

基于影像组学特征的胰腺神经内分泌肿瘤肝转移患者对索凡替尼的治疗反应预测模型研究

阿卜杜喀迪尔江·阿卜杜热合曼^{1,2}, 李玉婷^{1,2}, 刘容锐^{1,2}, 赵传华², 徐建明²

¹解放军医学院, 北京 100853; ²解放军总医院第一医学中心肿瘤内科, 北京 100853

摘要: **背景** 索凡替尼在晚期胰腺神经内分泌肿瘤(pancreatic neuroendocrine tumors, pNETs)中显示出良好的疗效。然而, 准确且无创地识别出从索凡替尼治疗中获益的患者仍具挑战性。**目的** 建立基于影像组学的机器学习模型, 用于预测索凡替尼治疗pNETs肝转移患者治疗反应, 解析关键影像特征。**方法** 回顾性收集来自7个中心的pNETs肝转移患者的增强CT门静脉期(portal venous phase, PVP)影像及临床资料。采用最小绝对收缩和选择算子(least absolute shrinkage and selection operator, LASSO)回归筛选出非零系数关键特征, 并结合Bootstrap重采样筛选稳定的影像特征。采用嵌套交叉验证策略, 基于堆叠集成算法构建机器学习模型, 以支持向量机(support vector machine, SVM)、随机森林(random forest, RF)、XGBoost(eXtreme gradient boosting)3种算法作为初级分类器, 次级分类器采用多层感知机(multilayer perceptron, MLP)进行融合建模, 用于预测索凡替尼治疗反应(靶病灶缩小>15%定义为应答)。同时基于临床变量, 通过多变量逻辑回归建立临床模型, 与机器学习模型进行对比。通过受试者工作特征曲线下面积(area under the curve, AUC)、特异度、敏感度和准确率、校准曲线、决策曲线等方面评估模型的性能。此外, 利用LASSO回归特征系数分析影像组学特征与治疗反应间的相关性。**结果** 最终102例患者纳入本研究, 应答组和无应答组的患者各51例。基于LASSO回归筛选出的5个Bootstrap重采样法中高频入选的核心影像组学特征构建的机器学习模型显示出较好的判别能力, AUC为0.805、准确率为76.47%, 敏感度为84.31%, 其整体表现优于对应临床模型(AUC=0.623, 准确率=68.40%, 敏感度=74.62%)。在总体数据集模型预测的应答组无进展生存期(progression-free survival, PFS)较无应答组(HR=0.573, 95% CI: 0.349~0.941, P<0.05)更长。LASSO回归筛选的关键影像学特征提示治疗前肿瘤负荷较小、血供更丰富、均质性高的患者更可能从索凡替尼治疗中获益。**结论** 基于增强CT影像特征构建的机器学习模型显示出预测pNETs肝转移患者对索凡替尼疗效的潜力, 并提示部分影像组学特征可能与预后分层相关。

关键词: 计算机断层扫描; 胰腺神经内分泌肿瘤; 肝转移; 索凡替尼; 影像组学; 治疗反应

中图分类号: R735.9

文献标志码: A

文章编号: 2095-5227(XXXX)XX-0001-11

DOI: 10.12435/j.issn.2095-5227.25122601

引用本文: 阿卜杜喀迪尔江·阿卜杜热合曼, 李玉婷, 刘容锐, 等. 基于影像组学特征的胰腺神经内分泌肿瘤肝转移患者对索凡替尼的治疗反应预测模型研究 [J]. 解放军医学院学报, XXXX, XX (XX): 1-11.

Radiomics model for predicting treatment response to surufatinib in patients with pancreatic neuroendocrine tumor liver metastases

ABUDUKADIERJIANG Abudurehman^{1,2}, LI Yuting^{1,2}, LIU Rongrui², ZHAO Chuanhua²

¹Chinese PLA Medical School, Beijing 100853, China; ²Senior Department of Oncology of PLA General Hospital, Beijing 100853, China

Corresponding author: XU Jianming. Email: jmxu2003@163.com

Abstract: Background Surufatinib has demonstrated favorable efficacy in advanced pancreatic neuroendocrine tumors (pNETs). However, accurately and noninvasively identifying patients who will benefit from surufatinib remains challenging. **Objective** To develop a radiomics-based machine learning (ML) model to noninvasively predict treatment response to surufatinib in patients with pNETs liver metastases and interpret the key imaging features associated with response. **Methods** Clinical data and portal venous - phase contrast-enhanced CT images of patients with pNETs liver metastases from seven centers were retrospectively collected. A machine learning (ML) model was constructed using a stacking ensemble algorithm within a nested cross-validation framework. Support vector machine (SVM), random forest (RF), and eXtreme gradient boosting (XGBoost) served as base classifiers, and a multilayer perceptron (MLP) was used as the meta-classifier for fusion to predict response to surufatinib treatment (response defined as >15% reduction in the target lesion size). In parallel, a clinical model was built using multivariable logistic regression based on clinical variables and compared with the ML model. Model performance was evaluated using the area

收稿日期: 2025-12-26

第一作者: 阿卜杜喀迪尔江·阿卜杜热合曼, 硕士. Email: abudukadierjiang@163.com

通信作者: 徐建明, 博士, 主任医师. Email: jmxu2003@163.com

under the curve (*AUC*), specificity, sensitivity, accuracy, calibration curves, and decision curve analysis. Feature coefficients derived from least absolute shrinkage and selection operator (LASSO) regression were analyzed to evaluate the association between radiomic features and treatment response. **Results** A total of 102 patients were ultimately included in this study, with 51 patients in the responder group and 51 patients in the non-responder group. The machine learning model constructed using five core radiomics features selected by LASSO regression and frequently retained across Bootstrap resampling iterations achieved a mean *AUC* of 0.805, an accuracy of 76.47% and a sensitivity of 84.31%, with overall performance superior to that of the corresponding clinical model (*AUC*=0.623, accuracy of 68.40%, and sensitivity of 74.62%). In the overall cohort, the model-predicted responder group had a significantly longer progression-free survival (PFS) than the non-responder group (*HR*=0.573, 95% *CI*: 0.349 - 0.941, *P*<0.05). The LASSO-selected key imaging features suggested that patients with lower pretreatment tumor burden, richer blood supply, and higher tumor homogeneity were more likely to benefit from surufatinib. **Conclusion** A CT radiomics - based machine learning model can effectively predict the efficacy of surufatinib in patients with pNETs liver metastases and identify interpretable radiomic biomarkers with prognostic significance.

Keywords: computed tomography; pancreatic neuroendocrine tumors; liver metastasis; surufatinib; radiomics; treatment response

Cited as: ABUDUKADIERJIANG Abudurehman, LI Yuting, LIU Rongrui, et al. Radiomics model for predicting treatment response to surufatinib in patients with pancreatic neuroendocrine tumor liver metastases [J]. Acad J Chin PLA Med Sch, XXXX, XX(XX): 1-11.

