

论 著

探讨DCE-MR影像组学预测模型鉴别HER2低表达与HER2阳性乳腺癌的应用价值*

陈贤飞* 李蕊蕊 刘旭东
张英博 刘世佳 周云琪
周愉钦 曾雯雯

海南省海南医学院第一附属医院放射科
(海南 海口 570102)

【摘要】目的 探讨MR影像组学模型鉴别HER2低表达与HER2阳性乳腺癌的应用价值。**方法** 回顾性分析我院2018年1月至12月确诊的浸润性乳腺癌233例,其中HER2阳性乳腺癌103例,HER2低表达乳腺癌130例,按8:2随机拆分为训练集186例及测试集47例。基于MR第2期增强图像提取组学特征,数据经过归一化,降维,筛选特征,构建逻辑回归机器学习模型,并于测试集中验证及评估其诊断效能。**结果** 训练集中AUC为0.87,准确率为0.80,敏感性0.89,特异性0.72,PPV0.72, NPV0.89,测试集中AUC值为0.77,准确率0.77,敏感度0.76,特异性0.77, PPV0.73, NPV0.80,基于DCE-MR影像组学的预测模型不仅具有较好的诊断效能,还具有良好的稳定性。**结论** 基于DCE-MR影像组学预测模型鉴别HER2低表达与HER2阳性乳腺癌具有较好的诊断效能,有望为后期临床制定精准化及个性化治疗决策提供参考依据。

【关键词】 HER2低表达; HER2阳性乳腺癌; 影像组学; 机器学习

【中图分类号】 R737.9

【文献标识码】 A

【基金项目】 海南省自然科学基金(822QN464); 海南医学院大学生创新创业训练计划项目(X202111810005); 海南省卫生健康行业科研项目(22A200245)

DOI:10.3969/j.issn.1672-5131.2023.10.027

To Explore the Application Value of DCE-MR Radiomics Prediction Model in Identifying HER2 Low Expression and HER2 Positive Breast Cancer*

CHEN Xian-fei*, LI Rui-rui, LIU Xu-dong, ZHANG Ying-bo, LIU Shi-jia, ZHOU Yun-qi, ZHOU Yu-qin, ZENG Wen-wen.

Department of Radiology, The First Affiliated Hospital of Hainan Medical University, Haikou 570102, Hainan Province, China

ABSTRACT

Objective To explore the application value of DCE-MR radiomics prediction model in identifying HER2 low expression and HER2 positive breast cancer. **Methods** 233 cases of invasive breast cancer diagnosed in our hospital from January to December 2018, including 103 cases of HER2-positive breast cancer and 130 cases of HER2 low expression breast cancer were retrospectively analyzed who were randomly split into a training set (186 cases) and testing set (47cases) according to 8:2. The radiomics features were extracted based on the enhanced images of the second phase of MR. The data were normalized, dimensionally reduced, and features were selected to establish a logistic regression (LR) machine learning prediction model, which was evaluated on the test set. **Results** The AUC of the prediction model established by LR machine learning method in the training group was 0.87, the accuracy rate was 0.80, the sensitivity was 0.89, and the specificity was 0.72. PPV0.72, NPV0.89, in testing set ,AUC value 0.77, accuracy rate 0.77, sensitivity 0.76, specificity 0.77, PPV0.73, NPV0.80, the prediction model based on DCE-MR radiomics not only has good diagnostic efficiency, but also has good stability. **Conclusion** The DCE-MR radiomics prediction model has good diagnostic efficiency in identifying HER2 low expression and HER2 positive breast cancer, and which is expected to provide a reference basis for making accurate and personalized treatment decisions.

Keywords: HER2 Low Expression; HER2 Positive Breast Cancer; Radiomics; Machine Learning.

