

基于 ABUS 和机器学习模型对乳腺结节良恶性判断的对比研究

欧 笛,姚劲草,李 伟,李 萌,徐 栋

(中国科学院大学附属肿瘤医院(浙江省肿瘤医院),中国科学院基础医学与肿瘤研究所,浙江 杭州 310022)

摘 要: [目的] 探讨基于乳腺全容积成像 ABUS 机器学习模型对乳腺结节良恶性的诊断效能。[方法] 收集因“乳腺结节”就诊的 164 例患者,均行手术治疗且病理诊断明确,采用回顾性研究,根据患者手术病理结果分为良性组和恶性组,对患者进行乳腺 ABUS 检查后勾画图像,提取影像组学特征,使用四种机器学习方法(LR、RF、SVM 和 DL)进行独立识别测试,检测 DL 的阳性预测值、阴性预测值、阳性似然比、阴性似然比等。[结果] 使用 LR、RF、SVM 和 DL 对乳腺结节进行良恶性判断的 ROC 曲线下面积分别为 0.592、0.627、0.715 和 0.737,选取 ROC 曲线下面积最大的 LD 模型进行进一步诊断效能分析,当阈值=0.6 时,LD 识别的约登指数最大(0.44),对应的阳性预测值为 69.81%、阴性预测值为 77.48%、阳性似然比为 3.80,阴性似然比为 0.48。[结论] 深度学习在目前使用的机器学习方法中对乳腺结节进行良恶性判断有一定临床诊断意义,未来如果训练集样本数量提高,理论上深度学习方法可以获得更好的识别效果。

关键词: 乳腺全容积成像;机器学习模型;乳腺结节

中图分类号: R737.9 **文献标识码:** A **文章编号:** 1671-170X(2021)12-1001-05

doi: 10.11735/j.issn.1671-170X.2021.12.B005

Differential Diagnosis of Benign and Malignant Breast Nodules Based on ABUS and Machine Learning Model

OU Di, YAO Jing-cai, LI Wei, LI Meng, XU Dong

(The Cancer Hospital of the University of Chinese Academy of Sciences(Zhejiang Cancer Hospital), Institute of Basic Medicine and Cancer(IBM), Chinese Academy of Sciences, Hangzhou 310022, China)

Abstract: [Objective] To explore the application of a full-volume breast imaging(ABUS) machine learning model for the differential diagnosis of benign and malignant breast nodules. [Methods] One hundred and sixty-four patients with breast nodules who underwent surgical treatment were divided into benign group(n=60) and malignant group(n=104) according to the surgical and pathological results. After breast ABUS examination, the image omics characteristics were extracted. Four machine learning methods (LR, RF, SVM and DL) were used for diagnosis of breast nodules, the positive predictive value, negative predictive value, positive likelihood ratio and negative likelihood ratio were evaluated. [Results] The areas under the ROC curve(AUC) of LR, RF, SVM and DL was 0.592, 0.627, 0.715 and 0.737 respectively. The LD model with the largest area under the ROC curve was selected for further diagnostic efficiency analysis. Through calculation, when the threshold=0.6, the Yoden index of LD recognition was the largest(0.44), and the corresponding positive predictive value was 69.81%, the negative predictive value was 77.48%, the positive likelihood ratio was 3.80 and the negative likelihood ratio was 0.48. [Conclusion] Deep learning has certain clinical diagnostic significance in differentiating the benign and malignant of breast nodules in the currently used machine learning methods. In the future, if the number of samples in the training set is increased, the deep learning method may achieve better differentiating effect.

