

基于影像组学与3D深度学习的多模态模型在重症急性胰腺炎预测中的应用

丁祥林¹, 陈鑫¹, 陈美玉², 沈祎萍², 王玉^{3a}, 殷民月⁴, 赵凯^{3b}, 朱锦舟²

1 苏州永鼎医院消化内科, 江苏苏州 215200

2 苏州大学附属第一医院消化内科, 江苏苏州 215006

3 江苏大学附属金坛医院 a. 肝胆外科, b. 消化内科, 江苏常州 213200

4 首都医科大学附属北京友谊医院消化内科, 国家消化系统疾病临床医学研究中心, 北京 100050

通信作者: 赵凯, zk@jtrmyy.com (ORCID: 0000-0003-0866-6504); 朱锦舟, jzzhu@zju.edu.cn (ORCID: 0000-0003-0544-9248)

摘要: 目的 探索融合影像组学特征、深度学习特征及临床结构化数据的多模态模型在重症急性胰腺炎(SAP)预测中的应用价值, 以期临床早期识别SAP提供更精准的工具。方法 纳入2017年1月1日至2023年12月31日就诊于苏州大学附属第一医院、江苏大学附属金坛医院及苏州永鼎医院的急性胰腺炎(AP)患者, 收集其人口学信息、既往史、病因、入院24 h内实验室检查数据及72 h内影像学资料, 并评估全身炎症反应综合征(SIRS), 同时计算Ranson评分、改良CT严重指数(MCTSI)、床边急性胰腺炎严重度指数(BISAP)和急性胰腺炎风险(SABP)评分。模型构建流程如下: (1) 利用三维CT图像提取并筛选影像组学特征, 基于极端梯度提升(XGBoost)算法建立影像组学分类模型; (2) 采用U-Net对三维CT图像进行语义分割, 随后将分割结果输入3D ResNet50构建深度学习分类模型; (3) 基于XGBoost算法融合上述2种模型的预测值与临床结构化数据, 建立多模态模型。采用变量重要性排序图和局部可解释性图对模型进行可视化解释。符合正态分布的计量资料组间比较采用成组 t 检验; 不符合正态分布的采用Mann-Whitney U 检验。计数资料组间比较采用 χ^2 检验或Fisher精确检验。绘制各模型和已有评分系统的受试者操作特征曲线(ROC曲线), 并计算曲线下面积(AUC), 以评估其性能, AUC间比较采用DeLong检验。结果 共纳入609例符合标准的患者, 其中114例(18.7%)发生SAP。本研究以苏州大学附属第一医院数据作为训练集($n=426$), 江苏大学附属金坛医院和苏州永鼎医院数据作为独立测试集($n=183$)。多模态模型在测试集中的AUC为0.914, 显著高于MCTSI、Ranson评分、BISAP及SABP评分等传统评分系统(AUC分别为0.827、0.675、0.791、0.648), 且较深度学习分类模型(AUC=0.685)及影像组学分类模型(AUC=0.739)性能亦有显著提升(基于DeLong检验的 Z 值分别为-3.23、-4.83、-3.48、-4.92、-4.31和-4.59, P 值均 <0.01)。多模态模型中变量重要性排名前10位的变量依次为胸腔积液、深度学习模型预测值、影像组学模型预测值、甘油三酯、钙离子、SIRS、白细胞计数、年龄、血小板及C反应蛋白, 提示上述变量对模型预测SAP具有重要贡献。结论 本研究基于结构化数据、影像组学特征及深度学习特征, 构建了基于XGBoost算法的多中心SAP预测模型, 其预测性能优于现有传统评分系统及单模态模型。

关键词: 胰腺炎, 急性坏死性; 极端梯度提升算法; 影像基因组学; 多模态成像; 深度学习

基金项目: 国家自然科学基金(82000540); 苏州市消化病临床医学中心(Szleyxzx202101); 苏州市科教兴卫项目(KJXW2019001); 肝脾外科教育部重点实验室开放基金资助课题(GPKF202304); 常州市卫生健康委员会科技项目前沿技术(QY202309); 常州市科技局科技计划项目(CJ20230002); 吴江区科教兴卫项目(WWK202513)

Application of a multimodal model based on radiomics and 3D deep learning in predicting severe acute pancreatitis

DING Xianglin¹, CHEN Xin¹, CHEN Meiyu², SHEN Yiping², WANG Yu^{3a}, YIN Minyue⁴, ZHAO Kai^{3b}, ZHU Jinzhou²

1. Department of Gastroenterology, Suzhou Yongding Hospital, Suzhou, Jiangsu 215200, China; 2. Department of Gastroenterology, The First Affiliated Hospital of Soochow University, Suzhou, Jiangsu 215006, China; 3. a. Department of

Hepatobiliary Surgery, b. Department of Gastroenterology, Jintan Hospital Affiliated to Jiangsu University, Changzhou, Jiangsu 213200, China; 4. Department of Gastroenterology, Beijing Friendship Hospital, Capital Medical University, National Clinical Research Center for Digestive Diseases, Beijing 100050, China

Corresponding authors: ZHAO Kai, zk@jirmy.com (ORCID: 0000-0003-0866-6504); ZHU Jinzhou, jzhu@zju.edu.cn (ORCID: 0000-0003-0544-9248)