继肺癌之后,乳腺癌逐渐成为全球最常见的恶性肿瘤,位居女性恶性肿瘤死亡率首位^[1]。人类表皮生长因子受体2(human epidermal growth factor receptor 2, HER2)是乳腺癌的原癌基因,也是HER2靶向治疗的重要靶点。传统HER2靶向药物^[2]曲妥珠单抗与帕妥珠单抗明显改善HER2阳性乳腺癌即免疫组化(immunohistochemistry, IHC)3+,或者IHC2+且原位杂交(in situ hybridization, ISH)检验基因有扩增患者,这类人群占乳腺癌15%-20%。然而40%-50%的乳腺癌患者为HER2低表达^[3],即免疫组化1+或IHC2+且ISH检验基因无扩增,在临床中,这类病人归为HER2阴性乳腺癌,既往研究中^[4]显示这类病人在传统靶向药物治疗中没有获益。近年来以DS-8201a、SYD985为代表的新型抗体靶向HER2药物的研发与探索^[3, 5-6],在HER2低表达乳腺癌患者中显示非常好的靶向治疗效果。这掀起了对HER2低表达乳腺癌研究的热潮。如何精准区分HER2阳性乳腺癌与HER2低表达乳腺癌,是开展精准治疗的前提。影像组学^[7]即从标准的影像图像中高容量提取海量的定量特征(如肿瘤大小、形态、强度、纹理)等,这些特征反映整个肿瘤的异质性,再通过对这些特征深度挖掘、量化分析、模型构建,可间接反应肿瘤癌症表型和肿瘤微环境。因此,本研究评估MR影像组学预测模型鉴别HER2低表达与HER2阳性乳腺癌的应用价值,以为后期临床制定治疗决策,实施精准、靶向、个性化治疗提供参考依据。

1 材料与方法

1.1 临床病理资料 回顾性分析我院2018年1月至12月收治的279例首诊为浸润性乳腺癌的患者。

纳入标准: 免疫组化及FISH检验结果完整; 术前行磁共振动态增强扫描(DCE-MR),图像质量好。排除标准: 术前接受过新辅助、内分泌治疗; 病理结果提示免疫组化HER2蛋白为0表达患者。病理结果从报告中获取。HER2分组标准: HER2阳性组: IHC 3+, 或IHC 2+, FISH基因扩增, HER2低表达组: IHC 1+, 或者IHC 2+, FISH检验基因无扩增。最后纳入233例,其中HER2阳性乳腺癌103例(44.2%), HER2低表达乳腺癌130例(55.8%)。本研究通过了本院医学伦理委员会的批准,豁免知情同意。

1.2 MRI扫描 设备: 3.0T磁共振(型号: GE Discovery 750), 8通道乳腺专用相控阵线圈。体位摆放: 患者俯卧位,双乳自然悬垂于线圈。扫描序列及参数: T₁WI快速梯度回波序列FGRE: TR 640ms, TE7.6s。T₂WI脂肪抑制序列:TR 2587ms, TE 85ms。FOV 320mm × 320mm, 矩阵512×512, 层厚4mm, 层距1mm。增强扫描采用容积成像序列: TR3.9ms, TE1.7ms, FOV360mm×360mm, 层厚4mm, 层距1mm。先扫描蒙片,然后以2.5mL/s速率注射钆喷酸葡胺0.2mL/kg,无间隔连续采集8期增强图像,每期采集时间为53s。

【第一作者】陈贤飞,女,主治医师,主要研究方向:乳腺癌影像诊断。E-mail: chenxianfei258@126.com

【通讯作者】陈贤飞

1.3 肿瘤ROI勾画及影像组学特征提取 肿瘤(region of Interest, ROI)勾画采用ITK-SNAP version 3.6.0软件,由一名具有10年乳腺MR诊断医师对第2期增强图像逐层勾画,避开液化坏死区,每一层尽量包括病灶全部区域1个月后再随机抽取30份MR图像,由同一医师及另外一名具有5年乳腺MR诊断经验的医师按照相同的方法完成第二次ROI勾画,并评估影像组学特征提取的组内及组间一致性、可重复性。

使用开源Python3.8.1平台Pyradiomics模块提取特征,函数库符合(IBSI, image biomarker standardization initiative)标准。对原始图像分别进行^[8]高斯滤波器的拉普拉斯算子变换、小波滤波变换,输入原始图像和ROI图像,根据ROC区提取一阶统计特征、形状特征、纹理特征灰度共生矩阵(gray level cooccurrence matrix, GLCM)、相邻灰度色调差分矩阵(neighbouring grey tone difference matrix, NGTDM)、灰度微分矩阵(grey level dependence matrix, GLDM)、灰度运行长度矩阵(grey level run length matrix, GLRLM)、灰度大小区域矩阵(grey level size zone matrix, GLSZM)等特征类型,共提取1046个特征。