Subject words: ABUS; machine learning model; breast nodules

据世界卫生组织(WHO)统计^[1],乳腺癌是女性

基金项目: 国家自然科学基金(82071946);浙江省自然科学基金(LSD19H180001)

通信作者: 徐栋,科主任,主任医师,博士研究生导师,博士;中国科学院大学附属肿瘤医院(浙江省肿瘤医院)超声医学科,浙江省杭州市拱墅区半山山东路 1 号(310022);E-mail: xudong@zjcc.org.cn

收稿日期: 2021-10-08; **修回日期:** 2021-11-05

发病率最高的恶性肿瘤,早发现、早诊断、早治疗成为乳腺癌诊治过程中的重中之重。乳腺全容积成像 ABUS 在乳腺占位成像方面具有巨大潜力,由于其独特的可重复性和可靠性在短期内受到临床医生和患者的一致认可^[2-3],但是,由于 3D ABUS 中的切片

数量很多,这需要超声医学科医生进行长时间的筛查,一些细微的特征可能容易被忽略,从而影响诊断效能。近年来,机器学习对于图像信息的深度挖掘工作取得了长足的进展,本研究采用四种机器学习方法分别对 ABUS 检查中的结节进行回顾性的诊断判断。

1 资料与方法

1.1 一般资料

回顾性收集 2019 年 7 月至 2020 年 1 月因“乳腺占位”在浙江省肿瘤医院乳腺肿瘤外科就诊并行手术的 164 例患者,均为女性,年龄 32~82 岁,平均(55.17±5.42)岁,最终术后病理良性 104 例,恶性 60 例,两组患者在年龄($P=0.118$)和结节大小($P=0.390$)等一般资料方面差异无统计学意义(Table 1)。

Table 1 General patients data of the two groups

Group	N	Effective number of layers*	Age(years old)	Size(cm)
Benign	104	243	55.13±4.73	2.36±1.33
Malignant	60	152	56.28±6.36	2.54±1.21

Notes: *The effective number of layers refers to the number of ultrasound images showing breast nodules on the ABUS cross-sectional image. The minimum diameter of the nodules on the secondary image is> 5 mm

1.2 纳入标准与排除标准

纳入标准:(1)患者均经术后病理明确;(2)术前均未行放、化疗及其他任何治疗;(3)均行术前一周内乳腺自动 ABUS 检查。排除标准:(1)患者一侧乳房有多个占位且良恶性不一;(2)患者拒绝进行 ABUS 检查或因皮肤有创面等问题无法进行全面扫查。

1.3 仪器与采集图像方法

患者平卧位,充分暴露双乳,嘱患者平静呼吸,采用 GE Invenia ABUS 全自动乳腺超声诊断仪进行图像采集,使探头紧贴于乳腺表面,需扫描双侧乳腺的正中位、内侧位、外侧位,分别从底部到顶部进行扫描,一个方向的扫描时间约为 60 s,采集图像的层间距为 0.5 mm。扫查结束后,将图像进行保存,传输至工作站。

1.4 识别方法

1.4.1 ABUS 图像标注和预处理

由一名具有 10 年以上工作经验的医生,使用 IBEX 软件对所有 ABUS 图像进行病灶感兴趣区域(ROI)勾画,每个结节勾画约 3~7 幅图片,输出 MAT 格式 ROI 文件(Figure 1)。