Abstract: Objective To investigate the application value of a multimodal model integrating radiomics features, deep learning features, and clinical structured data in predicting severe acute pancreatitis (SAP), and to provide more accurate tools for the early identification of SAP in clinical practice. **Methods** The patients with acute pancreatitis (AP) who attended The First Affiliated Hospital of Soochow University, Jintan Hospital Affiliated to Jiangsu University, and Suzhou Yongding Hospital from January 1, 2017 to December 31, 2023 were included. Related data were collected, including demographic information, previous medical history, etiology, laboratory test data, and systemic inflammatory response syndrome (SIRS) within 24 hours after admission, as well as imaging data within 72 hours after admission, while related scores were calculated, including Ranson score, modified CT severity index (MCTSI), bedside index for severity in acute pancreatitis (BISAP), and systemic inflammatory response syndrome, albumin, blood urea nitrogen and pleural effusion (SABP) score. The model was constructed in the following process: (1) three-dimensional CT images were used to extract and identify radiomics features, and a radiomics classification model was established based on the extreme gradient Boost (XGBoost) algorithm; (2) U-Net is used to perform semantic segmentation of three-dimensional CT images, and then the results of segmentation were imported into 3D ResNet50 to construct a deep learning classification model; (3) the predicted values of the above two models were integrated with clinical structured data to establish a multimodal model based on the XGBoost algorithm. The variable importance plot and local interpretability plot were used to perform visual interpretation of the model. The independent samples *t*-test was used for comparison of normally distributed continuous data between groups, and the Mann-Whitney *U* test was used for comparison of non-normally distributed continuous data between groups; the chi-square test or Fisher's exact test was used for comparison of categorical data between groups. The receiver operating characteristic (ROC) curve was plotted for each model and existing scoring systems, and the area under the ROC curve (AUC) was calculated to assess their performance; the Delong test was used for comparison of AUC. **Results** A total of 609 patients who met the criteria were included, among whom 114 (18.7%) developed SAP. In this study, the data of 426 patients from The First Affiliated Hospital of Soochow University was used as the training set, and the data of 183 patients from Jintan Hospital Affiliated to Jiangsu University and Suzhou Yongding Hospital were used as the independent test set. The multimodal model had an AUC of 0.914 in the test set, which was significantly higher than the AUC of traditional scoring systems such as MCTSI (AUC=0.827), Ranson score (AUC=0.675), BISAP (AUC=0.791), and SABP score (AUC=0.648); in addition, the multimodal model showed a significant improvement in performance compared with the radiomics classification model (AUC=0.739) and the deep learning classification model (AUC=0.685) (the Delong test: $Z = -3.23, -4.83, -3.48, -4.92, -4.31, \text{ and } -4.59$, all $P < 0.01$). The top 10 variables in terms of importance in the multimodal model were pleural effusion, predicted value of the deep learning model, predicted value of the radiomics model, triglycerides, calcium ions, SIRS, white blood cell count, age, platelets, and C-reactive protein, suggesting that the above variables had significant contributions to the performance of the model in predicting SAP. **Conclusion** Based on structured data, radiomic features, and deep learning features, this study constructs a multicenter prediction model for SAP based on the XGBoost algorithm, which has a better predictive performance than existing traditional scoring systems and unimodal models.

Key words: Pancreatitis, Acute Necrotizing; XGBoost Algorithm; Imaging Genomics; Multimodal Imaging; Deep Learning

Research funding: National Natural Science Foundation of China (82000540); Suzhou Clinical Center of Digestive Diseases (Szleyzx202101); Education and Science for Health Development Program, Suzhou Health Committee (KJXW2019001); The Open Fund of Key Laboratory of Hepatosplenic Surgery, Ministry of Education (GPKF202304); Frontier Technologies of Science and Technology Projects of Changzhou Municipal Health Commission (QY202309); Changzhou Municipal Health Commission Science and Technology Project (CJ20230002); Education and Science for Health Development Program, Wujiang Health Committee (WWK202513)

急性胰腺炎(acute pancreatitis, AP)是一种常见的胰腺外分泌炎症性疾病,目前全球年发病率为30/10万~40/10万,且呈逐年上升趋势^[1]。在AP患者中,约20%会发展为重症急性胰腺炎(severe acute pancreatitis, SAP),其死亡率较轻度AP显著升高,达10%~20%^[2]。因此,及时诊断和治疗SAP对于获得良好预后至关重要。目前,病情进展的预测主要依赖传统评分系统,如Ranson评分、急性生理与慢性健康评估(acute physiology and chronic health evaluation, APACHE)、改良CT严重指数(modified CT severity index, MCTSI)、床边急性胰腺炎严重度指数(bedside index for severity in acute pancreatitis, BISAP)等^[3]。

近来,人工智能,尤其是基于卷积神经网络(convolutional neural network, CNN)的深度学习,凭借其强大的计算与学习能力,在消化系统疾病的诊断、病情预测和预后评估中表现优异^[4]。特别是在医学影像分类^[5-6]、语义分割^[7]和目标检测^[8-9]等领域已取得成熟应用。现有研究大多采用二维CNN进行图像处理,但存在局限于单一平面、难以充分捕捉三维空间细节信息等不足^[10]。本研究旨在结合临床特征、影像组学特征及三维深度学习特征,构建一个基于极端梯度提升(extreme gradient boosting, XGBoost)的SAP多模态预测模型,并与传统评分系统进行性能对比。