1.4 特征降维、筛选及组学模型构建 将数据按8:2随机拆分为训练集186例及测试集47例。训练集用于特征筛选、模型构

建,于测试集中评估及验证模型。首先对特征进行最大绝对值归一化预处理如图2。特征筛选及降维选择方差阈值筛选器, threshold≥0。选择逻辑回归(logistic regression, LR)分类器建模,参数设置: class_weight(balanced):类平衡参数来平衡训练样本, Penalty: 选择L1正则化方式,参数C=15, C: 正则化强度的倒数, C越小,损失函数会越小,模型对损失函数的惩罚越重,正则化的效力越强,参数会逐渐被压缩得越来越小。为了避免模型过拟合,应用 5折交叉验证。绘制ROC,计算出训练集和验证集AUC值、准确性、敏感度、特异度, F1 Score、阳性预测值、阴性预测值。以上述指标来评估影像组学模型鉴别HER2低表达乳腺癌与HER2阳性乳腺癌的诊断效能。

1.5 统计学方法 组学特征数据处理及模型构建使用达尔文科研平台-医准智能。其余数据分析用SPSS 26.0软件。计数资料用频数和百分比(%)描述,符合正态分布的变量用($\bar{x} \pm s$)描述,对同一医师及不同医师前后两次手动分割病灶所提取的特征采用组内及组间相关性系数ICC进行一致性分析, ICC>0.75表示一致性及可重复性好。采用Mann-Whitney U比较定量数据的差异;对定性变量进行卡方检验, P<0.05表明差异有统计学意义。