1.4.2 机器学习模型

本研究采用了四种机器学习方法进行独立识别测试,包括支持向量机(SVM)、随机森

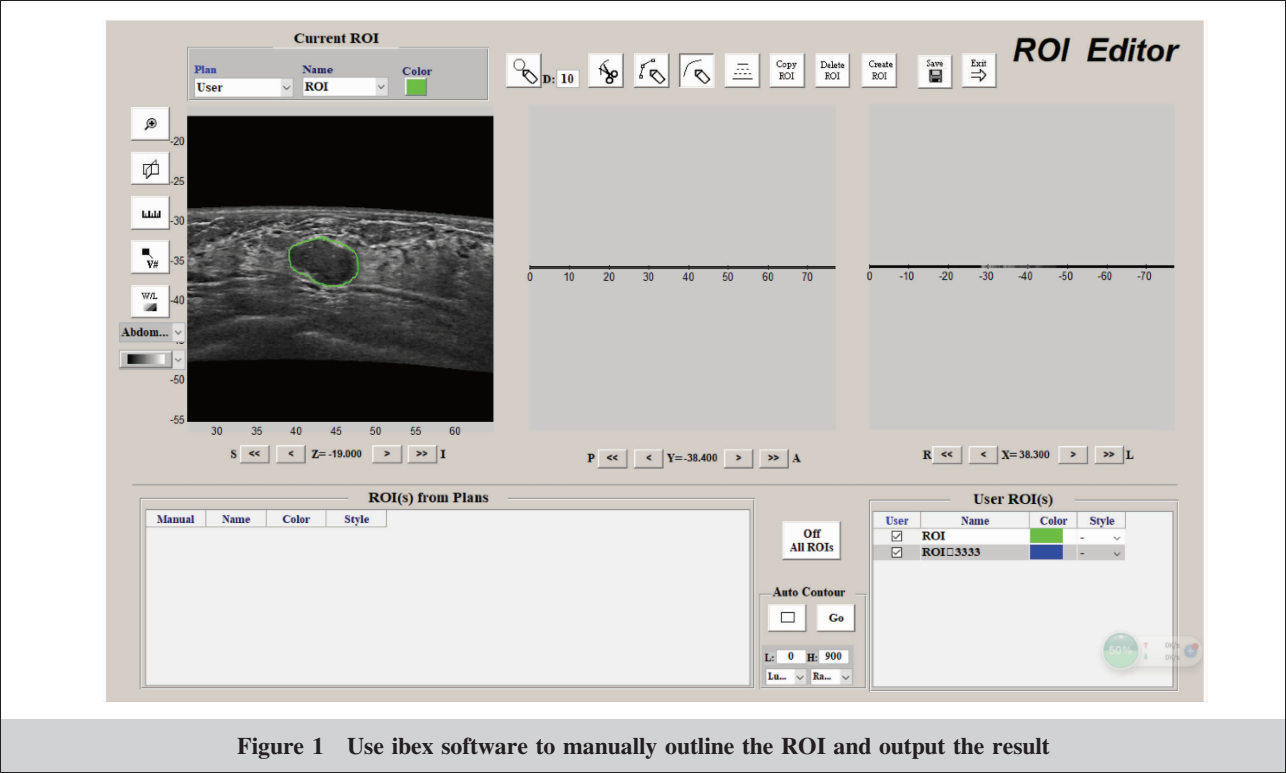


Figure 1 Use ibex software to manually outline the ROI and output the result

林(RF)、逻辑回归(LR)和深度学习(DL)模型 ResNet50。对于前三种传统机器学习方法需要预先提取特征,特征提取使用 IBEX 软件勾画并自动提取 730 种影像组学特征。通过主成分分析(principal component analysis,PCA)将特征维度降低到 35 维。对于深度学习方法,使用了 50 层的标准残差网络(ResNet 50),由于样本数量不足以训练整个深度学习网络,这里使用迁移学习的方法,即固定预训练的卷积层网络参数,重新训练全连接层用于识别。

1.4.3 训练与测试方法

对于传统机器学习方法,使用降维后的特征作为输入进行训练学习。在对 ResNet50 迁移学习时,将图像 ROI 作为输入,并固定卷积层参数,仅对全连接层进行训练。ResNet50 模型在训练时使用了数据增强方法,包括-20%~20%的随机尺度变换、-15 度~15 度的随机旋转、水平/垂直翻转、平移变换、高斯模糊。因为样本分布存在不确定性,且本文使用了多种机器学习模型,为了提高模型评估的客观性和准确性,本研究所有机器学习模型都统一使用留一法(leave-one-out,LOO)进行测试。LOO 每轮测试使用的训练集与初始数据集相比只少了一个样本,使得在不同类型模型在最终判别时,基于 LOO 的测试不受随机样本划分方式的影响,因此评估结果较为客观准确。

1.4.4 模型评估

所有模型识别结果统一归一化到概率 0~1,并通过绘制不同模型识别结果的 ROC 曲线对良恶性判别能力进行衡量,同时计算相应 AUC 值和置信区间。此外还对选择固定阈值下,不同方法的灵敏度、特异度、阳性预测值和阴性预测值进行统计。评估使用的计算机硬件安装了 NVIDIA GeForce TITAN Xp GPU 和 Intel® Core™i7-7740X CPU 4.30 GHz 处理器,

操作系统为 Ubuntu Linux 16.04 包含 Python 3.6.5、Keras 2.1.4 开源库和 TensorFlow 1.9,使用的 Python 语言公开库包括 pymc 库、pylab 库、sklearn 库以及 seaborn 库。