1 资料与方法

1.1 研究对象 本研究选取2017年1月1日—2023年12月31日就诊于苏州大学附属第一医院、江苏大学附属金坛医院及苏州永鼎医院的AP患者。所有患者的诊断均符合2012年亚特兰大AP分类标准^[11],诊断需满足以下至少2项条件:典型腹痛症状(急性发作且持续腹痛,伴或不伴放射痛);血清脂肪酶或淀粉酶水平达到正常值上限的3倍及以上;影像学检查显示胰腺水肿、渗出或坏死等特征。纳入标准:(1)年龄>18周岁;(2)符合2012年修订版亚特兰大AP分类标准的诊断要求,且为首次诊断的AP患者;(3)SAP诊断标准为伴有持续性器官衰竭超过48h;(4)入院后72h内完成CT检查,且可获取CT平扫图像(DICOM格式)。排除标准:(1)既往有复发性或慢性胰腺炎、胰腺肿瘤或胰腺切除史;(2)合并慢性肝病、慢性肾病或血液系统疾病;(3)既往接受过放疗或化疗;(4)妊娠期患者。

1.2 数据整理

1.2.1 资料收集 (1)人口学信息:年龄、性别;(2)既往史:基础疾病(高血压、糖尿病)、病因学(胆源性、高脂血症

性、酒精性或其他)等;(3)入院后24h内的实验室检查:血常规、肝肾功能、电解质、血脂、C反应蛋白等;(4)入院后72h内的CT特征,如胸腔积液等;(5)入院后72h内腹部平扫CT的DICOM格式文件。同时,根据患者临床数据,评估全身炎症反应综合征(systemic inflammatory response syndrome, SIRS),并计算以下评分:Ranson评分、MCTSI、BISAP和急性胰腺炎风险(systemic inflammatory response syndrome, albumin, blood urea nitrogen and pleural effusion, SABP)评分。

1.2.2 数据清洗 本研究以苏州大学附属第一医院的数据作为训练集,江苏大学附属金坛医院和苏州永鼎医院的数据作为独立测试集。对数据集进行数据清洗以避免数据泄露,使用mice包(版本3.14.0)中的回归树及分类树算法填补缺失数据(缺失比例<20%)。

1.3 研究流程

根据患者住院号收集其入院24h内的临床结构化数据和72h内的腹部平扫CT。以患者是否在72h后发展为SAP为结局变量,分别构建多个模态的预测模型,包括影像组学模型、深度学习模型,以及融合临床结构化数据、影像组学和深度学习的多模态模型。

1.3.1 影像组学模型的建立 (1)特征提取:将上述CT图像导入Slicer(版本5.0.3),手动勾勒病变胰腺区域以建立三维胰腺模型,并利用软件自带的Radiomics模块提取影像组学特征。这些特征共分为7大类,涵盖一阶统计特征、灰度共生矩阵、灰度相关矩阵、灰度游程长度矩阵、灰度尺寸区域矩阵、邻域灰度差矩阵及三维形态特征。(2)一致性分析:本研究采用组内相关系数(intraclass correlation coefficient, ICC)对ROI区域的勾勒和特征提取结果进行组间和组内一致性分析,以确保影像组学模型的稳定性。ICC的评价标准分为低(ICC<0.40)、中(0.40~0.75)、高(ICC>0.75),其取值范围为0~1。本研究由2位医师完成ROI区域的勾勒和特征提取,结果显示组间ICC为0.83,组内平均ICC为0.91,均符合一致性要求。(3)模型建立:通过差异性检验、Pearson相关性分析和LASSO回归对标准化数据进行特征筛选。筛选后的影像组学特征基于XGBoost算法(R CRAN版本1.7.8.1)进行建模,模型最终输出的预测值用于对AP患者进行二分类判断。

1.3.2 深度学习模型的建立 针对SAP的二分类深度学习预测模型的构建过程如下:建立平扫CT的胰腺三维分割模型;基于语义分割得到三维CT图像,构建分类模型。

(1)胰腺三维分割模型:本研究基于U-Net架构构建模型,图像预处理包括将图像裁剪为 228×228 像素大小,采用直方图均衡化增强对比度,同时最大化信息熵以提升图像质量。此外,对像素值进行Z-score归一化处理,即减去图像均值并除以标准差,使图像数据呈现均值为0、标准差为1的分布,从而提高模型对不同图像的适应性。在训练过程中,优化器选用Adam,批量大小设置为20,训练周期(epoch)为200。为了提升模型的泛化能力,训练时引入实时数据增强策略,具体包括随机旋转角度为 $[-30^\circ, 30^\circ]$ 、随机缩放比例为 $[0.8, 1.2]$ 、随机裁剪面积比例为 $[0.8, 1.0]$,以及随机水平或垂直翻转图像。上述预处理和数据增强措施共同保障了模型的鲁棒性。

(2)分割效果分析:采用Dice系数作为分割模型的评价指标,用以衡量图像分割的准确性。该系数通过比较2个集合的相似性,度量分割结果与真实分割的重叠程度,取值为0%~100%,其中0%表示完全不重叠,100%表示完全重叠。Dice系数越接近100%,表明分割效果越准确。

