

勾画方法对¹⁸F-FDG PET/CT 影像组学预测胰腺导管腺癌病理分化程度的影响

郝志鑫¹ 刘垒² 邢海群¹ 朱文佳¹ 张辉³ 霍力¹

¹中国医学科学院、北京协和医学院北京协和医院核医学科、核医学分子靶向诊疗北京市重点实验室 100730; ²赛诺联合医疗科技(北京)有限公司 100192; ³清华大学医学院生物医学工程系, 北京 100084

通信作者: 霍力, Email: huoli@pumch.cn

【摘要】 目的 探讨在¹⁸F-脱氧葡萄糖(FDG) PET/CT 图像中采用不同方法勾画胰腺导管腺癌(PDAC)肿瘤区域对使用影像组学特征预测病理分化程度的影响。方法 回顾性分析 2010 年 9 月至 2016 年 1 月间于北京协和医院经病理证实的 72 例 PDAC 患者(男 46 例、女 26 例, 年龄: 25~87 岁)的术前¹⁸F-FDG PET/CT 图像及病理资料。根据 PDAC 病理分化程度将患者分为高分化和非高分化组。入组患者按 3:1 的比例随机划分至训练集和验证集。所有病例由 2 位医师手动勾画感兴趣区(ROI; 记为 ROI_M1 和 ROI_M2), 再分别基于标准摄取值(SUV)梯度(记为 ROI_G)和 40%最大 SUV(SUV_{max})阈值(记为 ROI_S)半自动勾画 ROI。计算并比较 4 种勾画结果的体积、戴斯相似性系数(DSC)。从 PET/CT 原始和预处理图像中提取形状、一阶、纹理等特征, 并以组间相关系数(ICC)评估每个特征在不同勾画结果间的一致性。使用 Kruskal-Wallis 秩和检验、两独立样本 *t* 检验或 *z* 检验分析数据。采用受试者工作特征曲线下面积评估模型准确性, 并通过交叉验证评估模型泛化能力。结果 训练集共 55 例患者(高分化 14 例, 非高分化 41 例); 验证集共 17 例患者(高分化 4 例, 非高分化 13 例)。20 个特征组中共筛选出 44 个对 PDAC 分化程度有预测价值的特征。ROI_M1、ROI_M2、ROI_G 和 ROI_S 勾画的轮廓体积分别为 10.29(4.01, 19.43)、9.34(4.26, 17.27)、11.86(5.52, 19.74) 和 15.08(9.62, 27.44) cm³, 差异有统计学意义($H=18.641, P<0.05$)。手动勾画(ROI_M1 和 ROI_M2)的肿瘤轮廓重合度和特征一致性均较好[DSC=0.86(0.76, 0.90); ICC=0.86(0.74, 0.94)]。与手动勾画结果比较, ROI_G 的轮廓重合度和特征一致性较好[DSC:0.86(0.75, 0.91)、0.91(0.85, 0.96); ICC: 0.87(0.72, 0.94)、0.94(0.88, 0.98)]; ROI_M1 与 ROI_G 的预测模型准确性和泛化能力差异无统计学意义($z=1.052, t=0.712$, 均 $P>0.05$); ROI_M2 的预测模型准确性优于 ROI_G($z=3.031, P=0.002$), 但泛化能力不足($t=3.086, P=0.012$)。结论 基于手动勾画构建的预测模型准确性较高, 但模型性能不稳定; 基于梯度的半自动勾画可以达到与手动勾画相似的准确性, 且模型泛化能力更强。

【关键词】 癌, 胰腺管; 图像处理, 计算机辅助; 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

基金项目: 国家自然科学基金(82071967); 中国医学科学院医学与健康科技创新工程(CAMS-2018-I2M-3-001); 清华大学自主科研计划(20191080605)

DOI: 10.3760/cma.j.cn321828-20210511-00160

Impact of segmentation methods on pathological grade prediction in pancreatic ductal adenocarcinoma based on ¹⁸F-FDG PET/CT radiomics

Hao Zhixin¹, Liu Lei², Xing Haiqun¹, Zhu Wenjia¹, Zhang Hui³, Huo Li¹

¹Department of Nuclear Medicine, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences; Beijing Key Laboratory of Molecular Targeted Diagnosis and Therapy in Nuclear Medicine, Beijing 100730, China; ²Sinounion Healthcare Inc., Beijing 100192, China; ³Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing 100084, China