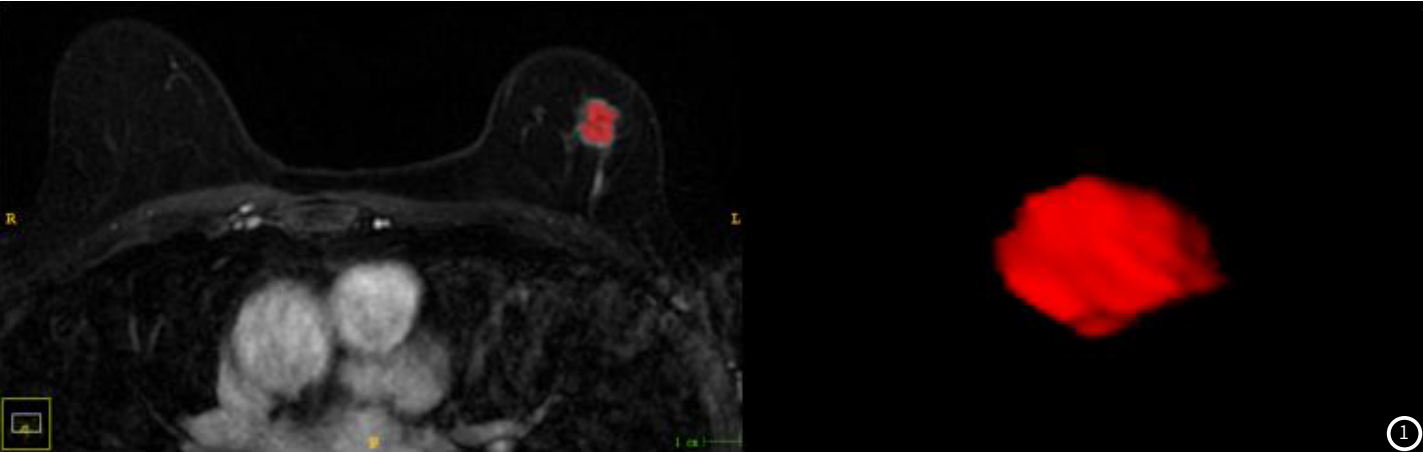


图1 肿瘤感兴趣勾画生成三维ROIs

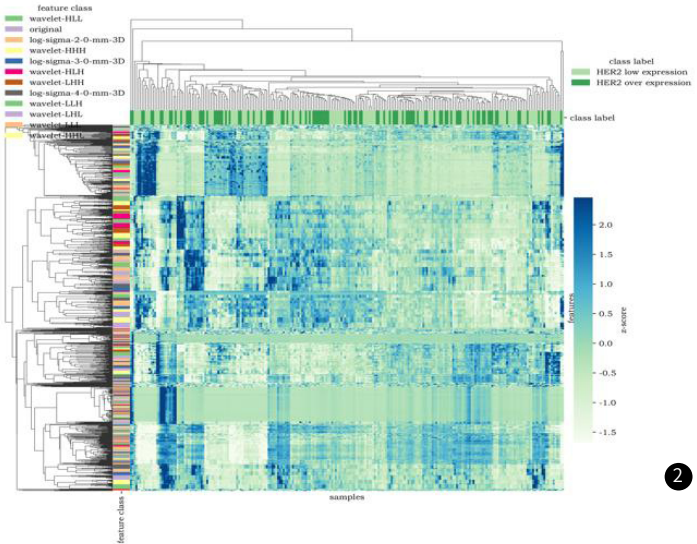


图2 特征标准化预处理

2 结果

2.1 临床病理特征 纳入233例病人中,平均年龄为49.92±10.68,浸润性导管癌伴导管原位癌52例(22.3%),浸润性癌170例(73.0%),其它类型11例(4.7%)(浸润性小叶癌等)。组织学分级 I + II 级109例(46.8%), III级56例(24.0%),未知72例(30.9%)。围绝经期134例(57.5%),绝经99例(42.5%)。年龄、月经状态、组织学分级及HER2阳性乳腺癌、HER2低表达乳腺癌在训练集及测试集两组中差异均无意义(P>0.05,表1)。

表1 训练集与测试集患者临床病理资料比较				
临床病理特征	训练集186(79.8%)	测试集47(20.2%)	统计值	P
年龄	49.52±10.48	51.52±11.39	-1.104	0.270
月经状态				
绝经前	110(59.1%)	24(51.1%)	1.001	0.317
绝经后	76(40.9%)	23(48.9%)		
组织学分级				
I + II	93(66.9%)	17(63.%)	0.157	0.692
III	46(33.1%)	10(37.0%)		
HER2阳性乳腺癌	82(44.1%)	21(44.7%)	0.005	0.942
HER2低表达乳腺癌	104(55.9%)	26(55.3%)		

2.2 影像组学特征提取的一致性及可重复性 同一医师两次提取影像组学特征的ICC值约0.902(95%CI: 0.797, 0.954), 不同医师间ICC为0.825, (95%CI: 0.548-0.927), 一致性及重复性好。经过特征降维及筛选后按特征贡献度排名前10的特征如下: original_firstorder_Skewness, log-sigma-4-0-mm-3D_firstorder_Maximum, wavelet-LLL_firstorder_Skewness, wavelet-HHL_glcml_Correlation, wavelet-LHL_firstorder_Skewness, wavelet-HHL_firstorder_Mean, log-sigma-3-0-mm-3D_firstorder_Maximum, log-sigma-4-0-mm-3D_firstorder_Skewness, log-sigma-2-0-mm-3D_firstorder_Skewness, wavelet-HHH_

firstorder_Median。其中高斯拉普拉斯滤波图像中的特征有4个, 小波变换图像特征5个, 一阶特征1个。

2.3 模型构建及预测性能 经LR机器学习方法建立预测模型在训练集为AUC为0.87(95%CI, 0.82, 0.92), 准确率为0.80, 敏感性0.89 (0.80,0.94), 特异性0.72 (0.63,0.80), PPV0.72(0.62,0.79), NPV0.89 (0.81,0.94), F1 Score0.79, 约登指数0.46, 测试集中AUC值为0.77(0.63,0.91), 准确率0.77, 敏感度0.76 (0.55,0.89), 特异性0.77(0.58,0.89), PPV0.73(0.52,0.87), NPV0.80 (0.61,0.91), F1 Score0.77, 约登指数0.48。结果如图3及表2。

表2 预测模型鉴别诊断HER2阳性乳腺癌及HER2低表达乳腺癌的性能

	AUC值	准确率	敏感性	特异性	PPV	NPV	F1 Score	youden
训练集	0.87(0.82,0.92)	0.80	0.89 (0.80,0.94)	0.72 (0.63,0.80)	0.72(0.62,0.79)	0.89 (0.81,0.94)	0.79	0.46
测试集	0.77(0.63,0.91)	0.77	0.76 (0.55,0.89)	0.77(0.58,0.89)	0.73(0.52,0.87)	0.80 (0.61,0.91)	0.77	0.48