(3)深度学习分类模型构建:模型以胰腺三维CT图像为输入,输出发生SAP的概率。主干网络采用在ImageNet数据集上预训练的3D ResNet50权重进行初始化;学习率采用余弦退火衰减策略,初始学习率设置为0.001;优化器选用Adam,初始学习率为0.001;批量大小为20,训练周期为200,同时根据模型性能动态调整训练周期。三维图像输入3D ResNet50编码器提取特征向量,经过全连接层后输出SAP发生的预测概率。

1.3.3 多模态融合模型的建立 通过上述方案构建基于影像组学和深度学习的SAP二分类模型,获取二者的特征预测值后,将其与临床结构化数据整合成新的特征集矩阵,再采用XGBoost算法融合上述预测值与临床结构化数据,训练并构建最终的多模态融合模型。其中,临床结构化数据的纳入标准为选择贡献度较高的前8个特征。

1.4 统计学方法和评价指标 符合正态分布的计量资料以 $\bar{x} \pm s$ 表示,组间比较采用成组 t 检验;不符合正态分布的计量资料以 $M(P_{25} \sim P_{75})$ 表示,组间比较采用Mann-Whitney U 检验。计数资料组间比较采用 χ^2 检验或Fisher精确检验。 $P < 0.05$ 为差异有统计学意义。采用混淆矩阵(包括真阳性、真阴性、假阳性和假阴性)计算敏感度和特异度;针对各模型和已有评分系统,绘制受试者操作特征曲线(ROC曲线),并计算曲线下面积

(AUC),以评估模型的区分能力,AUC间比较采用Delong检验。

1.5 软件版本及平台 所有统计分析均在R语言(版本4.1.0)和Python(版本3.9)中进行;基于tidymodels平台(版本0.2.0)进行机器学习,采用Keras Python平台(版本3.8.0, TensorFlow 2.8.0)进行深度学习分析。实验硬件条件为Mac mini,配备8核Apple M1处理器(内建GPU)及16 GB内存。

1.6 模型可视化解释 模型可视化采用变量重要性排序图和局部可解释性图。变量重要性排序图展示各变量在预测结局中的贡献程度;局部可解释性图则揭示对预测结局较重要的变量在具体预测中所起的支持或反对作用,以辅助解释模型的预测过程。

2 结果

2.1 临床特征 本研究训练集共纳入426例AP患者,其中72例发展为SAP;测试集共纳入183例AP患者,其中42例发展为SAP。在两个数据集间,性别、年龄、入院CT检查时间等主要特征差异均无统计学意义(P 值均 > 0.05)。在训练集中,非SAP组和SAP组的常见病因均为胆源性(分别为65.7%和56.9%),其次为高脂血症性(12.0%和15.3%)和酒精性(15.5%和19.4%),2组病因分布无统计学差异($P=0.556$)。在测试集中,胆源性同样为常见病因(分别为56.0%和52.3%),2组病因分布亦无统计学差异($P=0.898$)。本研究患者的临床特征详见表1。

2.2 多模态模型纳入的特征 本研究融合临床结构化数据、影像组学模型预测值及深度学习模型预测值,并建立多模态SAP预测模型。基于XGBoost算法,按变量重要性排序纳入的变量依次为:胸腔积液、深度学习模型预测值、影像组学模型预测值、甘油三酯、钙离子、SIRS、白细胞计数、年龄、血小板计数及C反应蛋白(图1)。

2.3 多模态模型的可解释化呈现 随机选取1例非SAP患者,其预测值为0.009,符合非SAP的判断。该预测中,重要性排名前3位的变量分别为无胸腔积液、深度学习模型(-0.168)及影像组学模型(-0.965),其对整体预测的贡献分别为-0.113、-0.035和-0.018,其余变量的总贡献为-0.016。随机选取1例SAP患者,其多模态模型预测值为0.963,符合SAP的判断。该预测中,重要性排名前3位的变量为有SIRS、深度学习模型(0.702)及钙离子(1.8 mmol/L),其对整体预测的贡献分别为+0.285、+0.211和+0.128,其余变量的总贡献为+0.148。多模态模型在随机患者中的可视化解释见图2。