Corresponding author: Huo Li, Email: huoli@pumch.cn

【Abstract】 Objective To investigate the segmentation methods of pancreatic ductal adenocarcinoma (PDAC) tumor regions in ¹⁸F-fluorodeoxyglucose (FDG) PET/CT images, as well as their impact on radiomic features-based pathological grade prediction. **Methods** A total of 72 patients (46 males, 26 females, age range: 25–87 years) with pathologically confirmed PDAC and a preoperative ¹⁸F-FDG PET/CT scan in Peking Union Medical College Hospital between September 2010 and January 2016 were enrolled

retrospectively. The cohort of patients was classified as well differentiated group and non-well differentiated group based on the pathological grade of PDAC, and patients were divided into training set and validation set in the ratio of 3:1 randomly. Two physicians performed manual contours in the tumor region (referred as region of interest (ROI)_M1 and ROI_M2) and semi-automatic ROIs based on standardized uptake value (SUV) gradient edge search (referred as ROI_G) and 40% threshold applied to the maximum SUV (SUV_{max} ; referred as ROI_S) were drawn. The four types of segmentation results were compared in terms of volume and Dice similarity coefficient (DSC). Shape, first-order, and texture features were extracted from PET/CT original and preprocessed images, and the interclass correlation coefficient (ICC) was used to assess each feature's consistency across all segmentations. Kruskal-Wallis rank sum test, independent-sample t test or z test were used to analyze the data. The area under the receiver operating characteristic curve was used to assess model accuracy, and cross validation was used to assess generalization ability. **Results** There were 55 patients in the training set (14 well differentiated cases and 41 non-well differentiated cases) and 17 patients in the validation set (4 well differentiated cases and 13 non-well differentiated cases). A total of 44 selected features were predictive of the pathological grade of PDAC among 20 feature groups. There was significant difference among the volumes of ROI_M1, ROI_M2, ROI_G and ROI_S (10.29(4.01,19.43), 9.34(4.26,17.27), 11.86(5.52,19.74) and 15.08(9.62,27.44) cm^3 ; $H=18.641$, $P<0.05$). The degree of contour coincidence and feature consistency between ROI_M1 and ROI_M2 were both higher (DSC=0.86(0.76,0.90), ICC=0.86(0.74,0.94)). Compared to manual contours, the degree of contour coincidence and feature consistency of ROI_G (DSC: 0.86(0.75,0.91), 0.91(0.85,0.96); ICC: 0.87(0.72,0.94), 0.94(0.88,0.98)) were better. There was no statistically significant difference in model accuracy or generalization ability between ROI_M1 and ROI_G ($z=1.052$, $t=0.712$, both $P>0.05$). The accuracy of ROI_M2 was better than ROI_G ($z=3.031$, $P=0.002$), but the generalization ability of ROI_M2 was insufficient ($t=3.086$, $P=0.012$). **Conclusions** Although the manual contour prediction models are highly accurate, their performance are unstable. Semi-automatic contouring based on gradient can achieve comparable accuracy to manual contouring, and the model's generalization ability is stronger.

【Key words】 Carcinoma, pancreatic ductal; Image processing, computer-assisted; Positron-emission tomography; Tomography, X-ray computed; Deoxyglucose; Forecasting

Fund program: National Natural Science Foundation of China (82071967); CAMS Innovation Fund for Medical Sciences (CAMS-2018-I2M-3-001); Tsinghua University Initiative Scientific Research Program (20191080605)

DOI:10.3760/cma.j.cn321828-20210511-00160

胰腺导管腺癌(pancreatic ductal adenocarcinoma, PDAC)恶性程度高,预后很差,10年生存率仅为3.9%^[1]。PDAC病理低分化是预后的危险因素^[2]。通过手术或穿刺获取肿瘤组织进行病理检查为有创操作,且成本较高。有研究表明,¹⁸F-脱氧葡萄糖(fluorodeoxyglucose, FDG) PET/CT影像组学可于术前对PDAC的病理分化程度进行无创评估^[3]。

影像组学可从医学图像中高通量地提取特征,具有无创捕获瘤内异质性的能力^[4]。既往大量研究表明,影像组学可辅助临床对疾病进行诊疗^[5-6]。然而,近期研究表明,图像采集、感兴趣区(region of interest, ROI)的勾画、特征提取参数的轻微变化,将导致某些特征发生显著变化^[7]。采用规范化的研究流程有助于提升研究的可重复性。ROI勾画是影像组学研究中的关键部分,勾画方法包括手动勾画、半自动勾画和自动勾画。勾画方法不统一会增加影像组学特征的不稳定。PET图像上肿瘤轮廓不清晰,进一步增加了ROI勾画的难度。本研究旨在探讨通过¹⁸F-FDG PET/CT影像组学特征预测PDAC病理分化程度过程中,ROI勾画方法对影像组学特征及模型预测能力的影响。

