

基于乳腺 X 线图像纹理特征建立机器学习模型在鉴别良恶性乳腺肿块中的价值

黄栌有, 高先聪, 尤传文

【摘要】 目的:探讨基于乳腺 X 线图像的纹理分析建立机器学习模型在鉴别乳腺肿块良恶性中的价值。方法:回顾性搜集经病理证实的 124 个乳腺良性肿块和 139 个乳腺恶性肿块的乳腺 X 线图像。并按照 7 : 3 的比例划将所有病灶随即分为训练集和验证集。使用 MaZda 软件,在 X 线图像上于乳腺病灶内手动勾画 ROI,提取 6 类共 133 个纹理特征,经降维处理后,利用训练集数据得到线性判别分析(LDA)、Logistic 回归(LR)、随机森林(RF)和支持向量机(SVM)共 4 种模型。在验证集中对这 4 种模型进行验证。通过符合率、Kappa 系数和 AUC 值分别评价 4 种模型在训练集和验证集中的表现,并通过 delong 法比较 4 种模型间 AUC 值的差异。结果:RF 模型在训练集和验证集中符合率、Kappa 系数和 AUC 值均高于其它模型;其中,RF 模型在验证集中的符合率为 94.9%、Kappa 系数为 0.896、AUC 值为 0.946,与 LDA 模型、LR 模型间 AUC 值的差异均具有统计学意义($P < 0.05$)。SVM 模型的符合率和 Kappa 系数仅次于 RF 模型;在验证集中,SVM 模型的 AUC 值高于 LDA 和 LR 模型,但差异无统计学意义($P > 0.05$)。结论:基于乳腺 X 线图像纹理特征建立的机器学习模型在鉴别乳腺肿块良恶性中具有一定优势。其中 RF 模型表现出较好的诊断效能,SVM 模型的表现仅次于 RF 模型。

【关键词】 乳腺肿瘤;乳房 X 线摄影术;纹理分析;机器学习;诊断效能

【中图分类号】 R814.41;R737.9 **【文献标志码】** A **【文章编号】** 1000-0313(2021)04-0480-04

DOI:10.13609/j.cnki.1000-0313.2021.04.012

开放科学(资源服务)标识码(OSID):



Application of machine learning model based on the texture features on mammography in the differentiation diagnosis of benign and malignant breast masses HUANG Li-you, GAO Xian-cong, YOU Chuan-wen. Department of Oncology, the Affiliated Suqian Hospital of Xuzhou Medical University, the Suqian People's Hospital of Nanjing Drum Tower Hospital Group, Jiangsu 223800, China

【Abstract】 Objective: The purpose of this study was to investigate the value of machine learning model based on the texture features of mammography image in the differentiation diagnosis of benign and malignant breast masses. **Methods:** Mammography images of 124 benign and 139 malignant breast masses were collected and analyzed retrospectively. All the masses were divided into two groups for training set and verification set according to the proportion of 7 : 3. The region of interest (ROI) was drawn, and 133 texture features of six types were extracted using MaZda software. After the extracted texture features were reduced in dimensionality, four models including linear discriminant analysis (LDA), logistic regression (LR), random forest (RF) and support vector machine (SVM) were obtained based on training set data; and verified in the verification set. The accuracy, Kappa coefficient, and AUC were used to evaluate the performance of the four models in the training and verification sets, and the AUC differences of the four models were analyzed by the delong test. **Results:** The accuracy, Kappa coefficient, and AUC of the RF model in the training and verification sets were higher than those of the other three models. In the verification set, the accuracy of the RF model was 94.9%, the Kappa coefficient was 0.896, and the AUC was 0.946, which was statistically different with the AUC value of LDA model and LR model ($P < 0.05$). The accuracy and Kappa coefficient of SVM model were second only to RF model. The AUC of SVM model was higher than that of LDA and logistic regression

作者单位:223800 江苏,徐州医科大学附属宿迁医院 南京鼓楼医院集团宿迁市人民医院肿瘤科(黄栌有,尤传文),放射科(高先聪)

作者简介:黄栌有(1990-),男,江苏宿迁人,主管技师,主要从事图像后处理工作。

通信作者:尤传文, E-mail: sqycw@126.com

model, but the difference was not statistically significant ($P > 0.05$). Conclusion: The machine learning model based on the texture features of mammography image has certain advantages in differentiating benign and malignant breast masses. RF model showed better classification performance, and SVM model was second only to RF model.