神经内分泌肿瘤 (neuroendocrine tumors, NETs) 是一类起源于弥散性神经内分泌细胞的罕见肿瘤, 累及全身多器官, 尤以胃肠道、胰腺多见。随着影像学检查普及、内镜筛查及病理诊断规范化, 该病种的发病率在全球多数地区呈上升趋势^[1-3]。胰腺神经内分泌肿瘤 (pancreatic neuroendocrine tumors, pNETs) 是 NETs 的重要亚型之一, 具有明显异质性, 早期症状隐匿, 多数患者确诊时已出现远处转移, 以肝转移最常见, 严重影响生存预后^[1,4-5]。

索凡替尼是一种新型口服多靶点酪氨酸激酶抑制剂, 兼具抗血管生成与免疫调节作用机制^[6-7]。既往临床研究结果显示, 索凡替尼在晚期 pNETs 患者中具有明确的抗肿瘤活性, 能够改善疾病控制且安全性可管理^[7-8]。然而, 当前仍缺乏可靠且无创的方法, 在治疗前或早期疗程中预测患者对索凡替尼的敏感性与获益程度, 传统临床指标与肉眼观察的影像评价在捕捉肿瘤微观异质性、反映抗血管治疗早期效应方面存在局限^[9]。值得注意的是, 我们在临床工作中发现索凡替尼治疗获益较好的患者, 其肿瘤在 CT 增强影像上往往表现为较高强化, 提示病灶富血供或血管生成活跃, 这与索凡替尼等以“抗血管生成”为核心的作用特点在机制层面相契合^[10-11]。基于此, 能够表征影像内部异质性的影像组学方法, 或可为索凡替尼疗效预测提供新的无创评估思路。

近年来影像组学与人工智能技术快速发展, 使得从常规影像中提取高通量定量特征成为可能^[12], 量化异质性、血供及微环境变化, 结合机器学习的方法有望实现对治疗反应的早期预测与

风险分层, 能够为药物疗效评估提供更客观、可量化和可解释的证据, 从而支持个体化治疗策略制定^[13-14]。因此, 本研究拟基于 pNETs 患者肝转移病灶增强 CT 门静脉期影像组学特征, 构建机器学习模型, 预测患者的治疗反应, 识别与索凡替尼治疗反应相关的影像标志物并探索其在个体化治疗中的决策支持价值。

1 对象与方法

1.1 研究对象

本研究回顾性收集以解放军总医院为主, 华西医院、上海市第一人民医院、西安交通大学第一附属医院、中日友好医院、郑州大学附属第一医院、北京肿瘤医院等全国 7 个中心 2015 年 12 月至 2022 年 11 月接受索凡替尼治疗的 pNETs 患者。纳入标准: (1) 年龄 ≥ 18 岁; (2) 确诊为不可切除或转移性; (3) 病理分级为高分化 (根据 WHO2019 分级 G1-G2) 的 NETs, 原发部位位于胰腺; (4) 患者的东部肿瘤合作组 (Eastern Cooperative Oncology Group, ECOG) 体能状态评分为 0 或 1。排除标准: (1) 腹部基线增强 CT 图像数据缺失或不完整, 或图像格式不支持分析; (2) 合并肝脏囊肿导致无法勾画准确病灶的数据; (3) 肝脏无可测量病变; (4) 缺失基线肿瘤学评估结果; (5) 临床随访数据不完整。本研究经解放军总医院医学伦理委员会批准 (编号: S2025-763-01)。

1.2 治疗方案

本研究通过病历系统筛选并调取接受索凡替尼口服治疗的患者资料。其标准给药方案为口服索凡替尼治疗, 标准剂量 300 mg/d, 连续给药, 以

28 d(4周)为一个治疗周期,持续用药至达到治疗终止事件(疾病进展或不可耐受的毒性)。

1.3 结局指标及临床指标采集

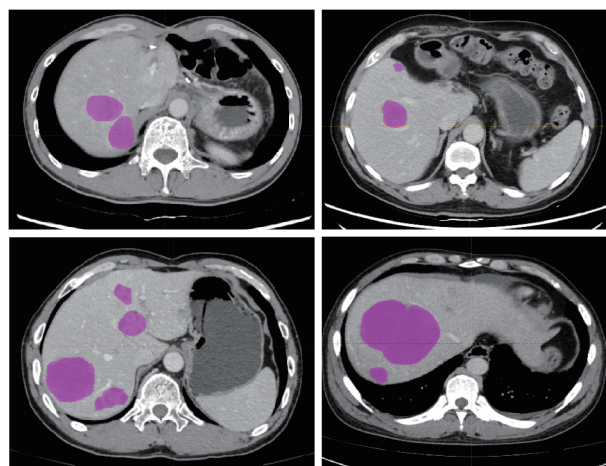
疗效评价在临床上常采用RECIST 1.1标准。考虑到SANET-p临床研究结果显示,pNETs索凡替尼治疗后多数病例表现为肿瘤稳定或轻度缩小^[7-8],为提高反应分层的敏感性,本研究参考既往惰性肿瘤的改良影像学评价思路及相关研究中较低缩小阈值的设置^[15-18],在RECIST 1.1基础上探索性增设“最小缓解”(minor response, MR):定义为靶病灶最大径较基线缩小15%~30%。完全缓解(complete response, CR)、部分缓解(partial response, PR)与疾病进展(progressive disease, PD)按RECIST 1.1判定,稳定疾病(stable disease, SD)定义为未达到MR且未出现PD。以最佳总体疗效(best overall response, BOR)分组: BOR为MR/PR/CR归为应答者, BOR为SD/PD归为非应答者。

收集患者基线腹部增强CT影像及其临床病理资料,包括研究者评估的肝转移灶最大直径、年龄、性别、ECOG评分(0或1)、确诊到接受索凡替尼治疗的时间、转移部位(仅有肝转移;肝转移合并肝外一个部位转移;肝转移合并肝外多个部位转移)、手术史(是否手术切除原发病灶)、病理分级(G1或G2)及Ki-67增殖指数(<2%或2%~20%)、随访终点等数据。

1.4 图像的采集与标准化

本研究采用标准化的影像标注与质控流程:2名资深放射科医师在盲法下独立标注门静脉期增强CT图像的病灶感兴趣区域(Region of Interest, ROI),随机抽取25%影像进行复核,争议病例由专家仲裁,确保标注一致性与质量。图像预处理包括CT值窗宽标准化(-1 000~3 000 HU)、直方图分箱(bin width=10)及各向同性体素重采样(1 mm×1 mm×1 mm,拉格朗日插值)。随后,基于LIFEx平台(v7.3.0)提取影像组学特征。ROI勾画的示例见图1。

我们优先选择门静脉期(portal venous phase, PVP)CT作为主要分析时相,因其在多中心真实世界中更易获得且扫描方案更一致,与动脉期或延迟期相比,PVP影像组学特征通常更稳定、可重复。同时PVP期病灶与肝实质之间的对比更清晰,有助于肿瘤分割更准确并降低不同时相带来的变异。对于多病灶患者,通过计算平均值将病灶水平的特征聚类至患者水平。



pNETs: 胰腺神经内分泌肿瘤。

图1 pNET肝转移患者肝脏病变感兴趣区域(ROI)勾画的典型示例
Fig. 1 Representative examples of ROI delineation for metastatic liver lesions in pNETs patients with liver metastases