资料与方法

1. 研究对象。本研究符合《赫尔辛基宣言》的原则。回顾性纳入自2010年9月至2016年1月间于北京协和医院行¹⁸F-FDG PET/CT显像的PDAC患者共72例[男46例、女26例,中位年龄60(25~87)岁]。纳入标准:经病理证实的PDAC患者;术前1个月内行¹⁸F-FDG PET/CT显像。排除标准:病灶最大径<15 mm;PET/CT图像上病灶显示不清。根据消化系统肿瘤WHO分类,PDAC病理分化程度分为高分化、中分化和低分化。本研究共纳入高分化患者18例,中分化患者49例,低分化患者5例;将中、低分化患者合并为非高分化组。使用分层抽样法将入组患者按3:1的比例随机划分至训练集和验证集。

2. ¹⁸F-FDG PET/CT显像方法。使用德国Siemens Biograph 64 TruePoint PET/CT仪进行全身PET/CT显像;显像剂为本科室合成的¹⁸F-FDG,放化纯>98%。检查前患者禁食6 h以上,检测血糖<11.1 mmol/L,按体质量静脉注射¹⁸F-FDG 5.55 MBq/kg,暗室休息40~60 min后行PET/CT扫描。首先行CT扫描,参数如下:管电流通过自动调制调节,管电压120 kV,层厚

3 mm, 矩阵尺寸 512×512。CT 扫描结束后立刻行 PET 扫描, 根据患者身高进行 5~6 个床位连续扫描, 扫描范围从颅底至股骨近端, 采集时间 2 min/床位。扫描结束后采用有序子集最大期望值迭代法重建 PET 图像, 参数如下: 层厚 3 mm, 分辨率 4.07 mm×4.07 mm, 矩阵尺寸 168×168。

3. 病灶勾画。肿瘤的 ROI 由 2 位具有 5 年以上影像诊断经验的核医学科医师通过 MIM 软件(版本 6.6, 美国 MIM software 公司)进行双盲勾画。总共进行 4 次勾画; 2 位医师分别基于 PET/CT 融合图像沿肿瘤边缘进行手动勾画, 记为 ROI_M1 和 ROI_M2; 1 位医师在 PET 图像上基于标准摄取值(standardized uptake value, SUV)梯度进行半自动勾画, 记为 ROI_G; 另 1 位医师选中高摄取区域以最大 SUV(maximum SUV, SUV_{max})的 40% 进行半自动勾画, 记为 ROI_S。以戴斯相似性系数(Dice similarity coefficient, DSC)分析不同勾画方法的一致性。DSC 的范围为 0~1, 数值越大表明 2 次勾画轮廓的一致性越高, 通常认为 DSC>0.7 表示重合度较好。

4. 特征提取。使用基于 Python 的开源工具包 Pyradiomics 在 PET 和 CT 图像的 ROI 区域分别提取影像组学特征^[8]。提取前通过线性插值方法分别对 PET、CT 及其对应的 ROI 进行重采样以使其各向同性, 重采样后体素变为 3 mm×3 mm×3 mm。分别采用 0.3 SUV 和 25 HU 的等宽分箱对 PET、CT 图像进行强度离散化。基于 4 次勾画提取对应的 4 组特征。从原始图像中提取的特征包括形状特征($n=14$)、一阶特征($n=18$)和纹理特征($n=75$)。从预处理图像中提取一阶特征和纹理特征, 预处理方式包括小波滤波和非线性强度变换(包括平方、平方根、对数、指数和梯度运算)。每个病灶的 PET 和 CT 图像分别提取 1 316 个影像组学特征, 其中 107 个来自原始图像, 744 个来自小波变换图像, 465 个来自非线性强度变换图像。

5. 特征筛选。通过计算不同勾画方式间每个特征的组间相关系数(interclass correlation coefficient, ICC)来评估特征的再现性, ICC \geq 0.75 表明特征具有良好的再现性。基于 3 个 ICC 阈值(0、0.8 和 0.9)分别筛选再现性较好的特征。为了分析不同勾画方式产生特征的预测价值, 分别基于 2 次手动勾画(ICC_M2)和所有勾画(ICC_A4)进行 ICC 分析, 最终得到 5 个特征集合(表 1), 每个特征集合均包含 4 组特征, 分别来自 ROI_M1、ROI_M2、ROI_G 和 ROI_S, 总共 20 组特征, 同一特征集合内不同特