【Key words】 Breast tumor; Mammography; Texture analysis; Machine learning; Diagnostic efficacy

乳腺癌作为发病率最高的女性恶性肿瘤, 目前已成为我国 45 岁以下女性死亡率最高的恶性肿瘤之一^[1]。尽早发现和诊断可以有效提高乳腺癌患者的生存率^[2]。目前, 乳腺 X 线检查在良恶性乳腺肿块的鉴别诊断中已得到广泛的临床应用, 但仅凭对形态学特征的主观分析来鉴别肿块良恶性仍存在较大困难。纹理分析能够提取并定量分析图像中的纹理特征参数, 提供图像中肉眼无法观察到的潜在信息^[3]。而机器学习作为一种人工智能算法, 可以利用纹理分析提供的纹理特征参数建立疾病的分类模型。目前基于机器学习的纹理分析的相关研究已越来越多^[4-6]。本研究拟探讨基于乳腺 X 线图像纹理特征所建立的机器学习模型在鉴别乳腺肿块良恶性中的应用价值。

材料与方 法

1. 研究对象

回顾性搜集 2017 年 1 月—2019 年 12 月在本院行乳腺 X 线检查且经病理证实的 243 例乳腺肿块患者(263 个病灶)的病例资料。所有患者为女性, 年龄 21~82 岁, 平均(50.4±15.41)岁。良性肿块患者 106 例共肿块 124 个, 包括纤维腺瘤 77 例(95 个肿块)、乳腺病 29 例(29 个肿块); 恶性肿块患者 137 例共 139 个肿块, 包括浸润性导管癌 122 例(124 个肿块)、导管

内癌 7 例(7 个肿块)、乳头状癌 4 例(4 个肿块)、黏液腺癌 4 例(4 个肿块)。纳入标准: ①有明确的病理和临床诊断结果; ②肿块内无明显钙化或空洞; ③肿块直径 0.5~3.0 cm。

2. 检查方法

使用 HoLogic Inc. Lorad 公司 Selenia 新型数字化乳腺 X 光机。常规摄取内外侧斜位(mediolateral oblique, MLO)和头尾位(cranial caudal, CC)X 光片。选用 Auto-Filter 曝光模式, 25~30 kV, 75~95 mAs, 图像大小 24 cm×29 cm。

3. 建立机器学习模型

选择图像: 由 2 位具有 5 年以上乳腺 X 线诊断经验的放射科医师采用盲法对每例患者的 MLO 及 CC 图像进行比较, 选取肿块显示较清晰的图像用于纹理分析。当两位医师的意见不同时, 经商议达成一致。将选定的图像以 bmp 格式保存至电脑。

图像分割及特征提取: 采用 MaZda 软件对图像进行后处理。为降低对比度和亮度对图像灰度值可能带来的影响, 在软件中选择 $\mu \pm 3\sigma$ 选项对每帧图像进行灰度归一化处理。由一位诊断医师在每个肿块灶内进行勾画 ROI(图 1、2)。MaZda 软件提供 6 种纹理分析方法: ①灰度直方图; ②灰度绝对梯度; ③游程矩阵; ④灰度共生矩阵; ⑤自回归模型; ⑥小波变换。每个病灶共