LIFEx(7.3.0版)提取影像学特征,包括形状、一阶直方图及纹理矩阵特征(如灰度共生矩阵、灰度游程长度矩阵等)。LIFEx特征提取遵循影像生物标志物标准化倡议(image biomarker standardisation initiative, IBSI)的建议^[13,19]。

本研究扫描主要采用相对一致的腹部门静脉期增强CT标准方案:管电压为120 kV并配合管电流自动调制,层厚1~2 mm,采用软组织重建核。对比剂为碘造影剂(300~370 mgI/mL),以3.0~3.5 mL/s注射,门静脉期采集时间窗约为注射后60~70 s。上述中心间差异总体处于常规临床范围。

鉴于数据集的多中心特性及各中心间可能存在的系统性差异(批次效应),我们对提取的影像组学特征应用ComBat批次校正方法,将各中心视为独立批次处理。校正后,所有连续性影像组学特征均采用Z分数标准化进行归一化处理,以消除尺度变异并促进模型收敛。

1.5 影像组学特征的选择

为实现高维特征数据的降维与精选,并识别与疗效应答相关的关键特征,本研究采用最小绝对收缩和选择算子(least absolute shrinkage and selection operator, LASSO)回归进行影像组学特征初筛,通过模型输出的非零系数特征,展示与疗效应答相关的影像组学特征及其系数方向。在模型构建过程中,为降低单次特征筛选的随机性并提升入选特征的稳健性,采用Bootstrap重抽样策略进行多轮迭代筛选,记录特征在重抽样LASSO模型中被选中的入选频率。最终,确定选中频率

前5的核心特征作为“稳定特征”，并以此作为构建后续预测模型训练的输入特征。

1.6 机器学习模型的构建

本研究采用重复100次的5折嵌套交叉验证策略进行机器学习模型训练与内部验证评估。在每一轮循环中，数据集被随机划分为5个大小相近的外层折，并依次以其中4折为训练折，用于模型构建，其中1折为验证折，用于性能评估。机器学习预测模型采用两级堆叠(Stacking)集成架构，初级分类器包含支持向量机(Support Vector Machine, SVM)、随机森林(Random Forest, RF)和极致梯度提升(eXtreme Gradient Boosting, XGBoost)3个基模型，次级分类器为多层感知机(Multi-Layer Perceptron, MLP)构成的元模型。在训练阶段，首先于4折训练折内通过内部5折交叉验证结合随机搜索对初级分类器(SVM、RF、XGBoost)进行超参数优化，并基于优化后参数生成折外预测概率值(Out-of-Fold predictions, OOF)。随后，将OOF概率值作为元特征与真实标签结合，用于训练次级分类器MLP。模型训练完成后，将留出的1折验证折依次通过初级与次级分类器，输出最终堆叠集成的机器学习模型对于应答的预测结果。在5折嵌套交叉验证过程中，每位患者均被分配至一个外层折，并在其所在折作为验证折时获得1次外层验证结果。因此，在100次重复实验中，每位患者作为外层验证折样本累计获得100个外层验证预测概

率。本研究取其算术平均值作为该个体的最终预测概率值，并报告核心指标(如AUC及其95% CI)，以评估模型的判别能力，并据此完成预测分组及其临床结局分层表现。模型构建流程见图2。

实施多重措施来降低过拟合并提升泛化能力：采用了重复100次的分层五折嵌套交叉验证，在每一轮循环中，通过分层抽样将数据集划分为相互隔离的训练折与验证折，验证折不参与该轮数据标准化、特征选择、模型训练和参数调优，以降低过拟合风险并提高内部验证结果的客观性。用LASSO进行特征筛选与降维，Bootstrap重采样稳定性分析，以少量精选特征建模来降低复杂度与噪声。采用堆叠集成策略，初级分类器中基模型用分层5折交叉验证生成折外预测值预测供次级分类器元模型训练，避免信息泄露，同时引入多类型基模型增强多样性。

1.7 模型性能评估

为评估模型的预测性能及稳定性，本研究通过整合每位患者100次验证循环中的外层预测概率，并取其平均值作为最终预测概率结果，计算受试者工作特征曲线下面积(area under the curve, AUC)、特异度、敏感度及准确率及其95%置信区间(采用Bootstrap重采样法)等指标对模型性能进行评估。通过计算本队列中的最优约登指数确定模型生成的个体平均预测概率的最佳截断值，从而将患者客观地划分为预测应答组与无应答组。通

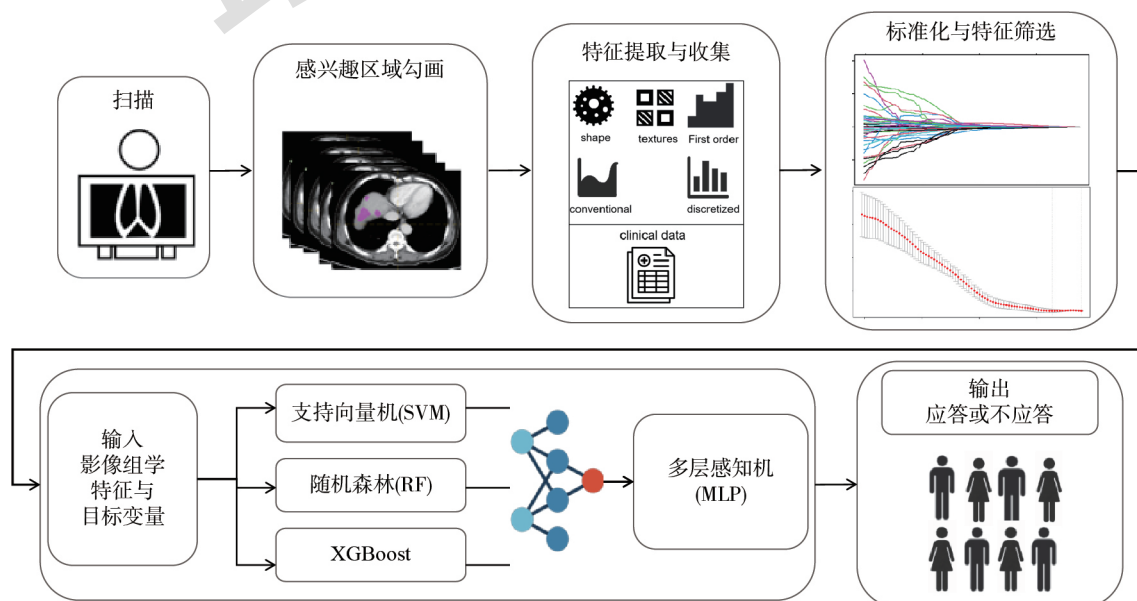


图2 索凡替尼治疗晚期胰腺神经内分泌肿瘤肝转移患者的疗效预测模型构建的流程

Fig. 2 Flowchart of the predictive model construction for surufatinib efficacy in advanced pancreatic neuroendocrine tumor patients with liver metastases

过校准曲线检验模型预测概率与实际观测概率的一致性, 并进一步计算 Brier 评分以量化预测概率的整体误差。模型临床实用性通过决策曲线分析 (decision curve analysis, DCA) 进行评价。

