674	0.893	0.901
2	0.5	0.809	0.833	0.819	0.180	0.836	0.636	0.841	0.880
3	0.5	0.778	0.833	0.802	0.198	0.817	0.602	0.846	0.885
4	0.5	0.825	0.833	0.829	0.171	0.845	0.654	0.842	0.869
5	0.5	0.762	0.896	0.819	0.180	0.827	0.642	0.864	0.892

\*Best performance of all models in training set; AP: Average precision; AUC: Area under the curve; Hamming loss: The fraction of labels that are incorrectly predicted; F1 score=2×(precision×recall)/(precision+recall); Kappa score: A score that expresses the level of agreement between two annotators on a classification problem.

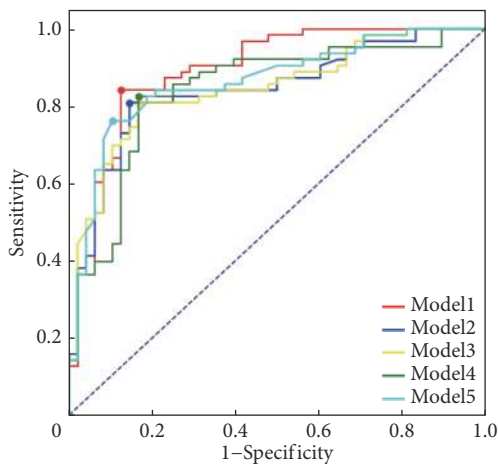


图 2 T分期模型在训练集的ROC曲线

Fig 2 ROC curves of T staging models applied on the training set

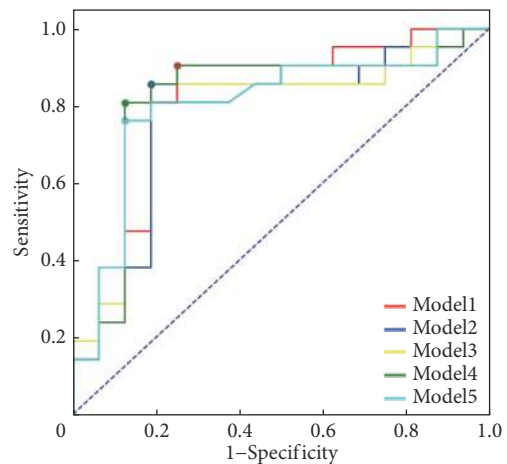


图 3 T分期模型在测试集的ROC曲线

Fig 3 ROC curves of T staging models applied on the test set

表 3 T分期模型在测试集上的评估指标  
Table 3 T staging model evaluation with the test set

Model	Threshold	Sensitivity	Specificity	Accuracy	AUC	AP
1*	0.5	0.809	0.812	0.811	0.809	0.823
2	0.5	0.667	0.812	0.729	0.783	0.821
3	0.5	0.714	0.875	0.784	0.803	0.843
4	0.5	0.619	0.875	0.729	0.815	0.837
5	0.5	0.571	0.875	0.703	0.802	0.837

\*Best performance of all models in training set; AP: Average precision; AUC: Area under the curve.

precision>0.7, 加权后的均值(AP)model3最大, 为0.843。最优模型(model1)在原数据集上的敏感度为0.810, 特异度为0.830, AUC为0.864, 如图5所示。

2.4 最优模型的特征

打印出model1模型6个特征名称及系数, 如图6所示, Original\_shape\_MinorAxisLength. 1是VOI 3D椭球第二长

的轴线, 此值越大, 代表肿瘤侵犯的范围越宽; Original\_shape\_Maximum2D Diameter Row及Original\_shape\_Maximum2D Diameter Row. 1指矢状2D平面上肿瘤表面顶点内最大距离, 肿瘤侵犯越深、累及肠道范围越大, 此值越大; Original\_glcm\_Contrast是在GLCM变换后计算的