征组中的特征相同。基于每组特征分别进行特征筛选, 首先采用两独立样本 t 检验或 Mann-Whitney U 检验进行单变量分析, 保留 $P<0.1$ 的特征; 然后通过最小绝对收缩与选择算子(least absolute shrinkage and selection operator, LASSO)模型进行特征筛选。

表 1 特征集合

特征集合	ICC 阈值	ICC 计算范围
Set1	0	无
Set2	0.8	ROI_M1、ROI_M2
Set3	0.8	ROI_M1、ROI_M2、ROI_G、ROI_S
Set4	0.9	ROI_M1、ROI_M2
Set5	0.9	ROI_M1、ROI_M2、ROI_G、ROI_S

注: ICC 为组间相关系数, ROI_G 为基于标准摄取值(SUV)梯度的半自动勾画感兴趣区(ROI), ROI_M1 和 ROI_M2 分别为 2 位医师手动勾画 ROI, ROI_S 为基于 40% 最大 SUV(SUV_{max}) 阈值半自动勾画 ROI。ICC 阈值取 0 表示纳入所有影像组学特征, ICC 阈值取 0.8 表示纳入 ICC \geq 0.8 的影像组学特征, ICC 阈值取 0.9 表示纳入 ICC \geq 0.9 的影像组学特征。

6. 模型构建与交叉验证。通过 logistic 回归构建 20 个预测模型。所有模型均在训练集中构建和训练, 并分别在训练集和验证集中评估预测性能。由于数据集内高分化和非高分化患者数量不平衡, 采用人工少数类过采样法对训练集中的高分化数据进行过采样。采用十折交叉验证法来防止模型过拟合。对同一特征集合内部的每组特征进行交叉验证以评估模型泛化能力。

7. 统计学处理。使用 Python(版本 3.7.8; www.python.org)对数据进行统计分析。符合正态分布的定量资料用 $\bar{x}\pm s$ 表示, 不符合正态分布的定量资料用 $M(P_{25}, P_{75})$ 表示; 定性资料以频数(百分比)表示。2 组间比较采用两独立样本 t 检验、 z 检验或 Mann-Whitney U 检验; 多组间比较采用 Kruskal-Wallis 秩和检验, 组间两两比较采用 Bonferroni 法校正显著性水平; 分类变量的比较采用 χ^2 检验。由于入组患者例数较少, 为了消除因数据集划分造成的数据分布不均对模型性能产生的随机影响, 将数据集随机划分 1 000 次, 计算曲线下面积(area under curve, AUC)。 $P<0.05$ 表明差异有统计学意义。

结 果

1. 患者一般情况。共 72 例患者纳入本研究, 其中高分化组 18 例, 非高分化组 54 例, 2 组的性别、年龄、淋巴结转移和术前血糖类抗原 19-9 水平的差异均无统计学意义(χ^2 值: 2.006, 0.472, $t=1.359$, $U=526.0$, 均 $P>0.05$)。训练集共 55 例, 其中高分化组

14 例,非高分化组 41 例;验证集共 17 例,其中高分化组 4 例,非高分化组 13 例。

2.不同勾画方法对 ROI 轮廓的影响。ROI_M1、ROI_M2、ROI_G 和 ROI_S 勾画的轮廓体积分别为 10.29(4.01, 19.43)、9.34(4.26, 17.27)、11.86(5.52, 19.74)、15.08(9.62, 27.44) cm^3 , 差异有统计学意义 ($H=18.641, P<0.05$); 其中, ROI_S 与 ROI_M1、ROI_M2、ROI_G 之间的差异均有统计学意义(调整后均 $P<0.05$), 其他组之间差异均无统计学意义(调整后均 $P>0.05$)。

如表 2 所示, ROI_M1 与 ROI_M2 间、ROI_G 与手动勾画间具有较好的重合度, 而 ROI_S 与手动勾画重合度较差。ROI_G 比 ROI_S 更接近于手动勾画的结果, 准确性更高。

表 2 不同勾画方法间的 DSC 和 ICC [$M(P_{25}, P_{75})$]