此外, 我们基于 SANET-p 亚组分析中与疗效相关, 且在多中心病历中记录较完整的变量 (Ki-67 指数、病理分级、ECOG 评分、诊断至用药时间及转移类型) 构建多变量逻辑回归临床模型 (同样采用重复 100 次的 5 折嵌套交叉验证策略), 并与对应机器学习模型进行比较。上述变量均为临床试验中常规收集的基线信息。考虑到症状、既往治疗及血清生物标志物等变量在各中心数据完整性与一致性不足可能加剧小样本下的不稳定性, 未纳入主分析。

1.8 机器学习模型预测分组的无进展生存期差异

在完成机器学习模型构建与性能评估后, 进一步在数据集开展无进展生存期 (progression-free survival, PFS) 分析, 以评估模型预测分组对临床结局的分层能力。本研究将 PFS 定义为自开始索凡替尼治疗至影像学进展或任何原因死亡的时间, 末次随访未进展者予以删失。按机器学习模型预测应答分组, 基于模型预测的个体平均预测概率, 将患者分为应答和无应答组, 绘制 Kaplan-Meier 生存曲线, 并采用 log-rank 检验及 Cox 比例风险模型比较组间差异, 报告 HR 及其 95% CI, 以评估模型预测分组对临床结局的区分能力。在相同数据集中依据改良评价标准按实际研究者评估疗效应答状态, 采用相同方法进行 PFS 分析, 评价本研究疗效分组标准是否具有一定的临床相关性。以研究者实际评估分组为参照, 综合呈现其与模型预测分组的生存曲线及 HR 结果。直观考察模型预测结果与后续临床结局之间的一致性及其在疗效预测和临床结局的分层中的潜在应用价值。

1.9 统计学方法

使用 RStudio (R 版本 4.5.1) 以及 PyCharm (Python 版本 3.13) 软件。服从正态分布的计量资料采用均数 \pm 标准差 ($\bar{x} \pm s$) 表示, 组间比较采用 Student's *t* 检验; 非正态分布的计量资料以中位数 *M* (IQR) 表示, 组间比较采用 Mann-Whitney *U* 检验。分类资料以例数或率表示, 组间比较采用 χ^2 检验或 Fisher 确切概率法。双侧 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 纳入患者的临床基线特征

来自 7 个中心的 118 例患者中, 排除 16 例患者 (2 例基线 CT 图像缺失, 12 例 BOR 缺失, 1 例囊肿干扰, 1 例未见肝转移灶) 最终纳入 102 例进行分析, 总体平均年龄 (47.9 ± 11.9) 岁, 其中男性 56 例, 女性 46 例。按照治疗反应评估标准, 51 例 (50.0%) 患者达到治疗应答。应答组平均年龄高于无应答组为 [(50.1 ± 11.3) 岁 vs (45.7 ± 12.1) 岁, $P = 0.061$], 但组间差异无统计学意义。性别构成在两组间差异亦无统计学意义 (男性占比: 49.0% vs 60.8%, $P = 0.233$)。两组在诊断至用药时间、Ki-67 指数、病理分级、ECOG 评分、用药前是否行手术切除原发灶、转移模式等方面差异亦无统计学意义 ($P > 0.05$), 见表 1。

2.2 影像组学特征筛选结果

以治疗应答为结局指标进行特征筛选。从每例患者门静脉期的 CT 图像中提取 55 个影像组学特征, 并纳入研究者评估的肿瘤最大径。采用 LASSO 回归并通过交叉验证确定最优正则化参数, 按最小误差准则选择 $\lambda_{\min} = 0.031$, 筛选出 10 个非零系数、与疗效应答相关的关键特征 (图 3A ~ 3C)。其中, 灰度游程长度矩阵-长游程高灰度强调 (GLRLM_LRHGE) 获得最大正系数; 此外, 离散化强度直方图-能量 [= 均匀性] (DISCRETIZED_HISTO_Energy [= Uniformity])、邻域灰度差异矩阵-粗糙度 (NGLDM_Coarseness)、常规 HU-偏度 (CONVENTIONAL_HUSkewness)、灰度区带长度矩阵-高灰度区带强调 (GLZLM_HGZE) 及邻域灰度差异矩阵-对比度 (NGLDM_Contrast) 均为正系数, 提示上述特征升高与应答概率增高相关。相反, 研究者评估的肿瘤最大径与形状学-体积 (体素计数) (SHAPE_Volume) 呈负系数, 同时, 离散化 HU-峰度 (DISCRETIZED_HUKurtosis) 与离散化 HU-超额峰度 (DISCRETIZED_HUExcessKurtosis) 亦为负系数, 提示峰度相关指标升高与非应答倾向相关。进一步评估与疗效相关特征的稳定性与一致性, 在模型构建过程中, 采用 Bootstrap 法进行多次重复抽样, 并记录特征的选择频率。最终发现了总计入选频率前 5 的核心“稳定特征” (图 3D), 即研究者评估的肿瘤最大径、GLRLM_LRHGE、SHAPE_Volume (vx)、NGLDM_Coarseness 及 NGLDM_Contrast 为预测模型的核心变量。所有特

表 1 pNETs肝转移总数据集中应答与无应答患者基线临床病理特征

Tab. 1 Baseline characteristics of responders and non-responders in the entire cohort of pNETs with liver metastases

指标	总计(n=102)	应答组(n=51)	无应答组(n=51)	t/U/ χ^2 值	P值
年龄/(岁, $\bar{x} \pm s$)	47.9 \pm 11.9	50.1 \pm 11.3	45.7 \pm 12.1	1.898 ^a	0.061
男性/(例, %)	56(54.9)	25(49.0)	31(60.8)	1.426 ^c	0.233
诊断至用药时间/[月, <i>M(IQR)</i>]	17.9(5.5, 41.3)	11.3(4.5, 44.6)	21.6(10.9, 37.5)	1.495 ^b	0.194
Ki-67指数[% , <i>M(IQR)</i>]	6.5(3.0, 10.0)	8.0(3.0, 10.0)	5.0(3.0, 10.0)	1.241 ^b	0.689
病理分级/(例, %)					
G1	14(13.7)	7(13.7)	7(13.7)	0 ^c	1.000
G2	88(86.3)	44(86.3)	44(86.3)		
东部肿瘤协作组(ECOG)/(例, %)					
0	67(65.7)	33(64.7)	34(66.7)	0.044 ^c	0.835
1	35(34.3)	18(35.3)	17(33.3)		
用药前手术治疗切除原发灶/(例, %)					
是	43(42.2)	23(45.1)	20(39.2)	0.362 ^c	0.547
否	59(57.8)	28(54.9)	31(60.8)		
转移/(例, %)					
仅肝转移	46(45.1)	28(54.9)	18(35.3)		
肝内及一处肝外转移	5(4.9)	2(3.9)	3(5.9)	3.962 ^c	0.138
肝内肝外多处转移	51(50.0)	21(41.2)	30(58.8)		

^a为 Student's *t* 检验, ^b为 Mann-Whitney U 检验, ^c为 χ^2 检验。pNETs:胰腺神经内分泌肿瘤。

征的提取与数学定义等详细信息可在影像生物标志物标准化倡议(Image Biomarker Standardisation Initiative, IBSI)国际标准网站上查找(<https://ibsi.readthedocs.io>)。