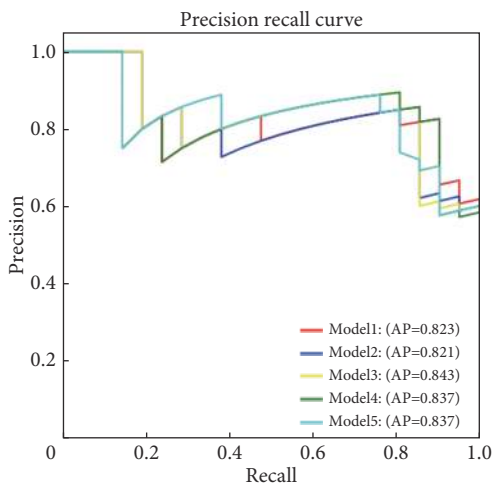


图 4 T分期模型在测试集的精确率-召回率曲线

Fig 4 The precision-recall curves of T staging models applied on the test set

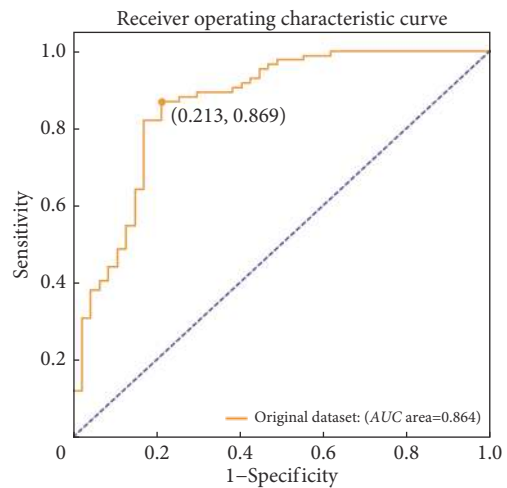


图 5 最优T分期模型 ( model1 ) 在原数据集的ROC曲线

Fig 5 The ROC curve of the optimal T staging model (model1) applied on the original dataset

The information in the parenthesis shows the specificity and sensitivity of the optimal cutoff point.

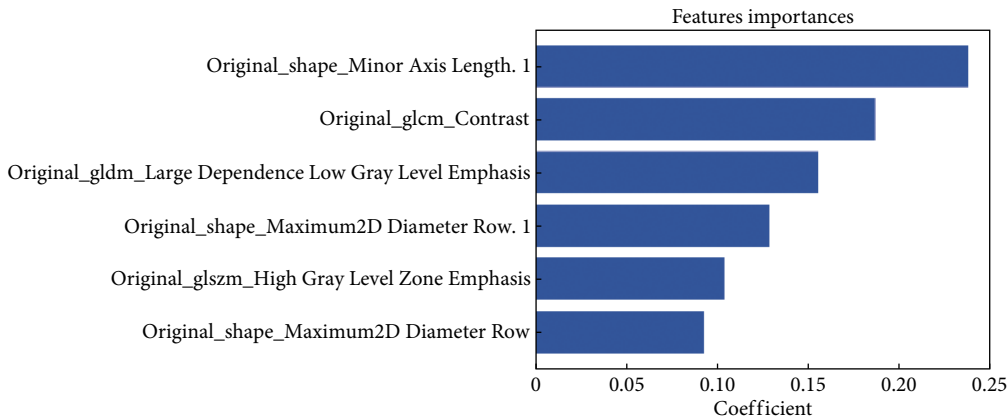


图 6 T分期model1模型特征的系数

Fig 6 Coefficient of T staging model1 features

Feature name with suffix ".1" were from ADC, and other features were from the T2WI.

特征, 值越大, 代表纹理沟纹越深, 视觉效果越清晰; Original\_gldm\_Large Dependence Low Gray Level Emphasis灰度依赖矩阵低灰度依赖程度, 值越大, 表示在低灰度范围内, 纹理越均匀; Original\_glszm\_High Gray Level Zone Emphasis高灰度区域分布程度, 是测量较高灰度值的分布, 较高的值表示较高的灰度值占图像中的大小区域的比例更大。这6个特征中, 2个特征来自于ADC图, 4个来自于T2WI图; 有3个形状类特征, 3个矩阵变换特征; ADC图像的特征比重大于T2WI的特征比重。表4中, Original\_shape\_Minor Axis Length.1的重要性系数排第一, 系数值大于0.2, Original\_glcm\_Contrast和 Original\_gldm\_Large DependenceLow Gray Level