组别	DSC	ICC
ROI_M1 与 ROI_M2	0.86(0.76, 0.90)	0.86(0.74, 0.94)
ROI_M1 与 ROI_G	0.86(0.75, 0.91)	0.87(0.72, 0.94)
ROI_M2 与 ROI_G	0.91(0.85, 0.96)	0.94(0.88, 0.98)
ROI_M1 与 ROI_S	0.69(0.53, 0.82)	0.76(0.57, 0.89)
ROI_M2 与 ROI_S	0.67(0.49, 0.80)	0.80(0.63, 0.91)

注: DSC 为戴斯相似性系数, ICC 为组间相关系数, ROI_G 为基于标准摄取值(SUV)梯度的半自动勾画感兴趣区(ROI), ROI_M1 和 ROI_M2 分别为 2 位医师手动勾画 ROI, ROI_S 为基于 40% 最大 SUV (SUV_{max}) 阈值半自动勾画 ROI

3.不同勾画特征的一致性分析。如表 2 所示, ROI_M1 和 ROI_M2 间、ROI_G 与手动勾画间提取的特征具有良好的一致性, 而基于 ROI_S 提取的特征与手动勾画间一致性相对较差。ROI_G 比 ROI_S 提取的特征更接近于手动勾画的特征。

4.特征选择。经 ICC 阈值筛选后, 5 个特征集中特征数量分别为 2 632、1 661、1 563、1 004 和 789 个。随着 ICC 阈值增大, 特征数量逐渐减少。从原始、小波变换和梯度变换图像中得到的稳定特征较多, 平均约占原有特征的 57%、52% 和 64%; 而从平方、平方根、对数和指数变换图像中得到的稳定特征较少, 平均约占原有特征的 38%、38%、25% 和 40%。经 LASSO 筛选后, 20 个特征组中共筛选出 44 个对 PDAC 分化程度有预测价值的特征, 从原始、小波变换、平方、平方根、对数、指数和梯度变换图像中筛选出的特征个数分别为 3、30、4、0、4、1 和 2。

5.模型性能分析。基于每个特征组中筛选出的特征分别构建预测模型。通过控制变量法分析不同因素对模型准确性和泛化能力的影响。(1) ICC 计

算范围对模型性能的影响。在原模型上和交叉验证时, 基于 ICC_M2 构建的模型都比 ICC_A4 构建的模型有更高的准确性, 且在所有的对比中, 2 种模型 AUC 值的差异都具有统计学意义 [ICC = 0.8: 0.80(0.77, 0.84) 与 0.79(0.76, 0.82)、0.72(0.69, 0.76) 与 0.71(0.67, 0.75); ICC = 0.9: 0.76(0.74, 0.79) 与 0.70(0.67, 0.74)、0.70(0.65, 0.74) 与 0.66(0.61, 0.70), z 值: 11.2~59.4, 均 $P<0.05$]。交叉验证时模型准确性均下降约 9%, 且不同 ICC 计算范围模型间准确性下降的幅度差异均无统计学意义 (ICC = 0.8: 训练集 $U=63.0$, 验证集 $U=57.0$; ICC = 0.9: 训练集 $U=66.0$, 验证集 $U=70.0$, 均 $P>0.05$), 提示 2 种模型泛化能力没有差别。所以基于 2 位医师手动勾画的结果进行 ICC 分析更优, 后续分析排除 Set3、Set5。

(2) ICC 阈值对模型性能的影响。在原模型上阈值越小, 模型准确性越高 [ICC 取 0、0.8 和 0.9 时的 AUC 分别为 0.86(0.80, 0.93)、0.80(0.77, 0.84) 和 0.76(0.74, 0.79)]; 在交叉验证时, ICC 阈值取 0.8 时模型准确性最高 [ICC 取 0、0.8 和 0.9 时的 AUC 分别为 0.70(0.65, 0.75)、0.72(0.69, 0.76)、0.70(0.65, 0.74)], 两两比较差异都具有统计学意义 (z 值: -31.0~72.4, 均 $P<0.05$)。ICC 阈值取 0 时, 交叉验证过程中模型准确性下降幅度最大; 与 ICC 阈值为 0 相比, ICC 阈值为 0.8、0.9 时模型准确性下降的幅度差异有统计学意义 (训练集: U 值: 25.0, 33.0; 验证集: U 值: 30.0, 35.0, 均 $P<0.05$); 而 ICC 阈值为 0.8 和 0.9 时模型准确性下降的幅度差异无统计学意义 (训练集: $U=57.0$; 验证集 $U=63.0$, 均 $P>0.05$)。综上, ICC 阈值取 0.8 和 0.9 时, 模型泛化能力相似, 但取 0.8 时模型准确性最高, 因此将 ICC 阈值设为 0.8 更优, 后续分析排除 Set1、Set4。