2.3 模型的建立及性能

基于每位患者在100次重复5折嵌套交叉验证中获得的外层预测概率平均值,两级分类器堆叠集成的机器学习模型表现出良好的判别效能:*AUC*为0.805(95% *CI*: 0.719 ~ 0.884),准确率为76.47%(95% *CI*: 67.65% ~ 83.33%),敏感度为84.31%(95% *CI*: 74.51% ~ 94.12%),特异度为68.63%(95% *CI*: 56.86% ~ 80.39%),*F1*值为0.781(95% *CI*: 0.708 ~ 0.849)。与临床模型相比(*AUC*为0.623,准确率为68.4%,敏感度为74.6%,特异度为62.6%,*F1*值为0.684),机器学习模型整体表现更优,提示机器学习模型在识别索凡替尼治疗应答方面具有更好的区分能力和综合预测效能。见表2,图4A。

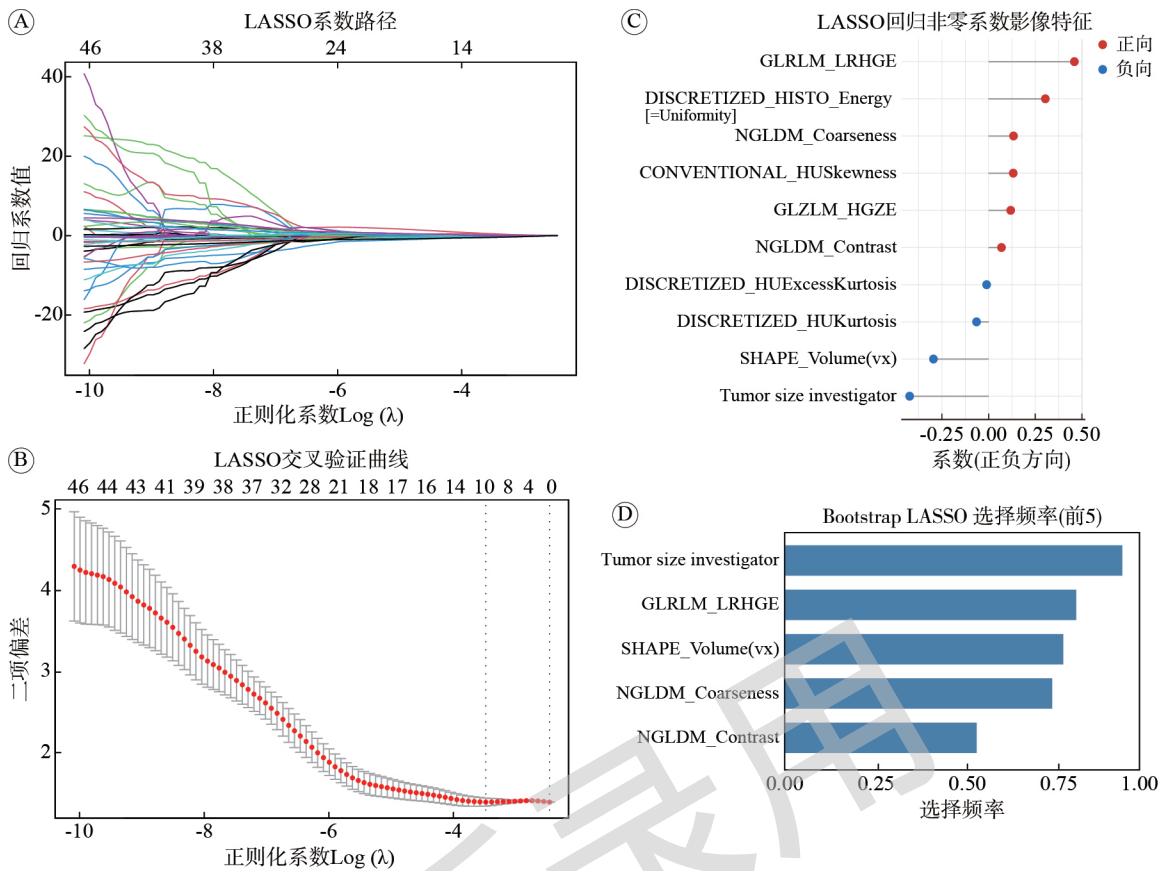
此外,基于最优约登指数阈值的混淆矩阵结果(图4B)进一步表明,模型在敏感度与特异度之间取得了相对均衡的判别表现。机器学习模型校准曲线总体接近理想线(图4C)Brier score为0.194,提示模型整体预测误差较低,模型具有一定的校准能力。决策曲线分析(DCA)(图4D)显示,该模型

在多数阈值概率范围内可获得正向净获益,提示其具有一定临床应用价值。

2.4 模型预测分组的PFS分析

根据治疗应答分组的PFS分析中, Kaplan-Meier生存曲线显示应答组与无应答组存在明显分离,应答组的中位PFS为16.8个月,无应答组9.3个月,与无应答组相比,应答组疾病进展风险显著降低(*HR*=0.391, 95% *CI*: 0.233 ~ 0.655, *P*<0.001),提示研究人员评估的疗效分组能够较好地区分患者预后。

基于机器学习模型预测分组的PFS分析同样显示两组生存曲线明显分离,且差异具有统计学意义(应答组 vs 无应答组: *HR*=0.573, 95% *CI*: 0.349 ~ 0.941, *P*=0.027<0.05)。图中所示两组中位PFS分别为16.6个月和11.0个月,提示机器学习模型预测分组亦可在一定程度上区分预后差异。综合来看,无论采用治疗应答分组还是机器学习模型预测分组的分析均提示,疗效应答与更佳预后相关。上述结果表明,基于治疗前增强CT构建的机器学习模型展现出较好的结局分层能力,其预测的预后分层与实际治疗后预后分层总体呈一致趋势。机器学习模型在一定程度上具有结局分层能力,可有助于治疗前识别潜在获益人群,为早期疗效预测及个体化治疗决策提供参考。



A: LASSO 系数路径图, 横轴为正则化参数 $\text{Log}(\lambda)$, 纵轴为各候选特征的回归系数。随着 λ 增大(惩罚增强), 各特征系数逐渐收缩并趋于 0; 上方数字表示对应 λ 下模型中非零系数(入模特征)数量; B: LASSO 交叉验证曲线, 横轴为 $\text{Log}(\lambda)$, 纵轴为交叉验证误差(二项偏差)。两条虚线分别表示 λ_{\min} (误差最小)与 $\lambda_{1\text{se}}$ (1-SE 规则下更简洁的模型)对应的 λ , 上方数字同样表示非零系数数量; C: 最优 λ 下的非零特征及系数, 展示在选定 $\lambda_{\min}=0.031$ 时保留下来的特征及其回归系数大小与方向, 红色为正向关联, 蓝色为负向关联(系数绝对值越大, 对模型贡献越大); D: Bootstrap LASSO 特征选择频率图(前 5)。pNETs: 胰腺神经内分泌肿瘤。