表1 纳入研究的患者临床特征
Table 1 The characteristics of the patients in the study

临床特征	训练集				测试集			
	非SAP组 (n=354)	SAP组 (n=72)	统计值	P值	非SAP组 (n=141)	SAP组 (n=42)	统计值	P值
性别[例(%)]			$\chi^2=2.600$	0.107			$\chi^2=0.258$	0.258
男	277(78.2)	50(69.4)			103(73.0)	29(69.0)		
女	77(21.8)	22(30.6)			38(27.0)	13(31.0)		
年龄(岁)	51.51±17.67	50.99±18.18	$t=0.225$	0.819	52.22±18.04	48.21±17.05	$t=1.318$	0.203
病因[例(%)]			$\chi^2=2.977$	0.395				0.898
胆源性	239(67.5)	41(56.9)			79(56.0)	22(52.4)		
高脂血症性	40(11.3)	11(15.3)			25(17.7)	7(16.7)		
酒精性	52(14.7)	14(19.4)			26(18.4)	10(23.8)		
其他	23(6.5)	6(8.3)			11(7.8)	3(7.1)		
入院CT检查时间(h)	8.05±4.63	7.23±3.19	$t=1.434$	0.152	8.40±4.77	7.91±3.50	$t=0.618$	0.537
血小板计数($\times 10^9/L$)	198.84±69.03	213.33±75.95	$t=-1.498$	0.111	202.43±66.39	215.43±80.19	$t=-0.957$	0.291
白细胞计数($\times 10^9/L$)	13.39±6.93	17.23±6.18	$t=-4.362$	<0.001	13.04±4.79	15.48±6.22	$t=-2.343$	0.008
钙离子(mmol/L)	2.14±0.16	1.98±0.28	$t=4.647$	<0.001	2.12±0.15	1.96±0.35	$t=3.011$	<0.001
甘油三酯(mmol/L)	1.30 (0.83~3.41)	2.15 (1.12~6.14)	$Z=208.606$	0.001	1.41 (0.85~3.31)	3.26 (1.71~13.78)	$Z=74.747$	<0.001
C反应蛋白(mg/L)	54.85 (8.70~140.93)	163.15 (16.22~258.14)	$Z=189.664$	<0.001	49.90 (6.64~144.00)	199.39 (22.98~299.10)	$Z=87.059$	<0.001
SIRS[例(%)]			$\chi^2=61.370$	<0.001			$\chi^2=31.726$	<0.001
无	248(70.1)	15(20.8)			96(68.1)	8(19.0)		
有	106(29.9)	57(79.2)			45(31.9)	34(81.0)		
胸腔积液[例(%)]			$\chi^2=86.861$	<0.001			$\chi^2=30.723$	<0.001
无	236(66.7)	5(6.9)			95(67.4)	8(19.0)		
有	118(33.3)	67(93.1)			46(32.6)	34(81.0)		
MCTSI(分)	2.00 (2.00~4.00)	4.00 (4.00~4.00)	$Z=117.105$	<0.001	2.00 (2.00~4.00)	4.00 (4.00~4.00)	$Z=45.780$	<0.001
Ranson评分(分)	1.00 (0.00~2.00)	2.00 (1.00~2.00)	$Z=169.007$	<0.001	1.00 (1.00~1.00)	2.00 (1.00~2.00)	$Z=86.654$	<0.001
BISAP(分)	1.00 (0.00~2.00)	2.00 (2.00~3.00)	$Z=88.789$	<0.001	1.00 (0.00~2.00)	2.00 (2.00~3.00)	$Z=55.616$	<0.001
SABP评分(分)	2.70 (-3.09~8.92)	10.09 (3.01~25.77)	$Z=176.081$	<0.001	3.89 (-2.80~11.76)	9.00 (1.15~21.31)	$Z=93.856$	0.004

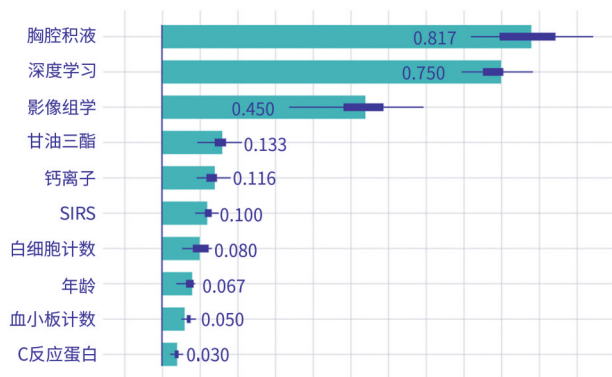


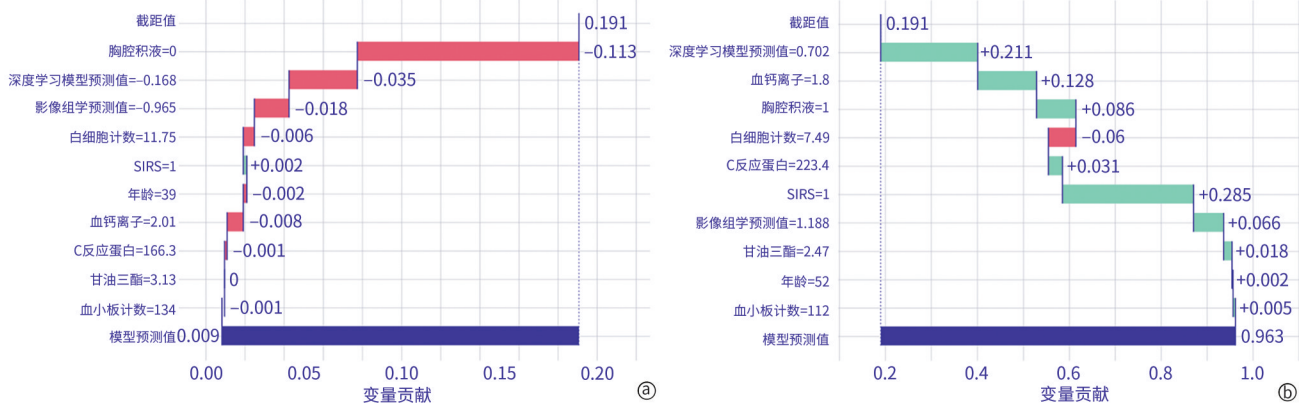
图1 多模态模型中变量重要性排序

Figure 1 Feature importance ranking of the multimodal model

2.4 各模型与现有评分的比较 本研究构建的多模态模型在训练集中的AUC达到0.926(图3), 优于现有评分系统(MCTSI、Ranson评分、BISAP、SABP评分的AUC分别为0.786、0.694、0.840和0.682), 以及影像组学模型和深度学习模型这2个单模态模型(AUC分别为0.692和0.655, P 值均<0.01)(表2)。在测试集中, 多模态模型的AUC为0.914, 同样优于其他模型(AUC分别为0.827、0.675、0.791、0.648、0.685和0.739, P 值均<0.01)。