(3) 特征来源对模型性能的影响。Set2 中模型 AUC 值见表 3。基于原模型分析, 训练集上 ROI_M2 准确性最高, 且各模型间的准确性差异有统计学意义 (z 值: -32.1~58.7, 均 $P<0.05$); 验证集上, ROI_M1 与 ROI_M2、ROI_G 之间的准确性差异无统计学意义 (z 值: -1.850, 1.052, 均 $P>0.05$), ROI_M2 准确性优于 ROI_G ($z=3.031, P=0.002$), ROI_S 准确性最低, 且与 ROI_M1、ROI_M2、ROI_G 之间的准确性差异有统计意义 (z 值: 11.7, 14.5, 10.8, 均 $P<0.05$)。

交叉验证时, ROI_M1 与 ROI_M2、ROI_G 之间准确性下降幅度差异无统计学意义 (t 值: -0.475, 0.712, 均 $P>0.05$); ROI_M2 的准确性比 ROI_G 下降更多 ($t=3.086, P=0.012$); ROI_S 的准确性下降

表 3 特征集合 2(Set2) 不同勾画方法模型在训练集和验证集的曲线下面积($\bar{x}\pm s$)

勾画方法	训练集(55 例)				验证集(17 例)			
	ROI_M1	ROI_M2	ROI_S	ROI_G	ROI_M1	ROI_M2	ROI_S	ROI_G
ROI_M1	0.798±0.040 ^a	0.727±0.044	0.718±0.044	0.707±0.042	0.772±0.139 ^a	0.706±0.144	0.690±0.149	0.684±0.146
ROI_M2	0.789±0.038	0.850±0.032 ^a	0.658±0.050	0.803±0.038	0.728±0.129	0.782±0.115 ^a	0.608±0.162	0.748±0.140
ROI_S	0.703±0.049	0.715±0.040	0.756±0.039 ^a	0.686±0.042	0.653±0.144	0.661±0.147	0.699±0.140 ^a	0.628±0.146
ROI_G	0.713±0.046	0.736±0.040	0.747±0.043	0.809±0.037 ^a	0.683±0.143	0.691±0.146	0.726±0.143	0.766±0.135 ^a

注:ROI_G 为基于标准摄取值(SUV)梯度的半自动勾画感兴趣区(ROI),ROI_M1 和 ROI_M2 分别为 2 位医师手动勾画的 ROI,ROI_S 为基于 40%最大 SUV(SUV_{max}) 阈值半自动勾画 ROI;^a 为原模型曲线下面积,其余为交叉验证曲线下面积

最少,且与 ROI_M1、ROI_M2 之间的差异有统计学意义(t 值:3.451, 2.952, 均 $P<0.05$),但与 ROI_G 之间差异无统计学意义($t=0.504$, $P=0.635$)。

综上,手动勾画模型虽准确性较高,但不同医师勾画存在差异,导致模型性能不稳定,且泛化能力较差;而 ROI_G 可达到与手动勾画相似的准确性,且泛化能力较强。

讨 论

分化程度是恶性肿瘤的重要生物学特征。目前肿瘤组织的病理学信息需要通过手术或活组织检查来确定。影像组学作为一种无创、全面的方法,将在临床中发挥越来越重要的作用。杨天红等^[9]使用 PET/CT 影像组学实现了对肺腺癌表皮生长因子受体基因表达状态的预测。Kaissis 等^[10]的研究表明,通过 MRI 影像组学特征可预测 PDAC 病理分型。Liang 等^[11]发现 CT 影像组学特征与临床特征相结合构建的列线图,可较好地预测胰腺神经内分泌肿瘤的分级。本研究发现,¹⁸F-FDG PET/CT 影像组学特征在预测 PDAC 分化程度方面具有潜在价值。

本研究评估了 ROI 勾画方法对影像组学特征及模型预测能力的影响,发现手动勾画存在不稳定因素,部分特征一致性较差。基于不同医师手动勾画结果构建的模型,预测性能不一致,可能是医师在判断病灶区域时存在主观因素。Owens 等^[12]研究表明,与手动勾画相比,半自动勾画的稳定性更高。本研究中,与手动勾画相比,ROI_G 比 ROI_S 更准确。Belli 等^[13]在基于头颈癌和胰腺癌的研究中得到了相同的结论。另外,SUV 的稳定性与 PET/CT 机型和图像重建算法相关^[14],因此不同中心基于 SUV 阈值勾画的 ROI 可能会存在差异。