Emphasis大于0.15。单独使用此特征预测直肠癌T分期在训练集测试集AUC均大于0.7, 不同T分期此特征的值差异均有统计学意义( $P < 0.05$ )。Original\_glcm\_Contrast、original\_shape\_Maximum2D DiameterRow.1和Original\_shape\_Maximum2D Diamete rRow在训练集不同T分期均值差异有统计学意义( $P < 0.05$ ), 在测试集上差异无统计学意义( $P > 0.05$ )。单独使用Original\_glszm\_High Gray Level Zone Emphasis或Original\_gldm\_LargeDependenceLow Gray Level Emphasis在训练集与测试集上AUC未超过0.6, 说明缺乏稳定性, 不同T分期此特征的值差异无统计学意义( $P = 0.892$ )。在图7中, 过采样后数据中, 重采样点均未出现在原数据集点外偏离数据集。

表 4 T分期model1前6个最重要特征在数据集上的分布  
Table 4 The distribution of the top 6 important features of model1

Dataset	Index	Top 6 featrues					
		1	2	3	4	5	6
		Original_shape_MinorAxis Length.1	Original_glcm_Contrast	Original_gldm_LargeDependence LowGrayLevelEmphasis	Original_shape_Maximum2D DiameterRow.1	Original_glszm_HighGrayLevel ZoneEmphasis	Original_shape_Maximum2D DiameterRow
Training set	Stage T <sub>1-2</sub>	25.11±8.25	22.17±28.07	1.19±1.30	39.14±14.86	200.45±264.37	37.79±14.66
	Stage T <sub>3-4</sub>	31.18±8.73	12.31±12.16	2.39±3.42	51.39±21.61	140.12±124.86	51.03±20.71
	P*	0.000	0.010	0.335	0.001	0.209	0.000
	AUC	0.739	0.644	0.554	0.677	0.570	0.693
Test set	Stage T <sub>1-2</sub>	22.98±5.98	17.22±14.01	2.45±3.47	34.98±16.06	129.72±87.07	32.11±14.80
	Stage T <sub>3-4</sub>	29.06±8.83	12.34±10.85	3.39±6.35	45.99±18.22	129.72±104.79	45.67±21.29
	P*	0.037	0.229	0.751	0.063	0.892	0.078
	AUC	0.702	0.619	0.533	0.682	0.515	0.673
Original dataset	Stage T <sub>1-2</sub>	25.26±8.38	22.82±28.05	1.38±2.32	40.76±15.59	208.06±265.76	39.32±14.55
	Stage T <sub>3-4</sub>	30.65±8.75	12.32±11.79	2.64±4.32	50.04±20.84	137.52±119.64	49.70±20.87
	P*	0.000	0.003	0.197	0.014	0.063	0.004
	AUC	0.710	0.657	0.568	0.630	0.598	0.654

\*With no multiple testing correction.

### 3 讨论

本研究使用基于自动化机器学习的影像组学分析来预测术前直肠癌T分期。以往研究表明, 高分辨率磁共振T2WI成像在直肠癌的临床分期和指导治疗中起到重要的作用, 但在T2WI上无法准确区分肠壁纤维化反应和肿瘤浸润, 可能导致肿瘤T分期诊断错误而影响治疗决策<sup>[14-15]</sup>。DWI主要反映了水分子在活体组织中的微观扩散运动<sup>[15]</sup>。与正常组织和炎性组织相比, 直肠癌细胞密度更高, 组织间质空间更小, 游离水分子在肿瘤组织中的扩散更少, DWI上突出病变及周围组织的差异, 有利于鉴别T<sub>2</sub>期和T<sub>3</sub>期的肠壁纤维化反应和肿瘤浸润。多项研究表明, 通过DWI图计算出的ADC值在不同分期之间存在显著差异, 并认为ADC值有可能成为肿瘤侵袭性的影像学标志

物<sup>[16-17]</sup>。常规DWI采用SS-EPI序列, 有较大的图像畸变, 会严重影响特征的提取。而本研究所采用的RS-EPI序列, 在以往研究表中被证实, 在图像质量、几何畸变和组织差异上明显优于SS-EPI DWI<sup>[18]</sup>。本研究结合T2WI和RS-EPI DWI数据, 得到更丰富的影像组学特征数据, 从而建立更加准确的机器学习模型。