注: AUC值: 受试者工作特征曲线下面积, PPV: 阳性预测值, NPV: 阴性预测值, Youden: 约登指数。

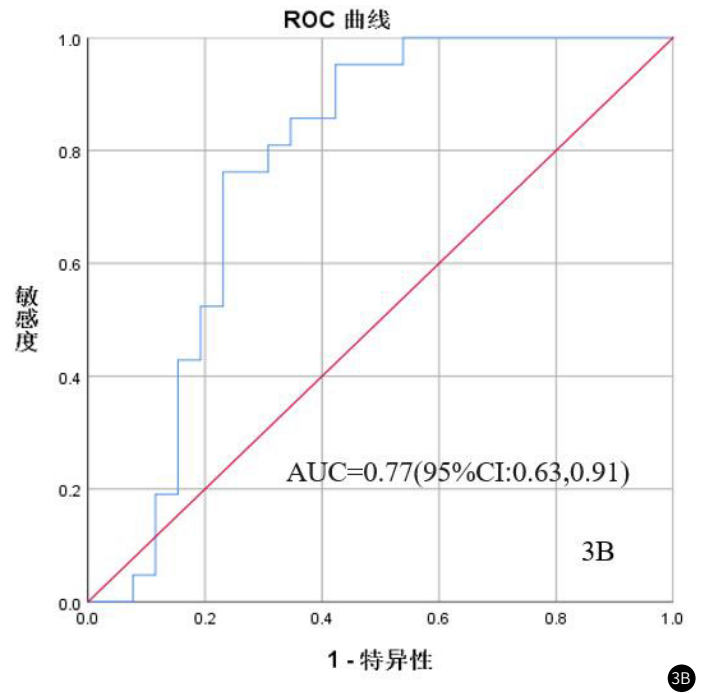
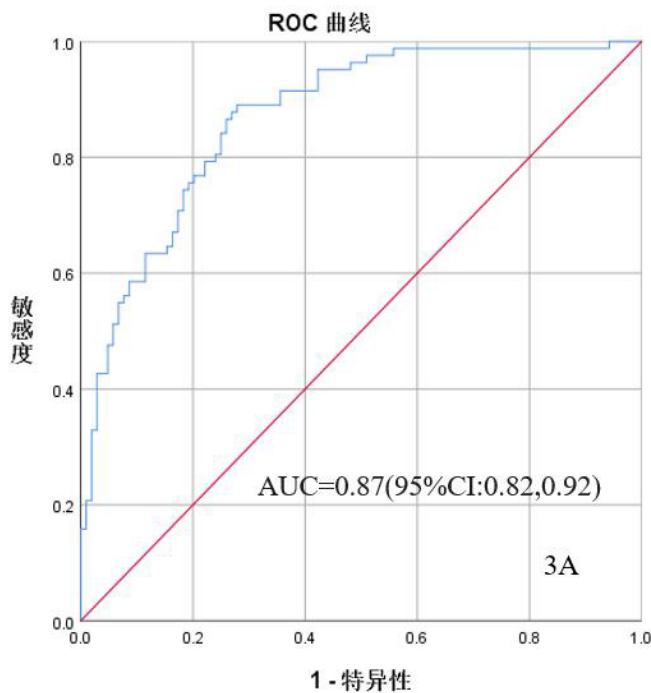


图3 A: 影像组学预测模型在训练集中的ROC曲线 B: 预测模型在测试集中的ROC曲线

3 讨论

HER2^[2]是乳腺癌的原癌基因, 也是抗HER2靶向药物治疗的最成功的靶点。曲妥珠单抗等抗HER2药物的临床应用显著改善了HER2阳性乳腺癌患者的临床结局和生存率, 是乳腺癌治疗的重要突破。此外, 随着抗HER2靶向药物的不断研发, 新的抗体偶联药物也使HER2低表达乳腺癌获益, 临床试验正在进行中^[3, 5-6]。规范治疗是保证相关疗效的前提, 如何做到“合适人群、合适时机、合适方案”对乳腺癌的临床治疗和预后判断至关重要。因此精准、正确检测和评定乳腺癌的HER2蛋白表达和基因扩增状态至关重要。影像组学^[7,9]通过高通量提取图像特征, 突破了传统的基于形态学和半定量分析的医学图像模式, 量化人眼难以识别的组织之间的差异, 深度挖掘和分析隐含在图像中的特征空间数据, 来捕获肿瘤内异质性, 预测肿瘤的分子特征, 在鉴别肿瘤良