本研究基于医师手动勾画结果而非半自动勾画结果进行 ICC 分析可筛选出更好的特征,从而提升模型性能。不进行 ICC 分析(ICC 为 0),原模型预测性能较好,但泛化能力差。将 ICC 阈值设置为 0.8,模

型准确性和泛化能力都较好,这与 Lu 等^[15]的研究结果一致。经 ICC 分析后发现,稳定特征的数量约占特征总量的 50%甚至更少,凸显了 ICC 分析在提升模型可复现性方面的重要性。需要注意的是,稳定的特征并不一定意味着具有较高的预测价值。Kocak 等^[16]对比了轮廓聚焦和边缘收缩 2 mm 这 2 种勾画方法,后者稳定的特征较多,但前者在预测肾透明细胞癌核分级方面性能更优。本研究中对数、梯度变换图像中得到的可再现特征分别约占原有特征数量的 25%和 64%,经 LASSO 筛选后,得到了 4 个有价值的对数变换特征,而有价值的梯度变换特征只有 2 个。

2 位医师手动勾画过程中没有进行讨论,也没有形成统一结果,最终 2 种模型的预测性能有较大差别。基于梯度的半自动勾画可以达到类似于手动勾画的预测性能,且泛化能力更优。需要强调,该结论的前提条件是使用的特征是基于手动勾画结果进行 ICC 分析筛选得到的。部分研究认为,半自动和自动勾画 ROI 的稳定性较高^[12]。本研究未基于梯度进行多次勾画,后续研究有必要基于梯度进行多次勾画,以进一步评估该方法的稳定性。

虽然部分模型在训练集和验证集上的准确性相对一致,但在使用其他来源特征进行交叉验证时,所有模型的性能下降幅度普遍较大,模型泛化能力较差,可重复性较低,进一步揭示了影像组学模型较差的可复现性。数据集较小且没有使用多中心数据进行外部验证时,很容易出现过拟合现象。在研究中应着重关注复现性问题,切勿在本数据集上盲目追求高准确性。

本研究存在以下局限性:(1)影像组学特征提取过程中的参数未纳入变量中,如离散化参数、插值方法、各向同性体素大小等,这些参数对影像组学特征也有较大影响^[17-18];(2)未使用多中心数据进行交叉验证;(3)非高分化组未进一步区分中、低分化,有待扩大患者数量,并分析影像组学特征预测

高、中、低分化患者的能力；(4) 部分影像组学特征存在体积效应^[19]，但没有对不同大小的病变进行单独的分组分析，考虑到入组患者数量较少，分组将加大过拟合风险。

综上所述，基于手动勾画构建的预测模型准确性较高，但模型性能不稳定；基于梯度的半自动勾画可以达到与手动勾画相似的准确性，且模型泛化能力更强。建议在今后的影像组学研究中，由多位医师进行手动勾画，并进行 ICC 分析，筛选出 ICC ≥ 0.8 的特征，以多位医师协商结果或基于梯度的半自动勾画结果作为最终的勾画方案，并进行组内交叉验证。