本研究使用T2WI和RS-EPI DWI图像特征数据建立5个预测模型, 使用准确率和AUC筛选出最优的模型, 在测试集上预测T分期的敏感度、特异度、AUC分别为0.810、0.813、0.810。5个预测模型是TPOT选择不同随机数计算得到的。5个模型使用的训练集虽然一致, 但是因为数据量较小, 在分割成更小的交叉验证的数据集时, 可能会存在偏移, 此时在某一个交叉验证数据集上对比得到的最优模型, 在整个训练集上不一定表现最佳。在本

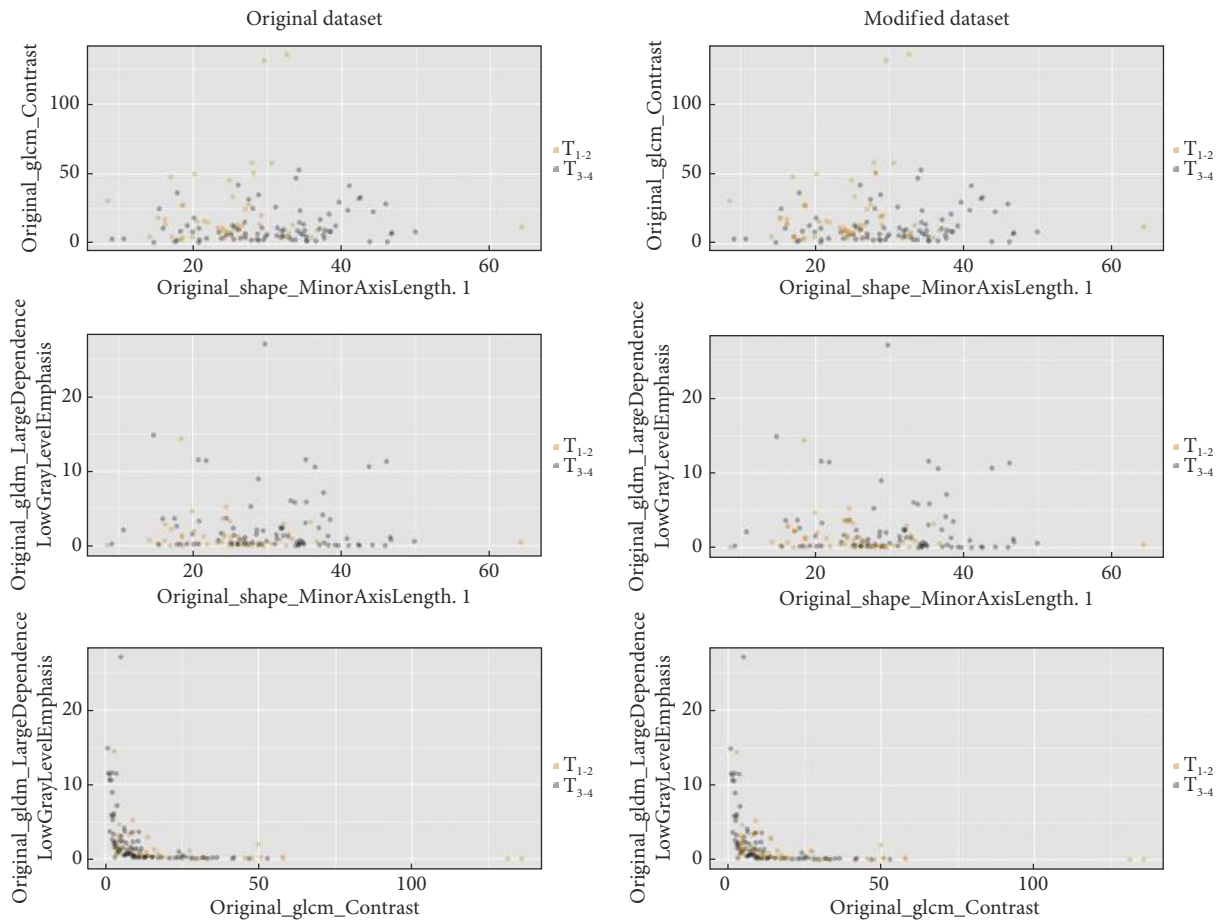


图 7 过采样修改后的数据在最重要的3个特征上的分布

Fig 7 The distribution of the oversampled modified dataset on the three most important features

The gray dots indicate  $T_{1,2}$  patients, while the yellow indicates  $T_{3,4}$  patients.