3 讨论

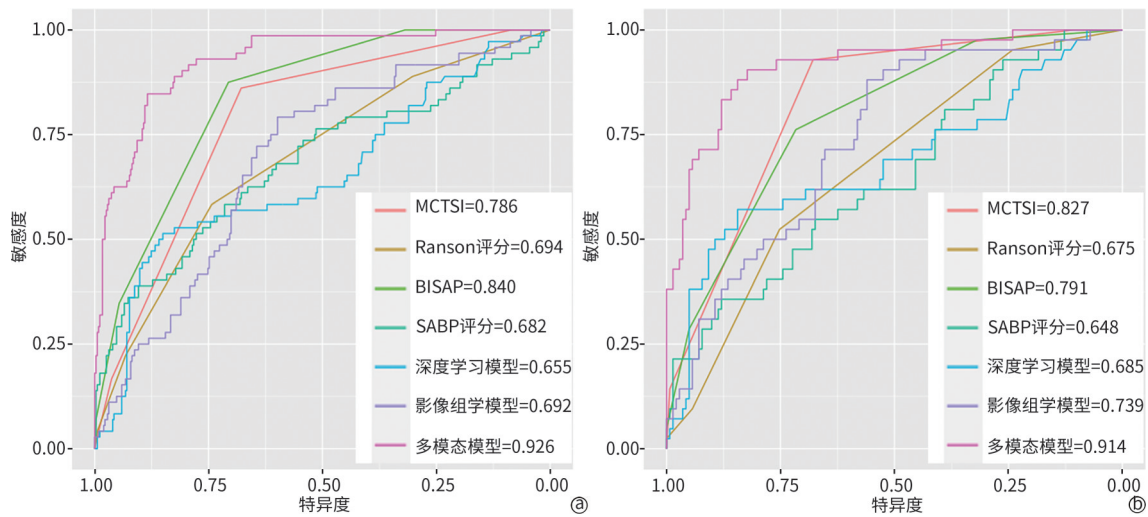
既往传统医学研究多侧重于单一数据模态, 如临床



注:a,非SAP患者;b,SAP患者。

图2 多模态模型在随机患者中的可视觉解释

Figure 2 Visualized explanation of the multimodal model in two random patients



注:a,训练集;b,测试集。

图3 各模型与已有评分系统在两个数据集上的ROC曲线

Figure 3 The ROC curve of each model and the existing scoring system in the two datasets

表2 各个模型与已有评分系统在数据集上的表现

Table 2 Performance of the models and existing scoring systems in datasets

分类	指标	MCTSI	Ranson 评分	BISAP	SABP 评分	深度学习	影像组学	多模态模型
训练集	AUC	0.786	0.694	0.840	0.682	0.655	0.692	0.926
	P值	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
	Z值	-6.11	-6.49	-4.32	-5.76	-7.22	-7.36	
	敏感度	0.861	0.583	0.875	0.583	0.514	0.792	0.847
	特异度	0.678	0.743	0.706	0.715	0.850	0.599	0.884
测试集	AUC	0.827	0.675	0.791	0.648	0.685	0.739	0.914
	P值	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
	Z值	-3.23	-4.83	-3.48	-4.92	-4.31	-4.59	
	敏感度	0.929	0.524	0.762	0.357	0.571	0.881	0.905
	特异度	0.679	0.751	0.716	0.879	0.844	0.560	0.823

注:P值和Z值是多模态模型与其他评分或模型AUC的DeLong检验结果。

结构化数据或影像数据等。但单一模态数据难以从多角度全面反映疾病特征,难以充分表征疾病的复杂性,无法满足精准医学的需求^[12]。近年来,随着医学领域数据类型的多样化和机器学习技术的快速发展,不同模态数据的融合成为可能,多模态模型逐渐成为医学研究的热点。

多模态模型的最大优势在于整合不同来源的信息,克服单模态模型的局限性,从而提供更全面和准确的预测^[13-15]。例如,在AP诊断过程中,单一的临床数据或影像数据可能存在信息不全面的问题,导致误判;而通过融合影像组学特征、深度学习模型和临床结构化数据,可充分发挥各自优势,弥补单模态的不足。其中,临床数据反映患者的生理状态和基础疾病,影像组学展现病灶区域与形态变化,深度学习则通过自动图像分割及特征提取挖掘复杂信息。通过多模态融合,模型能够充分识别潜在的疾病模式,与传统单模态模型相比,显著提升了预测的准确性和临床实用性^[16]。

SAP的早期诊断对患者预后至关重要,因此构建有效的预测模型成为AP研究的重点,以便为早期干预和治疗提供依据^[17]。近年来,人工智能方法在医学领域的应用迅速发展,尤其是机器学习技术的引入,显著提升了医学研究和临床决策的能力。基于该技术构建的SAP预测模型,其准确率较传统模型显著提高。高欣等^[18]利用临床特征(如基线信息、既往史等)和影像学描述(CT、B超),基于XGBoost算法构建了SAP早期预测的机器学习模型,其在测试集中的AUC显著高于Ranson评分、BISAP等传统评分系统。Chen等^[19]在AP严重程度预测中结合临床特征和深度学习特征,先利用单因素和多因素回归模型筛选独立危险因素,再应用CNN提取CT影像特征,将二者融合后通过全连接层获得预测结果。研究发现,融合模型的预测准确度明显优于单一模态模型。Yin等^[20]首次在SAP预测中融合临床特征、影像组学和二维深度学习,构建的多模态模型性能显著优于单模态模型,且具有较高的稳定性及泛化能力。