图 3 pNETs 肝转移患者影像组学特征的 LASSO 回归筛选及最优惩罚参数选择结果

Fig. 3 LASSO regression for radiomics feature selection and penalty parameter (λ) tuning of pNETs patients with liver metastases

表 2 机器学习模型与临床模型在 pNET 肝转移患者治疗应答预测中的性能对比

Tab. 2 Comparison of machine learning and clinical model performance for predicting treatment response in pNETs with liver metastases

评价指标	机器学习模型	临床模型
AUC	0.805(0.719 ~ 0.884)	0.623(0.554 ~ 0.757)
准确率/%	76.47(67.65 ~ 83.33)	68.40(57.14 ~ 80.05)
敏感度/%	84.31(74.51 ~ 94.12)	74.62(36.21 ~ 99.00)
特异度/%	68.63(56.86 ~ 80.39)	62.55(28.57 ~ 99.00)
F1 值	0.781(0.708 ~ 0.849)	0.684(0.500 ~ 0.858)

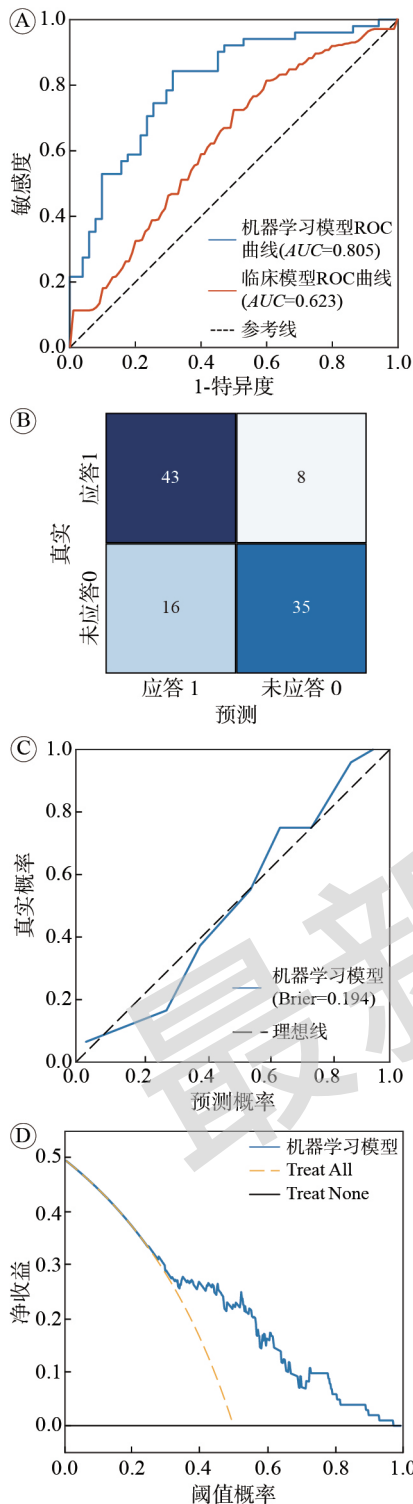
见图 5。

3 讨论

靶向治疗在当前肿瘤治疗中呈现快速发展趋势, 以索凡替尼为代表的新型抗血管生成药物在 pNETs 中显示出良好疗效。SANET-p III 期研究结

果表明, 索凡替尼可将晚期 pNETs 患者的中位 PFS 显著延长^[7-8]。然而, 仍有部分患者无法从索凡替尼的治疗中获益。如何在用药前通过无创方法甄别出哪些患者将从索凡替尼等抗血管生成药物中获益, 是亟待解决的临床难题。

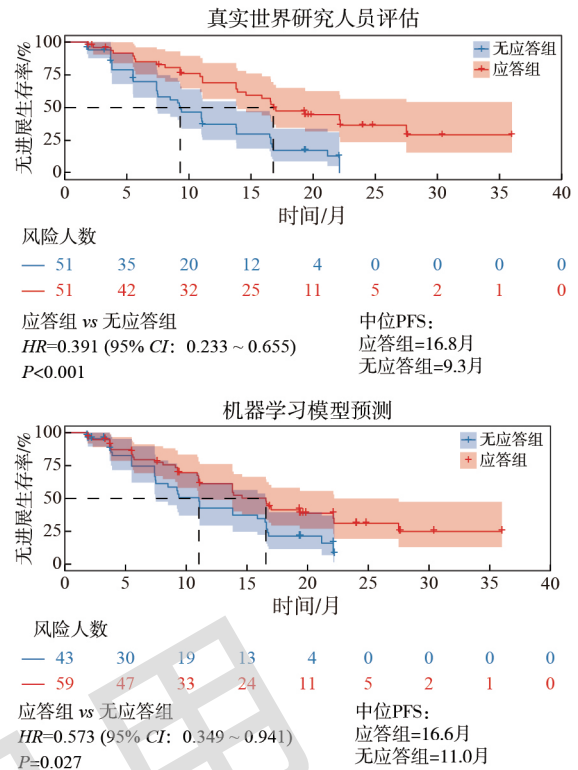
随着机器学习、深度学习等人工智能技术的发展, 影像组学能够从医学影像中提取肉眼难以察觉的肿瘤特征, 并建立模型进行疗效预测, 在多种肿瘤的预后和疗效预测中已体现出重要价值^[20]。本研究基于多中心患者影像数据, 严格遵循影像生物标志物标准化倡议标准, 通过 LASSO 回归结合 Bootstrap 重采样稳定性分析, 从高维特征空间中精选出稳定性最强的影像组学特征, 并基于这些特征构建机器学习模型, 采用重复 K 折交叉验证策略, 即在整个数据集上进行 100 次重复的 5 折嵌套交叉验证与评估。结果显示, 该模型在多次迭代验证中表现出优秀且稳健的预测效能,



pNETs: 胰腺神经内分泌肿瘤; AUC: 受试者工作特征曲线下面积。A: 基于个体最终预测概率绘制的机器学习模型与临床模型 ROC 曲线; B: 机器学习模型的混淆矩阵; C: 机器学习模型校准曲线; D: 决策曲线分析(DCA)。

图4 机器学习模型与临床模型在pNET肝转移患者治疗应答预测中的性能评估

Fig. 4 Performance evaluation of machine learning models and clinical models



pNETs: 胰腺神经内分泌肿瘤。

图5 pNETs肝转移患者基于真实世界评估与机器学习模型预测分组的无进展生存期Kaplan-Meier曲线

Fig. 5 Kaplan-Meier curves of PFS in pNETs with liver metastases stratified by real-world assessment and machine-learning model predicted groups

AUC达到0.805, 能够在一定程度上区分索凡替尼治疗的应答者与非应答者。

在pNETs影像组学研究中, 多数围绕分级、增殖活性、转移风险与复发结局展开^[21-22], 如Ye等^[23]开发了一种基于CT影像组学的可解释机器学习模型用于非侵入性评估病理分级, 其最佳随机森林模型在独立测试集中的AUC为0.779, 准确率为77.0%。Gu等^[24]基于CT的影像组学-深度学习联合模型用于预测非功能性pNETs的淋巴结转移, 在两中心验证中的AUC分别为0.88和0.91, 从而支持术式与淋巴清扫策略的术前决策。在长期结局方面, 也有研究基于37例患者术前CT影像组学建立复发风险或无复发生存预测模型, 并证实预后预测的价值, 但该研究样本量较小, 且主要基于生存分层分析, 未见验证结果^[25]。需要指出的是, 尽管pNETs影像组学在分级、转移、预后方面已有较多探索, 但与“治疗获益预测”直接相关的研究仍相对有限, 且不少工作纳入的是混合来源的NETs人群, 如有研究尝试用CT影像组学预测依维莫司治疗结局, 在25例转移性NET患者