研究结果中,这5个模型在训练集上的准确率差异不大,采用建模方式以朴素贝叶斯方法居多,差异主要来自于不同的交叉验证集分割。

很多直肠癌是一个弧形的肿块,大部分时候3D椭球最长的径线是弧形平面上的长轴,而这个方向仅仅代表肿瘤包绕直肠壁的情况。第二轴线很多时候是凸向肠壁,本研究也可以看到,图6中, $T_{3,4}$ 期的肿瘤第二轴线值(Original\_shape\_MinorAxisLength. 1)更大,也提示 $T_{3,4}$ 期的肿瘤在肠壁方向的深度更大。表4还显示, $T_{3,4}$ 期肿瘤的original\_glcm\_Contrast值比 $T_{1,2}$ 期值小,这可能提示 $T_{3,4}$ 期肿瘤的浸润比炎性浸润的纹理对比更为强烈。

在同类研究SUN等<sup>[19]</sup>的论文中,以柱状图特征VAR(histogram variance)和分形特征(FRACTAL\_DIMENSION)为主要特征。MA等<sup>[20]</sup>选择original\_shape\_Size以及多种小波变换后的特征建模。本研究未纳入小波变换特征。因为其特征数(1029)远超样本量(152人),且选用了大量的数学变换特征,结果的稳定性还有待验证。SUN的模型对T分期有较好的预测能力(测试集AUC为0.852),这

与本研究的结果大致相当。

此外,和MA和SUN的研究不同,本研究使用自动化机器学习框架,采用更多的建模模型来减小不同的建模技术可能对结果造成的干扰。SUN等<sup>[19]</sup>直接采用LASSO的算法,而MA等<sup>[20]</sup>对比了5种建模技术,最终选取了最佳的模型。多模型任务的复杂程度超越非机器学习专家研究者的能力,自动化机器学习可以自动确定某种优化策略下表现最佳的pipeline来解决多模型重建时的困难。TPOT作为一个开源的自动化学习pipeline,可以智能地探索数千种建模方式(自动化机器学习中最烦琐的部分),从而找到最适合当前数据情况的算法及其参数<sup>[21]</sup>,提高模型的准确率。

本研究存在一定的局限性。①纳入的患者在同一台MRI设备和采集,所得结果在不同设备采集图像上的泛化可能受限。②样本量( $n=131$ )相对较小,且为单中心数据,可能会存在一定的偏差。