1.5 统计指标

真正(true positive,TP)被模型预测为恶性的恶性样本;真负(true negative,TN)被模型预测为良性的良性样本;假正(false positive,FP)被模型预测为恶性的良性样本;假负(false negative,FN)被模型预测为良性的恶性样本。

2 结 果

2.1 SVM、RF、LR 和 ResNet50 法对肿瘤良恶性判别的灵敏度、特异度及其 ROC 曲线

选取不同的阈值 0.1~0.9,记录 SVM、RF、LR 和 ResNet50 法的灵敏度和特异度(Table 2),随后绘制 ROC 曲线,其中 LR、RF、SVM 和 ResNet50 的曲线下面积分别为 0.592、0.627、0.715、0.737(Figure 2)。

2.2 不同阈值时 ResNet50 识别良恶性的预测能力

选取 ROC 曲线下面积最大的 LD 模型进行进一步诊断效能分析,当阈值=0.6 时,ResNet50 识别的约登指数最大,为 0.44,对应的阳性预测值为 69.81%、阴性预测值为 77.48%、阳性似然比为 3.80,阴性似然比为 0.48(Table 3)。

3 讨 论

超声检查由于其无创、重复性高等优点在乳腺疾病的诊断中有独特的优势,近年来也成为了乳腺疾病筛查的首选检查手段。近年来,ABUS 作为一种新的超声检查技术在临床上的表现极佳,在部分地

Table 2 The sensitivity and specificity of different fixed thresholds(%)

Threshold	LR		RF		SVM		ResNet50	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
0.1	100.00	0	100.00	0	100.00	0	93.55	4.90
0.2	90.32	3.92	96.77	4.90	95.16	1.96	88.71	22.55
0.3	77.42	19.61	79.03	21.57	88.71	20.59	82.26	38.24
0.4	61.29	38.24	66.13	48.04	77.42	41.18	79.03	50.98
0.5	51.61	70.59	48.39	74.51	66.13	70.59	70.97	65.69
0.6	30.65	96.08	29.03	90.20	45.16	89.22	59.68	84.31
0.7	12.90	99.02	19.35	100.00	22.58	97.06	43.55	93.14
0.8	1.61	100.00	3.23	100.00	3.23	100.00	29.03	97.06
0.9	0	100.00	0	100.00	0	100.00	11.29	99.02