中探索了CT影像组学对依维莫司治疗相关PFS及死亡的预测价值,部分特征在内部及合成外部队列中的AUC约为0.84~0.90,但样本量有限且人群异质性较高,限制了结果的稳定性和外推性^[26]。Ligero等^[27]在抗血管生成方案中探索影像表型与疗效和生存的关联研究基于治疗前强CT对89例患者肝转移灶构建的影像学评分模型在训练集、内部验证集及外部舒尼替尼队列中分别取得AUC 0.75、0.67和0.67,整合Ki-67指数及原发部位后训练集AUC提升至0.79,发现高影像学评分与血浆VEGFR2/ANG2升高及PFS延长显著相关,不过该模型并非专门针对pNETs肝转移患者,外部验证队列药物与基线使用药物有差异。因此,围绕索凡替尼这类兼具抗血管生成与免疫微环境调控机制的多靶点TKI,在pNETs肝转移灶层面构建无创治疗反应预测模型,具有明确的临床意义。且多数研究为单中心、样本量不足100例,使用传统的逻辑回归构建模型,在pNETs人群中的影像组学预测证据仍相对缺乏。

与既往研究相比,本研究的优势在于通过LASSO回归以及标准化流程,进行特征精选与参数优化,进一步提升了模型的预测能力。更在特征维度与模型构建逻辑上有所侧重:一方面,通过三维ROI勾画捕捉肿瘤整体体积特征,减少了二维图像分析可能引入的异质性误差,使提取的影像组学特征更贴近肿瘤的真实生物学特性。另一方面,采用堆叠式两层分类器的集成学习策略,通过异构算法的协同效应有助于提升模型的预测准确性与性能。此外,本研究明确对比了临床模型与影像组学模型的效能差异,更直观地证实了机器学习算法的影像模型在捕获肿瘤微环境复杂信息及提升临床预后评估效能方面的潜在优势。

同时,我们还对患者进行了疗效相关的生存分析。结果表明,无论根据治疗后影像评估的临床应答分层,还是根据机器学习模型预测的应答分层,应答组患者均表现出更长的PFS和更低的疾病进展风险。这提示了早期应答状态与长期获益之间的一致关联。基于影像的临床应答分层一定程度上反映出真实世界的疗效差异,若模型预测可在基线时获得,则有望更早、更客观地识别出潜在获益人群,为治疗决策和风险分层提供前瞻性的依据。

LASSO保留的10个非零系数影像学特征主要刻画肿瘤血供及强化相关表型,同时反映肿瘤负

荷与内部异质性,提示其可能与索凡替尼疗效差异存在关联。GLRLM_LRHGE、GLZLM_HGZE等“高灰度强调”纹理特征及部分一阶统计量(如HU偏度)呈正向系数,提示肝转移灶门静脉期高强化成分更突出、分布更成片时更易应答,这与临床上pNETs及其肝转移灶常见的高血管性强化特征相吻合^[28-29]。从生物学基础看,pNETs与血管生成密切相关,血管内皮生长因子等促血管生成通路在其发生发展中具有重要作用^[30],索凡替尼可同时抑制血管内皮生长因子受体1/2/3与成纤维细胞生长因子受体1介导的血管生成,并通过抑制集落刺激因子-1受体影响肿瘤相关巨噬细胞等免疫微环境,因此“血管生成更活跃、血供更丰富”的病灶更可能对其作用机制敏感^[7]。同时,直方图能量(均匀性)及与邻域结构相关的纹理特征(如NGLDMCoarseness/Contrast)同样呈正向贡献,提示强化更均质的病灶更可能获益。相反,肿瘤体积(SHAPE_Volume)与肿瘤大小在模型中呈负向贡献,提示肿瘤负荷越大,应答概率越低。该结果与临床认识一致,基线肿瘤负荷或肿瘤体积更低通常与更长的PFS或更高的治疗反应概率相关^[31]。此外,峰度和超额峰度等呈负向系数,提示灰度分布高、低强化成分并存更倾向非应答,可能反映肿瘤内部异质性与灌注不均增加^[32]。总体而言,这些特征从血供、强化分布与结构异质性、肿瘤大小等多维度综合反映肿瘤微环境,可为索凡替尼获益人群的无创识别提供一定依据。

本研究存在一定局限性。首先,样本量102例,尽管在罕见的pNETs中相对可观,但对复杂模型仍偏小,外推评估可能存在不稳定性。我们通过Bootstrap方法进行特征筛选并构建模型,并采用重复嵌套交叉验证策略等多种策略降低过拟合并提升模型泛化能力,显示模型具有一定稳健性,后续仍需在独立外部队列中进一步验证模型的稳定性和泛化能力。其次,手工勾画ROI存在主观性与重复性不足,未来可采用半自动或深度学习自动分割提高一致性。虽已统一预处理并校正批次效应,不同设备与扫描参数差异仍可能存在。最后,回顾性多中心研究不可避免存在临床信息缺失及选择偏倚,可能限制临床模型泛化。未来可开展前瞻性研究,完善临床变量并评估与影像组学整合价值,扩展至其他抗血管生成药物或联合方案验证可迁移性,同时探索多时相、多模态及多组学融合以提升模型性能与可解释性。

综上所述,本研究基于增强CT影像组学构建的堆叠机器学习模型,显示出在治疗前以非侵入性方式预测晚期pNETs肝转移患者对索凡替尼的治疗反应的潜力,从而辅助识别可能获益的人群。模型所涉及的血供特征、肿瘤负荷及均质性等影像学指标,或可作为预测疗效与预后的潜在影像标志物,为pNETs肝转移患者的个体化治疗决策提供依据。

作者贡献 阿卜杜喀迪尔江·阿卜杜热合曼: 收集数据,数据分析与绘图,撰写初稿;李玉婷: 收集数据,数据分析;刘容锐、赵传华: 方法学指导,文稿审读和修订;徐建明: 课题设计,学术指导,文稿修订。