利益冲突 所有作者均声明不存在利益冲突

参 考 文 献

- [1] Paniccia A, Hosokawa P, Henderson W, et al. Characteristics of 10-year survivors of pancreatic ductal adenocarcinoma [J]. *JAMA Surg*, 2015, 150(8): 701-710. DOI: 10.1001/jamasurg.2015.0668.
- [2] Stark AP, Sacks GD, Rochefort MM, et al. Long-term survival in patients with pancreatic ductal adenocarcinoma [J]. *Surgery*, 2016, 159(6): 1520-1527. DOI: 10.1016/j.surg.2015.12.024.
- [3] Xing H, Hao Z, Zhu W, et al. Preoperative prediction of pathological grade in pancreatic ductal adenocarcinoma based on ¹⁸F-FDG PET/CT radiomics [J]. *EJNMMI Res*, 2021, 11(1): 19. DOI: 10.1186/s13550-021-00760-3.
- [4] 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.
- [5] Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: clinical challenges and applications [J]. *CA Cancer J Clin*, 2019, 69(2): 127-157. DOI: 10.3322/caac.21552.
- [6] 顾丙新, 谢一兆, 王碧芸, 等. 治疗前¹⁸F-FDG PET/CT 摄取异质性在 HER2 阳性转移性乳腺癌疗效预测中的价值 [J]. *中华核医学与分子影像杂志*, 2021, 41(2): 76-81. DOI: 10.3760/cma.j.cn321828-20191112-00256.
Gu BX, Xie YZ, Wang BY, et al. Value of pretreatment ¹⁸F-FDG PET/CT uptake heterogeneity for early prediction of response to targeted therapy in patients with HER2 positive metastatic breast cancer [J]. *Chin J Nucl Med Mol Imaging*, 2021, 41(2): 76-81. DOI: 10.3760/cma.j.cn321828-20191112-00256.
- [7] Reiazi R, Abbas E, Famiyeh P, et al. The impact of the variation of imaging parameters on the robustness of computed tomography radiomic features: a review [J]. *Comput Biol Med*, 2021, 133: 104400. DOI: 10.1016/j.combiomed.2021.104400.
- [8] van Griethuysen J, 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.
- [9] 杨天红, 张胤, 李淑仪, 等. ¹⁸F-FDG PET/CT 影像组学在预测胰腺癌患者 EGFR 突变中的价值 [J]. *中华核医学与分子影像杂志*, 2021, 41(2): 65-70. DOI: 10.3760/cma.j.cn321828-20191108-00255.
Yang TH, Zhang Y, Li SY, et al. Ability of ¹⁸F-FDG PET/CT radiomics features to differentiate EGFR mutation status in patients with lung adenocarcinoma [J]. *Chin J Nucl Med Mol Imaging*, 2021, 41(2): 65-70. DOI: 10.3760/cma.j.cn321828-20191108-00255.
- [10] Kaissis G, Ziegelmayer S, Lohöfer F, et al. A machine learning model for the prediction of survival and tumor subtype in pancreatic ductal adenocarcinoma from preoperative diffusion-weighted imaging [J]. *Eur Radiol Exp*, 2019, 3(1): 41. DOI: 10.1186/s41747-019-0119-0.
- [11] Liang W, Yang P, Huang R, et al. A combined nomogram model to preoperatively predict histologic grade in pancreatic neuroendocrine tumors [J]. *Clin Cancer Res*, 2019, 25(2): 584-594. DOI: 10.1158/1078-0432.CCR-18-1305.
- [12] Owens CA, Peterson CB, Tang C, et al. Lung tumor segmentation methods: impact on the uncertainty of radiomics features for non-small cell lung cancer [J]. *PLoS One*, 2018, 13(10): e0205003. DOI: 10.1371/journal.pone.0205003.
- [13] Belli ML, Mori M, Broggi S, et al. Quantifying the robustness of [¹⁸F]FDG-PET/CT radiomic features with respect to tumor delineation in head and neck and pancreatic cancer patients [J]. *Phys Med*, 2018, 49: 105-111. DOI: 10.1016/j.ejmp.2018.05.013.
- [14] Lodge MA. Repeatability of SUV in oncologic ¹⁸F-FDG PET [J]. *J Nucl Med*, 2017, 58(4): 523-532. DOI: 10.2967/jnumed.116.186353.
- [15] Lu L, Lv W, Jiang J, et al. Robustness of radiomic features in [¹¹C]choline and [¹⁸F]FDG PET/CT imaging of nasopharyngeal carcinoma: impact of segmentation and discretization [J]. *Mol Imaging Biol*, 2016, 18(6): 935-945. DOI: 10.1007/s11307-016-0973-6.
- [16] Kocak B, Ates E, Durmaz ES, et al. Influence of segmentation margin on machine learning-based high-dimensional quantitative CT texture analysis: a reproducibility study on renal clear cell carcinomas [J]. *Eur Radiol*, 2019, 29(9): 4765-4775. DOI: 10.1007/s00330-019-6003-8.
- [17] Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping [J]. *Radiology*, 2020, 295(2): 328-338. DOI: 10.1148/radiol.2020191145.
- [18] Ligerio M, Jordi-Ollero O, Bernatowicz K, et al. Minimizing acquisition-related radiomics variability by image resampling and batch effect correction to allow for large-scale data analysis [J]. *Eur Radiol*, 2021, 31(3): 1460-1470. DOI: 10.1007/s00330-020-07174-0.
- [19] Papp L, Rausch I, Grahovac M, et al. Optimized feature extraction for radiomics analysis of ¹⁸F-FDG PET imaging [J]. *J Nucl Med*, 2019, 60(6): 864-872. DOI: 10.2967/jnumed.118.217612.

(收稿日期: 2021-05-11)