恶性、肿瘤的分级、放化疗效果评估、预后及个体化治疗等方面提供信息。机器学习^[10]作为一个热门话题, 在医学领域展示了良好的应用前景。因此本研究探讨MR影像组学预测模型鉴别HER2低表达乳腺癌与HER2阳性乳腺癌的应用价值, 旨在为临床治疗HER2蛋白不同表达的乳腺癌患者选择合适的靶向药物, 制定合适方案、精准治疗提供参考依据。

本研究采用人工分割方法在二维ROIs图像基础上生成了三维ROIs, 3D ROIs保留了肿瘤的所有特征, 并在此基础上提取3D影像特征, 充分反映肿瘤生物学特性和内部异质性。前期有学者^[11]比较2D和3D影像组学特征构建的模型对多器官癌症预测性能, 结果发现相对于2D图像, 3D图像组学预测模型预测性能较好, 部分3D图像组学特征可能与患者的预后相关, 建议将3D图像组学特征

应用到影像组学研究中。另一项研究^[12]也显示,相对于2D影像组学特征,3D影像组学特征显示更好的ICCs,基于三维图像的组学模型在训练集和测试集的预测性能均优于二维图像。而本研究中与鉴别HER2阳性乳腺癌与HER2低表达乳腺癌权重相关的10个特征中就有4个特征为3D图像中获取。

本研究除了在原始图像中提取特征外,还在图像预处理中进行高斯滤波器的拉普拉斯算子变换及小波滤波变换。小波变换^[13]是一种对空间频率的局部分析,可以有效地提取图像中的高、低频信号,更全面、详细地分析纹理变化,这是常规纹理无法获得的。高斯拉普拉斯(laplacian of gaussian, LoG)^[14]是对图像二阶空间导数的各向同性测量,用于突出图像中强度变化迅速的区域。而上述特征均属于高阶特征,能够更可信的解释肿瘤空间的差异。

机器学习可以深度挖掘多维影像信息并将其转化为与肿瘤异质性高度相关的定量特征,从而实现准确诊断和预测^[15],而选择合适的机器学习算法可以提高模型的稳定性和预测效率。logistic回归模型作为临床最常用的机器学习算法,操作简便,适用于二分类变量的模型构建。之前的研究表明^[16],在具有二分类结果的临床预测模型中,LR的性能优于其他机器学习算法。同时,为了提高预测模型的性能,减少模型训练时间,避免模型过拟合,提高模型泛化能力,LR算法中采用了L1^[17]正则化(即LASSO回归),将线性模型系数的H范数作为惩罚项添加到损失函数中,使弱项相关特征对应的系数为0。从而实现影像组学特征的压缩和降维。前期有^[18]学者基于LR分类器构建的乳腺癌良恶性预测模型,训练集中AUC为0.910,准确率0.815,敏感性0.769,特异性0.857,PPV0.833,NPV0.800,测试集中AUC为0.750,准确率0.714,敏感性0.400,特异性0.600,PPV0.667,NPV0.750,本研究中经LR机器学习方法建立预测模型在训练集为AUC为0.87(95%CI, 0.82, 0.92),准确率为0.80,敏感性0.89(0.80,0.94),特异性0.72(0.63,0.80),PPV0.72(0.62,0.79),NPV0.89(0.81,0.94),F1 Score0.79,约登指数0.46,测试集中AUC值为0.77(0.63,0.91),准确率0.77,敏感度0.76(0.55,0.89),特异性0.77(0.58,0.89),PPV0.73(0.52,0.87),NPV0.80(0.61,0.91),F1 Score0.77。基于LR机器学习的预测模型不仅具有良好的诊断性能,而且具有良好的稳定性。

本研究存在以下局限性:(1)本研究为回顾性、单中心研究,样本量小,尚需大样本、多中心外部验证。(2)本研究未分析临床特征,影像组学预测模型的诊断性能是否优于临床特征,尚需进一步研究。