基于T2WI和RS-EPI DWI的影像组学数据,通过自动化机器学习建立的模型在预测直肠癌T分期上有较高的

准确率, 可为临床预测直肠癌病理分型以及决策临床治疗等提供依据。

\* \* \*

**利益冲突** 所有作者均声明不存在利益冲突

### 参 考 文 献

- [1] BOYLE P, FERLAY J. Mortality and survival in breast and colorectal cancer. *Nat Clin Pract Oncol*, 2005, 2(9): 424-425.
- [2] PARKIN D, BRAY F, FERLAY J. Global Cancer Statistics, 2002. *CA Cancer J Clin*, 2005, 55(2): 74-108.
- [3] FERLAY J, SHIN H, BRAY F, *et al*. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer*, 2010, 127(12): 2893-2917.
- [4] GLYNNE-JONES R, WYRWICZ L, TIRET E, *et al*. Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*, 2017, 28(Suppl 4): iv22-iv40.
- [5] SEBAG-MONTEFIORE D, STEPHENS R J, STEELE R, *et al*. Preoperative radiotherapy versus selective postoperative chemoradiotherapy in patients with rectal cancer (MRC CR07 and NCIC-CTG C016): A multicentre, randomised trial. *Lancet*, 2009, 373(9666): 811-820.
- [6] VAN GIJN W, MARIJNEN C A, NAGTEGAAL I D, *et al*. Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer: 12-year follow-up of the multicentre, randomised controlled TME trial. *Lancet Oncol*, 2011, 12(6): 575-582.
- [7] CARLOS F, PERICAY C, APARICIO J, *et al*. Phase II, randomized study of concomitant chemoradiotherapy followed by surgery and adjuvant capecitabine plus oxaliplatin (CAPOX) compared with induction CAPOX followed by concomitant chemoradiotherapy and surgery in magnetic resonance imaging-defined, locally advanced rectal cancer: GrupoCáncer de Recto 3 Study. *J Clin Oncol*, 2010, 28(5): 859-865.
- [8] BIRGISSON H, PÅHLMAN L, GUNNARSSON U, *et al*. Adverse effects of preoperative radiation therapy for rectal cancer: Long-term follow-up of the Swedish Rectal Cancer Trial. *J Clin Oncol*, 2005, 23(34): 8697-705.
- [9] STEPHENS R J, THOMPSON L C, QUIRKE P, *et al*. Impact of short-course preoperative radiotherapy for rectal cancer on patients' quality of life: Data from the Medical Research Council CR07/National Cancer Institute of Canada Clinical Trials Group C016 randomized clinical trial. *J Clin Oncol*, 2010, 28(27): 4233-4239.
- [10] BENSON A B, VENOOK A P, AL-HAWARY M M, *et al*. NCCN Guidelines insights: colon cancer, version 2. 2018. *J Natl Compr Canc Netw*, 2018, 16(4): 359-369.
- [11] AL-SUKHNI E, MILOT L, FRUITMAN M, *et al*. Diagnostic accuracy of MRI for assessment of T category, lymph node metastases, and circumferential resection margin involvement in patients with rectal cancer: A systematic review and meta-analysis. *Ann Surg Oncol*, 2012, 19(7): 2212-2223.
- [12] OLSON R S, BARTLEY N, URBANOWICZ R J, *et al*. Evaluation of a tree-based pipeline optimization tool for automating data science//GECCO'16: Proceedings of Genetic and Evolutionary Computation Conference, 2016: 485-492. <https://doi.org/10.1145/2908812.2908918>.
- [13] FORTIN F A, RAINVILLE F M D, GARDNER M A, *et al*. DEAP: Evolutionary algorithms made easy. *J Mach Learn Res*, 2012, 13(7): 2171-2175.
- [14] KEANE C, YOUNG M. Accuracy of magnetic resonance imaging for preoperative staging of rectal cancer. *ANZ J Surg*, 2014, 84(10): 758-762.
- [15] BEETS-TAN R G, BEETS G L, VLIEGEN R F, *et al*. Accuracy of magnetic resonance imaging in prediction of tumour-free resection margin in rectal cancer surgery. *Lancet*, 2001, 357(9255): 497-504.
- [16] YAMASHITA Y, TANG Y, TAKAHASHI M. Ultrafast MR imaging of the abdomen: Echo planar imaging and diffusion-weighted imaging. *J Magn Reson Imaging*, 1998, 8(2): 367-374.
- [17] CURVO-SEMEDO L, LAMBREGTS D M J, MAAS M, *et al*. Diffusion-weighted MRI in rectal cancer: Apparent diffusion coefficient as a potential noninvasive marker of tumor aggressiveness. *J Magn Reson Imaging*, 2012, 35(6): 1365-1371.
- [18] SUN Y, TONG T, CAI S, *et al*. Apparent diffusion coefficient (ADC) value: A potential imaging biomarker that reflects the biological features of rectal cancer. *PLoS One*, 2014, 9(10): e109371[2020-09-12]. <https://doi.org/10.1371/journal.pone.0109371>.
- [19] XIA C C, LIU X, PENG W L, *et al*. Readout-segmented echo-planar imaging improves the image quality of diffusion-weighted MR imaging in rectal cancer: Comparison with single-shot echo-planar diffusion-weighted sequences. *Eur J Radiol*, 2016, 85(10): 1818-1823.
- [20] SUN Y, HU P, WANG J, *et al*. Radiomic features of pretreatment MRI could identify T stage in patients with rectal cancer: Preliminary findings. *J Magn Reson Imaging*, 2018, 48(3): 615-621.
- [21] MA X L, SHEN F, JIA Y, *et al*. MRI-based radiomics of rectal cancer: Preoperative assessment of the pathological features. *BMC Med Imaging*, 2019, 19(1): 86.

(2020-10-13收稿, 2021-04-16修回)

编辑 吕熙