本研究采用多模态数据融合的方法,充分利用影像组学、深度学习和临床数据的优势,构建高度集成的预测模型,旨在提高SAP的预测精度和可解释性。与传统单一模型相比,本研究的多模态模型通过XGBoost算法融合不同来源的数据,使模型能够从多个维度捕捉疾病的复杂性和多样性。通过多模态融合,模型在训练集和测试集中均达到较高的AUC(分别0.926和0.914),明显优于现有评分系统(如MCTSI、Ranson评分、BISAP和SABP评分)。更重要的是,本研究在影像数据处理中采用三维

模型,突破了传统二维影像处理的局限,更真实地还原了胰腺的形态和结构特征,提升了模型对复杂情况的适应性^[21-22]。此外,通过变量重要性排序和局部可解释性方法,增强了模型的可解释性,使研究人员能够更好地理解模型的预测过程和依据,为其临床应用奠定了基础。

本研究仍存在一定局限性。首先,尽管采用多中心数据,但患者均来自江苏南部地区,样本的地域局限性可能影响模型的外部泛化能力。因此,需引入其他地区的患者数据进行独立外部验证,以进一步评估模型的适用性和鲁棒性。其次,研究对影像数据的质量和清晰度存在一定依赖性,可能会影响模型性能。同时,影像数据的手动标注过程不可避免地存在一定主观性和误差,可能对影像特征提取的稳定性和模型预测的准确性产生潜在影响。再次,研究结局仅采用二分类设计,聚焦于SAP的预测,未对中度SAP进行单独分类。这主要是由于SAP患者的临床管理更复杂且资源需求更高,而XGBoost算法在多分类问题中的可解释性有限。未来研究将探索三分类模型,以更全面地覆盖AP严重程度的分布,为临床决策提供更精准的支持。最后,研究仅收集了入院后短期时间窗内的变量和结局情况,未涉及远期结局(如病死率)。未来研究将延长随访时间,采集更多临床数据和结局信息,以实现AP远期预后的预测。

综上所述,本研究提出的多模态模型在SAP预测中展现出较高的准确性和显著的临床应用潜力。未来研究可进一步优化模型精度,探索更多临床数据源和影像数据维度,以提升模型的泛化能力。此外,随着新型影像技术和生物标志物的发展,未来的多模态模型有望整合更多数据源,进一步提高预测准确性,推动个性化医疗的实现。

伦理学声明:本研究已通过参与单位的伦理委员会审批:苏州大学附属第一医院批号为2022062;苏州永鼎医院批号为202449;江苏大学附属金坛医院批号为2023007。

利益冲突声明:本文不存在任何利益冲突。

作者贡献声明:丁祥林、赵凯、朱锦舟负责课题设计;陈美玉、沈祎萍负责资料分析;陈鑫撰写论文;丁祥林、陈鑫、王玉、殷民月参与收集数据;赵凯、朱锦舟拟定写作思路,指导撰写文章并最后定稿。

参考文献:

- [1] PETROV MS, YADAV D. Global epidemiology and holistic prevention of pancreatitis[J]. Nat Rev Gastroenterol Hepatol, 2019, 16(3): 175-184. DOI: 10.1038/s41575-018-0087-5.