利益冲突 所有作者声明无利益冲突。

数据共享声明 本论文相关数据可依据合理理由从作者处获取, Email: abudukadierjiang@163.com

参考文献

- 中国临床肿瘤学会神经内分泌肿瘤专业委员会. 中国胃肠胰神经内分泌肿瘤专家共识(2016年版) [J]. 临床肿瘤学杂志, 2016, 21 (10): 927-946.
- Panzuto F, Ramage J, Pritchard DM, et al. European Neuroendocrine Tumor Society (ENETS) 2023 guidance paper for gastroduodenal neuroendocrine tumours (NETs) G1-G3 [J]. J Neuroendocrinol, 2023, 35 (8): e13306.
- Dasari A, Shen C, Halperin D, et al. Trends in the incidence, prevalence, and survival outcomes in patients with neuroendocrine tumors in the United States [J]. JAMA Oncol, 2017, 3 (10): 1335.
- Sharma A, Muralitharan M, Ramage J, et al. Current management of neuroendocrine tumour liver metastases [J]. Curr Oncol Rep, 2024, 26 (9): 1070-1084.
- Shimata K, Sugawara Y, Hibi T. Liver transplantation for unresectable pancreatic neuroendocrine tumors with liver metastases in an era of transplant oncology [J]. Gland Surg, 2018, 7 (1): 42-46.
- Xu JM. Current treatments and future potential of surufatinib in neuroendocrine tumors (NETs) [J]. Ther Adv Med Oncol, 2021, 13: 17588359211042689.
- Xu JM, Shen L, Bai CM, et al. Surufatinib in advanced pancreatic neuroendocrine tumours (SANET-p): a randomised, double-blind, placebo-controlled, phase 3 study [J]. Lancet Oncol, 2020, 21 (11): 1489-1499.
- Xu JM, Shen L, Li J, et al. Surufatinib in advanced neuroendocrine tumours: Final overall survival from two randomised, double-blind, placebo-controlled phase 3 studies (SANET-ep and SANET-p) [J]. Eur J Cancer, 2025, 222: 115398.
- Ko CC, Yeh LR, Kuo YT, et al. Imaging biomarkers for evaluating tumor response: RECIST and beyond [J]. Biomark Res, 2021, 9 (1): 52.
- Hudson JM, Bailey C, Atri M, et al. The prognostic and predictive value of vascular response parameters measured by dynamic contrast-enhanced-CT, -MRI and-US in patients with metastatic renal cell carcinoma receiving sunitinib [J]. Eur Radiol, 2018, 28 (6): 2281-2290.
- Facchin C, Perez-Liva M, Garofalakis A, et al. Concurrent imaging of vascularization and metabolism in a mouse model of paraganglioma under anti-angiogenic treatment [J]. Theranostics, 2020, 10 (8): 3518-3532.
- Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data [J]. Radiology, 2016, 278 (2): 563-577.
- Zwanenburg A, Vallières M, Abdalah M A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping [J]. Radiology, 2020, 295 (2): 328-338.
- Adili D, Mohetaer A, Zhang WB. Diagnostic accuracy of radiomics-based machine learning for neoadjuvant chemotherapy response and survival prediction in gastric cancer patients: A systematic review and meta-analysis [J]. Eur J Radiol, 2024, 173: 111249.
- Choi H. Response evaluation of gastrointestinal stromal tumors [J]. Oncologist, 2008, 13 (S2): 4-7.
- Solis-Hernandez MP, Fernandez del Valle A, Carmona-Bayonas A, et al. Evaluating radiological response in pancreatic neuroendocrine tumours treated with sunitinib: comparison of Choi versus RECIST criteria (CRIPNET_GETNE1504 study) [J]. Br J Cancer, 2019, 121 (7): 537-544.
- van Vliet EI, Krenning EP, Teunissen JJ, et al. Comparison of response evaluation in patients with gastroenteropancreatic and thoracic neuroendocrine tumors after treatment with [¹⁷⁷Lu-DOTA⁰, Tyr³] octreotate [J]. J Nucl Med, 2013, 54 (10): 1689-1696.
- Lamarca A, Barriuso J, Kulke M, et al. Determination of an optimal response cut-off able to predict progression-free survival in patients with well-differentiated advanced pancreatic neuroendocrine tumours treated with sunitinib: an alternative to the current RECIST-defined response [J]. Br J Cancer, 2018, 118 (2): 181-188.
- Whybra P, Zwanenburg A, Andrearczyk V, et al. The image biomarker standardization initiative: standardized convolutional filters for reproducible radiomics and enhanced clinical insights [J]. Radiology, 2024, 310 (2): e231319.
- Zhang YP, Zhang XY, Cheng YT, et al. Artificial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling [J]. Mil Med Res, 2023, 10 (1): 22.
- 夏婷, 王明亮, 易宗键, 等. 基于增强CT影像组学模型术前预测胰腺神经内分泌肿瘤病理分级 [J]. 中国医学影像技术, 2021, 37 (3): 396-400.
- Pantelis AG, Panagopoulou PA, Lapatsanis DP. Artificial intelligence and machine learning in the diagnosis and management of gastroenteropancreatic neuroendocrine neoplasms: a scoping review [J]. Diagnostics, 2022, 12 (4): 874.
- Ye JY, Fang P, Peng ZP, et al. A radiomics-based interpretable model to predict the pathological grade of pancreatic neuroendocrine tumors [J]. Eur Radiol, 2024, 34 (3): 1994-2005.
- Gu WC, Chen YL, Zhu HB, et al. Development and validation of CT-based radiomics deep learning signatures to predict lymph node metastasis in non-functional pancreatic neuroendocrine tumors: a multicohort study [J]. eClinicalMedicine, 2023, 65: 102269.
- Homps M, Soyer P, Coriat R, et al. A preoperative computed tomography radiomics model to predict disease-free survival in

- patients with pancreatic neuroendocrine tumors [J]. *Eur J Endocrinol*, 2023, 189 (4): 476-484.
- 26 Caruso D, Polici M, Rinzivillo M, et al. CT-based radiomics for prediction of therapeutic response to Everolimus in metastatic neuroendocrine tumors [J]. *La Radiol Med*, 2022, 127 (7): 691-701.
- 27 Ligerio M, Hernando J, Delgado E, et al. Radiomics and outcome prediction to antiangiogenic treatment in advanced gastroenteropancreatic neuroendocrine tumours: findings from the phase II TALENT trial [J]. *BJC Rep*, 2023, 1: 9.
- 28 Worhunsky DJ, Krampitz GW, Poullos PD, et al. Pancreatic neuroendocrine tumours: hypoenhancement on arterial phase computed tomography predicts biological aggressiveness [J]. *HPB*, 2014, 16 (4): 304-311.
- 29 Karmazanovsky G, Belousova E, Schima W, et al. Nonhypervascular pancreatic neuroendocrine tumors: Spectrum of MDCT imaging findings and differentiation from pancreatic ductal adenocarcinoma [J]. *Eur J Radiol*, 2019, 110: 66-73.
- 30 Lauricella E, Mandriani B, Cavallo F, et al. Angiogenesis in NENs, with a focus on gastroenteropancreatic NENs: from biology to current and future therapeutic implications [J]. *Front Oncol*, 2022, 12: 957068.
- 31 Nicolò E, Tarantino P, D' Ecclesiis O, et al. Baseline tumor size as prognostic index in patients with advanced solid tumors receiving experimental targeted agents [J]. *Oncologist*, 2024, 29 (1): 75-83.
- 32 Yuan J, Gong ZG, Liu K, et al. Correlation between diffusion kurtosis and intravoxel incoherent motion derived (IVIM) parameters and tumor tissue composition in rectal cancer: a pilot study [J]. *Abdom Radiol*, 2022, 47 (4): 1223-1231.

(责任编辑: 武建虎, 潘越)

最新录用