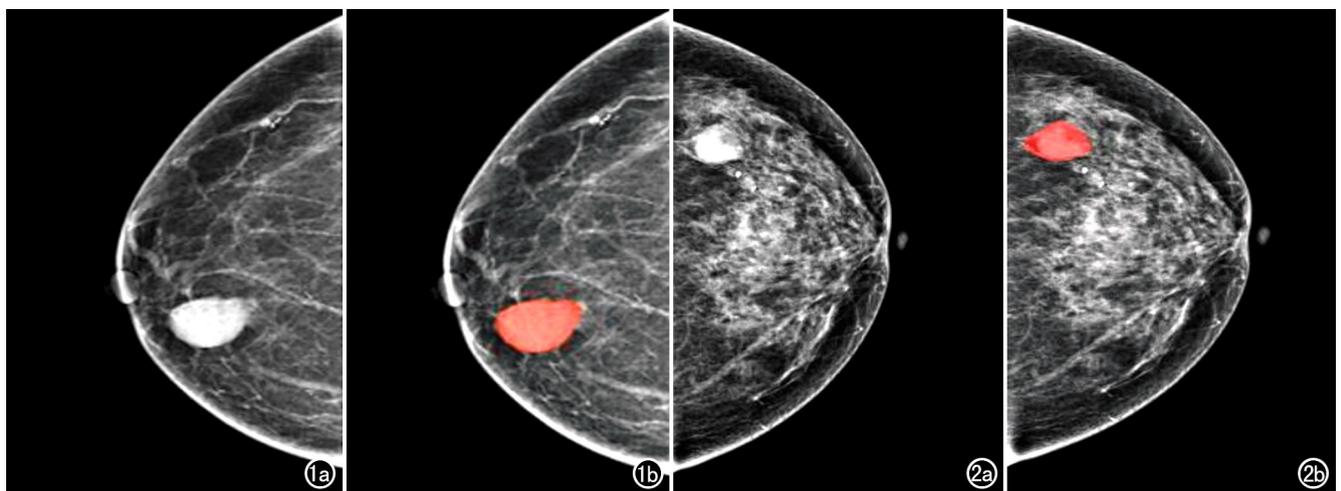


图 1 女, 80 岁, 右乳乳头状癌。a) 右侧乳腺 CC 位 X 线图像, 显示乳腺内椭圆形病灶, 大部分边界较清晰; b) 手动勾画 ROI(红色区域)后的图像。 图 2 女, 52 岁, 左乳纤维腺瘤。a) 左侧乳腺 CC 位 X 线图像, 显示乳腺内不规则形病灶, 部分边界欠清晰; b) 手动勾画 ROI(红色区域)后的图像。

提取 133 个纹理特征, 详见表 1。

表 1 纹理特征列表

分析方法	特征参数
灰度直方图	均值; 方差; 偏态系数; 峰度; 第 1、5、10、50、90 和 99 百分位(Pn)灰度值
灰度绝对梯度	均值; 方差; 偏态系数; 峰度; 非零梯度百分比
游程矩阵	行程长度非均匀性; 灰度级非均匀性; 长行程补偿; 短行程补偿; 行程图像分数
灰度共生矩阵	能量; 对比度; 自相关; 和平方; 和均值; 和方差; 总和熵; 熵; 差方差; 差异熵; 逆差矩矩(参数根据不同距离差分各计算一次; 其中 $d=1, 2, 3, 4, 5$)
自回归模型	噪声标准差 σ ; 模型参数向量
小波变换	低频小波系数能量; 高频垂直、水平和对角线分量小波系数能量(共有 5 级小波分析)

机器学习模型的建立及验证: 利用 R 语言(V3.5.1)的 caret 软件包对数据进行处理和建模。首先, 对提取到的纹理特参数进行初步降维, 然后利用递归特征删减法进行特征选择, 得到最终用于训练模型的纹理特征参数。将处理后的所有数据分为训练集和验证集两组, 按照 7 : 3 的比例进行随机分配。利用训练集数据分别训练获得线性判别分析(linear discriminant analysis, LDA)、Logistic 回归(logistic regression, LR)、随机森林(random forest, RF)和支持向量机(support vector machine, SVM)四种模型。其中 LDA 模型的 k 值为 7 时模型的精度最高, RF 模型在决策树为 500、2 叉树变量树为 2 时模型的精度最高。将训练集获得的 4 种模型分别在验证集中进行验证。

本研究分别通过符合率、Kappa 系数以及 ROC 曲线下面积(area under the curve, AUC)对模型在训练集和验证集中的诊断效能进行评价, 其中训练集采用 10 折交叉验证后所获得的符合率、Kappa 系数和 AUC。通过 Delong 法比较验证集中不同模型之间 AUC 的差异。以 $P < 0.05$ 为差异有统计学意义。

结果

1. 纹理特征的选择结果

通过对 133 个纹理参数进行降维处理, 筛选得到 33 个纹理特征。然后通过特征选择, 最终将 22 个纹理特征分别纳入 4 个模型中进行建模, 其中预测精度最高的 5 个自变量为梯度峰度、高高频小波转换系数 S6、偏斜度、长游程补偿和梯度偏斜度。

2. 机器学习模型效能的评价

四种机器学习模型在训练集和验证集中的符合率、Kappa 系数和 AUC 值见表 2。训练集中以 RF 模型的诊断效能最好, 诊断符合率、Kappa 系数和 AUC 值均高于其它 3 种模型; SVM 模型的诊断效能次之。验证集中以 RF 模型的符合率、Kappa 系数和 AUC 值最高。