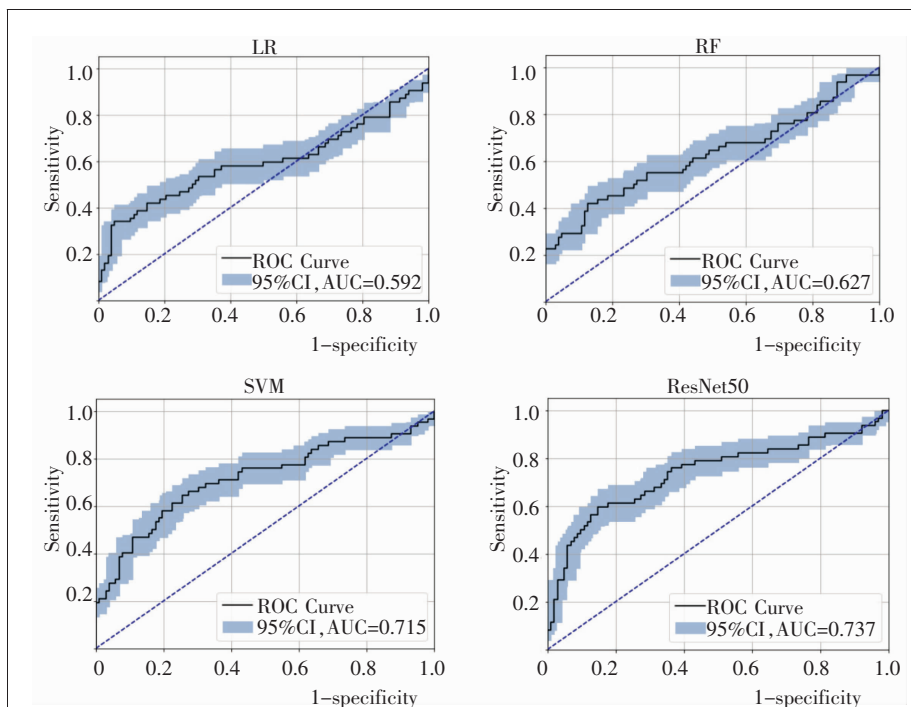


Figure 2 ROC curve of benign and malignant breast nodules detected by 4 methods

Table 3 Predictive results of positive predictive value and negative predictive value identified by ResNet50

Threshold	True positive	False positive	True negative	False negative	Yorden index	Positive predictive value(%)	Negative predictive value(%)	Positive likelihood ratio	Negative likelihood ratio
0.1	58	97	5	4	-0.02	37.42	55.56	0.98	1.32
0.2	55	79	23	7	0.11	41.04	76.67	1.14	0.50
0.3	51	63	39	11	0.21	44.74	78.00	1.33	0.46
0.4	49	50	52	13	0.30	49.49	80.00	1.61	0.41
0.5	44	35	67	18	0.37	55.70	78.82	2.06	0.44
0.6	37	16	86	25	0.44	69.81	77.48	3.80	0.48
0.7	27	7	95	35	0.37	79.41	73.08	6.34	0.61
0.8	18	3	99	44	0.26	85.71	69.23	9.87	0.73
0.9	7	1	101	55	0.10	87.50	64.74	11.52	0.90

区, ABUS 检查已经替代了常规超声检查来进行乳腺疾病筛查, ABUS 检查由于其独特的三维重建功能使肿瘤冠状面重现而使疾病的检出率明显提高^[4], 另外其还减轻了超声科医生的体力劳动, 效率明显提高, 最重要的是, 使用乳腺 ABUS 进行检查可以使乳腺超声图像标准化, 并且可重复读取, 这在大数据时代对临床诊疗具有里程碑意义^[3-4]。

近年来, 针对乳腺结节的人工智能图像研究日益增多, 但绝大部分研究建立在以磁共振图像为依托的图像分析上^[5-7], 但磁共振价格昂贵且可重复性不及超声检查, 因此基于 ABUS 的超声图像实现了

将乳腺结节切割、重建的过程, 也使得基于超声检查的人工智能诊断乳腺结节变为可能。

本研究选取了 164 例患者的 ABUS 图像, 共 164 个乳腺结节, 为了研究的统一性, 本次研究仅纳入乳房单发结节, 提取影像组学特征后使用四种方法进行模型训练及良恶性判别, 其中 ResNet50 效果明显较好, ROC 曲线下面积大于 0.7, 并且通过计算得出约登指数最高在阈值为 0.6 时, 特异度和灵敏度最高。在目前常用的集中机器学习手段中, 深度学习占有一定优势, 虽然在本次研究中, 深度学习的模型

作为一种判读方法, 仍然不能完全替代医生进行临床诊断, 但 ROC 为 0.737 的诊断结果提示可将此模型作为一种临床辅助判读工具, 作为初学者的临床诊断建议工具。

本研究虽然取得了一些初步结果,

但是仍有部分局限。首先, 本次研究共纳入了 164 个结节, 共 395 副 ABUS 图像, 由于进行人工智能分析数据量大结果更有说服力, 也将取得更好的学习效果, 因此在今后的研究中我们也将增加样本量进行研究; 其次, 本次研究的图像勾画全部为人工勾画, 本团队已经在自动分隔训练模型上取得相关进展, 今后也将继续研究, 得到精确的勾画方法, 减少医生的低效劳动。近年来, 研究者们针对基于手持超声的图像分析研究, 也得到一些初步结果, 但部分研究样本量过少, 更为重要的是, 使用手持超声难以像钼靶、磁共振图像将数据标准化^[8-9], 但 ABUS 可进

行图像标准化研究,更有利于今后统一的图像分割和信息提取^[10-11],这是本研究的独特优势。

综上,使用深度学习对基于 ABUS 检查乳腺结节进行良恶性判断有一定临床诊断价值,值得扩大样本量进一步深入研究。

参考文献:

- [1] Ervik M, Soerjornataram I, Ferlay J. Globocan 2012: estimated cancer incidence, mortality and prevalence worldwide in 2012[EB/OL]. 2021-05-06, <https://www.scienceopen.com>.
- [2] 袁娜, 赵巧玲, 王斐倩, 等. 自动乳腺超声诊断系统测量乳腺癌大小及预估乳腺癌 T 分期[J]. 西安交通大学学报(医学版), 2019, 40(3): 133-136.
Yuan N, Zhao QL, Wang FQ, et al. Automatic breast ultrasound diagnosis system to measure the size of breast cancer and estimate the T stage of breast cancer[J]. Journal of Xi'an Jiaotong University (Medical Edition), 2019, 40(3): 133-136.
- [3] 陶承颜, 冉素真, 林芸, 等. 自动乳腺超声诊断系统在乳腺微小结节良恶性鉴别诊断中的价值[J]. 临床超声医学杂志, 2020, 22(8): 605-608.
Tao CY, Ran SZ, Lin Y, et al. Value of automatic breast ultrasound diagnosis system in the differential diagnosis of benign and malignant breast micronodules[J]. Journal of Clinical Ultrasound in Medicine, 2020, 22(8): 605-608.
- [4] 陈翠京, 张彦, 史秀云, 等. 超声自动乳腺全容积成像与多普勒超声成像联合应用鉴别乳腺病灶良恶性[J]. 临床荟萃, 2012, 27(2): 106-107.
Chen CJ, Zhang Y, Shi XY, et al. Combined application of ultrasound automatic breast full volume imaging and Doppler ultrasound imaging to distinguish benign and malignant breast lesions [J]. Clinical Convergence, 2012, 27(2): 106-107.
- [5] 袁成成, 范明, 许茂盛, 等. 基于 DCE-MRI 影像的深度多任务学习联合预测乳腺癌病理信息研究[J]. 传感技术学报, 2021, 34(3): 354-360.
Yuan CC, Fan M, Xu MS, et al. Research on joint prediction of breast cancer pathological information based on DCE-MRI image-based deep multi-task learning[J]. Journal of Transducer Technology, 2021, 34(3): 354-360.
- [6] 陈辛元, 门阔, 唐玉, 等. 基于深度学习自动分割模型乳腺癌放疗临床应用与评价[J]. 中华放射肿瘤学杂志, 2020, 29(3): 197-202.
Chen XY, Men K, Tang Y, et al. Clinical application and evaluation of breast cancer radiotherapy based on deep learning automatic segmentation model[J]. Chinese Journal of Radiation Oncology, 2020, 29(3): 197-202.
- [7] 吕明慧, 周帅, 朱强. 基于深度学习乳腺超声计算机辅助诊断系统研究进展[J]. 中国医学影像技术, 2020, 36(11): 1722-1725.
Lyu MH, Zhou S, Zhu Q. Research progress of breast ultrasound computer-aided diagnosis system based on deep learning[J]. Chinese Medical Imaging Technology, 2020, 36(11): 1722-1725.
- [8] 宋张骏, 贺赛, 王柚熙, 等. 基于深度学习技术的乳腺健康智能检测系统在乳腺肿瘤检测中的应用[J]. 中华乳腺病杂志(电子版), 2019, 13(1): 37-43.
Song ZJ, He S, Wang YX, et al. Application of breast health intelligent detection system based on deep learning technology in breast tumor detection[J]. Chinese Journal of Breast Diseases(Electronic Edition), 2019, 13(1): 37-43.
- [9] Ma W, Zhao Y, Ji Y, et al. Breast cancer molecular subtype prediction by mammographic radiomic features [J]. Acad Radiol, 2019, 26(2): 196-201.
- [10] Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building [J]. Stat Med, 2010, 26(30): 5512-5528.
- [11] Bevilacqua V, Brunetti A, Guerriero A, et al. A performance comparison between shallow and deeper neural networks supervised classification of tomosynthesis breast lesions images[J]. Cog Sys Res, 2018, 53(4): 3-19.