- [2] TRIKUDANATHAN G, YAZICI C, EVANS PHILLIPS A, et al. Diagnosis and management of acute pancreatitis[J]. *Gastroenterology*, 2024, 167(4): 673-688. DOI: 10.1053/j.gastro.2024.02.052.
- [3] MEDEROS MA, REBER HA, GIRGIS MD. Acute pancreatitis[J]. *Jama*, 2021, 325(4): 382. DOI: 10.1001/jama.2020.20317.
- [4] YIN MY, ZHU JD, LIU L, et al. Research advances in machine learning models for acute pancreatitis[J]. *J Clin Hepatol*, 2023, 39(12): 2978-2984. DOI: 10.3969/j.issn.1001-5256.2023.12.034. 殷民月, 朱锦舟, 刘璐, 等. 急性胰腺炎机器学习模型的研究进展[J]. *临床肝胆病杂志*, 2023, 39(12): 2978-2984. DOI: 10.3969/j.issn.1001-5256.2023.12.034.
- [5] XIE F, ZHANG KQ, LI F, et al. Diagnostic accuracy of convolutional neural network-based endoscopic image analysis in diagnosing gastric cancer and predicting its invasion depth: A systematic review and meta-analysis[J]. *Gastrointest Endosc*, 2022, 95(4): 599-609. DOI: 10.1016/j.gie.2021.12.021.
- [6] HAGGENMÜLLER S, MARON RC, HEKLER A, et al. Skin cancer classification via convolutional neural networks: Systematic review of studies involving human experts[J]. *Eur J Cancer*, 2021, 156: 202-216. DOI: 10.1016/j.ejca.2021.06.049.
- [7] LI GW, LIU J, CAO H, et al. Research review of deep learning in colon polyp image segmentation[J]. *J Front Comput Sci Technol*, 2025, 19(5): 1198-1216. DOI: 10.3778/j.issn.1673-9418.2408012. 李国威, 刘静, 曹慧, 等. 深度学习在结肠息肉图像分割中的研究综述[J]. *计算机科学与探索*, 2025, 19(5): 1198-1216. DOI: 10.3778/j.issn.1673-9418.2408012.
- [8] PACAL I, KARABOGA D. A robust real-time deep learning based automatic polyp detection system[J]. *Comput Biol Med*, 2021, 134: 104519. DOI: 10.1016/j.compbiomed.2021.104519.
- [9] YU T, LIN N, ZHANG X, et al. An end-to-end tracking method for polyp detectors in colonoscopy videos[J]. *Artif Intell Med*, 2022, 131: 102363. DOI: 10.1016/j.artmed.2022.102363.
- [10] LIU H, ZHUANG YZ, SONG EM, et al. A 3D boundary-guided hybrid network with convolutions and Transformers for lung tumor segmentation in CT images[J]. *Comput Biol Med*, 2024, 180: 109009. DOI: 10.1016/j.compbiomed.2024.109009.
- [11] BANKS PA, BOLLEN TL, DERVENIS C, et al. Classification of acute pancreatitis: 2012: Revision of the Atlanta classification and definitions by international consensus[J]. *Gut*, 2013, 62(1): 102-111. DOI: 10.1136/gutjnl-2012-302779.
- [12] KLINE A, WANG HY, LI YK, et al. Multimodal machine learning in precision health: A scoping review[J]. *NPJ Digit Med*, 2022, 5(1): 171. DOI: 10.1038/s41746-022-00712-8.
- [13] LIPKOVA J, CHEN RJ, CHEN BW, et al. Artificial intelligence for multimodal data integration in oncology[J]. *Cancer Cell*, 2022, 40(10): 1095-1110. DOI: 10.1016/j.ccell.2022.09.012.
- [14] MOHSEN F, ALI H, EL HAJJ N, et al. Artificial intelligence-based methods for fusion of electronic health records and imaging data[J]. *Sci Rep*, 2022, 12(1): 17981. DOI: 10.1038/s41598-022-22514-4.
- [15] SHAO J, MA JC, ZHANG Q, et al. Predicting gene mutation status via artificial intelligence technologies based on multimodal integration (MMI) to advance precision oncology[J]. *Semin Cancer Biol*, 2023, 91: 1-15. DOI: 10.1016/j.semcancer.2023.02.006.
- [16] HUANG SC, PAREEK A, SEYYEDI S, et al. Fusion of medical imaging and electronic health records using deep learning: A systematic review and implementation guidelines[J]. *NPJ Digit Med*, 2020, 3: 136. DOI: 10.1038/s41746-020-00341-z.
- [17] GLIEM N, AMMER-HERRMENAUC, ELLENRIEDER V, et al. Management of severe acute pancreatitis: An update[J]. *Digestion*, 2021, 102(4): 503-507. DOI: 10.1159/000506830.
- [18] GAO X, LIN JX, WU AR, et al. Application of machine learning model based on XGBoost algorithm in early prediction of patients with acute severe pancreatitis[J]. *Chin Crit Care Med*, 2023, 35(4): 421-426. DOI: 10.3760/cma.j.cn121430-20221019-00930. 高欣, 林嘉希, 吴爱荣, 等. 基于XGBoost算法的机器学习模型在早期预测重症急性胰腺炎中的应用[J]. *中华危重病急救医学*, 2023, 35(4): 421-426. DOI: 10.3760/cma.j.cn121430-20221019-00930.
- [19] CHEN ZY, WANG Y, ZHANG HL, et al. Deep learning models for severity prediction of acute pancreatitis in the early phase from abdominal nonenhanced computed tomography images[J]. *Pancreas*, 2023, 52(1): e45-e53. DOI: 10.1097/MPA.0000000000002216.
- [20] YIN MY, LIN JX, WANG Y, et al. Development and validation of a multimodal model in predicting severe acute pancreatitis based on radiomics and deep learning[J]. *Int J Med Inform*, 2024, 184: 105341. DOI: 10.1016/j.ijmedinf.2024.105341.
- [21] KAMNITSAS K, LEDIG C, NEWCOMBE VFJ, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation[J]. *Med Image Anal*, 2017, 36: 61-78. DOI: 10.1016/j.media.2016.10.004.
- [22] LITJENS G, KOOI T, BEJNORDI BE, et al. A survey on deep learning in medical image analysis[J]. *Med Image Anal*, 2017, 42: 60-88. DOI: 10.1016/j.media.2017.07.005.

收稿日期: 2025-05-27; 录用日期: 2025-06-10

本文编辑: 朱晶

引证本文: DING XL, CHEN X, CHEN MY, et al. Application of a multimodal model based on radiomics and 3D deep learning in predicting severe acute pancreatitis[J]. *J Clin Hepatol*, 2025, 41(10): 2110-2117. 丁祥林, 陈鑫, 陈美玉, 等. 基于影像组学与3D深度学习的多模态模型在重症急性胰腺炎预测中的应用[J]. *临床肝胆病杂志*, 2025, 41(10): 2110-2117.