结论:基于机器学习的影像组学预测模型在HER2低表达和HER2阳性乳腺癌的鉴别诊断中具有良好的诊断效能和临床应用价值,有望作为临床决策的辅助工具。

参考文献

- [1]SUNG H,FERLAY J,SIEGEL R L,et al.Global cancer statistics 2020:GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J].CA Cancer J Clin,2021,71(3):209-249.
- [2]DENKERT C,SEITHER F,SCHNEEWEISS A,et al.Clinical and molecular characteristics of HER2-low-positive breast cancer:pooled analysis of individual patient data from four prospective,neoadjuvant clinical trials[J].Lancet Oncol,2021,22(8):1151-1161.
- [3]MODI S,PARK H,MURTHY R K,et al.Antitumor activity and safety of trastuzumab deruxtecan in patients with HER2-low-expressing advanced breast cancer:results from a phase Ib study[J].J Clin Oncol,2020,38(17):1887-1896.

- [4]FEHRENBACHER L,CECCHINI R S,GEYER C E JR,et al.NSABP B-47/NRG oncology phase III randomized trial comparing adjuvant chemotherapy with or without trastuzumab in high-risk invasive breast cancer negative for HER2 by FISH and with IHC 1+ or 2[J].J Clin Oncol,2020,38(5):444-453.
- [5]TARANTINO P,HAMILTON E,TOLANEY S M,et al.HER2-low breast cancer:pathological and clinical landscape[J].J Clin Oncol,2020,38(17):1951-1962.
- [6]RINNERHALER G,GAMPENRIEDER S P,GREIL R.HER2 directed antibody-drug-conjugates beyond T-DM1 in breast cancer[J].Int J Mol Sci,2019,20(5).
- [7]CHO N.Imaging features of breast cancer molecular subtypes:State of the art[J].J Pathol Transl Med,2021,55(1):16-25.
- [8]VAN GRIETHUYSEN J J M,FEDOROV A,PARMAR C,et al.Computational radiomics system to decode the radiographic phenotype[J].Cancer Res,2017,77(21):e104-e7.
- [9]马丽娜,哈宗兰,汪生颜.MRI影像组学在早期浸润性乳腺癌腋窝淋巴结转移诊断的价值[J].中国CT和MRI杂志,2022,20(7):88-91.
- [10]TAGLIAFICO A S,PIANA M,SCHENONE D,et al.Overview of radiomics in breast cancer diagnosis and prognostication[J].Breast,2020,49:74-80.
- [11]XU L,YANG P,YEN E A,et al.A multi-organ cancer study of the classification performance using 2D and 3D image features in radiomics analysis[J].Phys Med Biol,2019,64(21):215009.
- [12]XIE X J,LIU S Y,CHEN J Y,et al.Development of unenhanced CT-based imaging signature for BAP1 mutation status prediction in malignant pleural mesothelioma:Consideration of 2D and 3D segmentation[J].Lung Cancer,2021,157:30-39.
- [13]CHADDAD A,DANIEL P,NAIAZI T.Radiomics evaluation of histological heterogeneity using multiscale textures derived from 3D wavelet transformation of multispectral images[J].Front Oncol,2018,8:96.
- [14]NONATO L G,DO CARMO F P,SILVA C T.GLoG:laplacian of Gaussian for spatial pattern detection in spatio-temporal data[J].IEEE Trans Vis Comput Graph,2021,27(8):3481-3492.
- [15]HEO J,YOON J G,PARK H,et al.Machine Learning-based model for prediction of outcomes in acute stroke[J].Stroke,2019,50(5):1263-1265.
- [16]CHRISTODOULOU E,MA J,COLLINS G S,et al.A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models[J].J Clin Epidemiol,2019,110:12-22.
- [17]MASSAFRA R,BOVE S,LORUSSO V,et al.Radiomic feature reduction approach to predict breast cancer by contrast-enhanced spectral mammography images[J].Diagnostics (Basel),2021,11(4).
- [18]吴佩琪,杨雅俐,周妍璐,等.乳腺良恶性病变鉴别的MRI影像组学预测模型构建:基于随机森林、支持向量机、逻辑回归分类器[J].分子影像学杂志,2021,44(5):764-770.

(收稿日期:2022-12-08)

(校对编辑:谢诗婷)