表 2 机器学习模型 ROC 分析结果

模型	训练集			验证集		
	符合率	Kappa 系数	AUC	符合率	Kappa 系数	AUC
LDA	0.741	0.479	0.816	0.795	0.587	0.794
Logistic 回归	0.767	0.534	0.823	0.833	0.665	0.833
RF	0.907	0.814	0.934	0.949	0.896	0.946
SVM	0.837	0.673	0.901	0.872	0.741	0.869

通过 Delong 法比较 4 种模型间 AUC 的差异, 结果见表 3。RF 模型与 LDA、LR 模型的 AUC 的差异有统计学意义($P < 0.05$), 与 SVM 模型的 AUC 的差异无统计学意义($P > 0.05$)。其余各模型间 AUC 的差异无统计学意义($P > 0.05$)。RF 模型中各项自变量的重要性排序见图 3, 其中重要性排名前 5 的纹理特征为长游程补偿、梯度峰度、偏斜度、高高频小波转换系数 S6 和峰度。

表 3 四个模型间 AUC 的比较

统计指标	LDA	Logistic 回归	随机森林	SVM
LDA				
Z 值	—	1.746	3.126	1.734
P 值	—	0.081	0.002	0.082
Logistic 回归				
Z 值	1.746	—	2.542	0.847
P 值	0.081	—	0.011	0.397
RF				
Z 值	3.126	2.542	—	1.853
P 值	0.002	0.011	—	0.063
SVM				
Z 值	1.734	0.847	1.853	—
P 值	0.082	0.397	0.063	—

讨论

乳腺 X 线摄影作为乳腺癌首选的检查方法之一, 对钙化灶具有很强的分辨力。但对于软组织结节和肿块, 其诊断能力有限。目前, 已有研究结果显示通过提取乳腺 X 线图像的纹理特征建立预测模型, 可以辅助并提高对乳腺病灶良恶性诊断的准确性^[7-8]。但上述研究采用的模型算法并不统一, 且未对不同建模算法

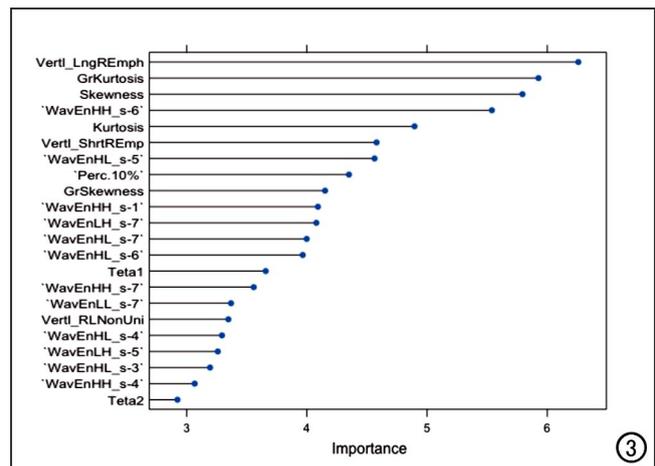


图 3 RF 模型中纹理特征重要性排序的线条图。

的准确性进行对比分析。

本研究探讨了乳腺 X 线图纹理像特征建立机器学习模型在鉴别乳腺肿块良恶性中的价值。研究发现 4 种预测模型的准确率均超过 74%，AUC 值均大于 0.79，表现出良好的准确率和诊断效能。该结果与既往研究相符。在此基础上本研究通过准确率、Kappa 值以及 AUC 值分别对 4 种不同模型算法的表现进行比较。其中准确率是指样本中正确分类的样本数的百分比。Kappa 值反映了分类精度的一致性。AUC 值反应了模型的诊断效能。研究发现 RF 模型在训练集与验证集中准确率、Kappa 值和 AUC 均高于其它模型。其中 RF 模型在验证集中的准确率为 94.8%，Kappa 系数为 0.896，AUC 值为 0.946。其与 LDA 模型、LR 模型间 AUC 值差异具有统计学意义。RF 模型在本研究中表现出高准确性、高度的分类一致性和显著的诊断效能，提示 RF 模型在鉴别乳腺肿块良恶性中最具潜力。RF 模型作为一种用随机方式建立的包含多个决策树的分类器，具有可以处理高维数据、模型泛化能力强、抗过拟合能力强以及鲁棒性等优点^[9]。因此，RF 模型可以基于图像纹理特征参数训练出较为稳定的模型。

SVM 模型的符合率和 Kappa 系数仅次于 RF 模型，验证集中 AUC 值高于 LDA 和 Logistic 回归模型，但差异不具有统计学意义。SVM 模型通过寻找对空间划分的最优超平面，对特征空间进行分类。其具有不依赖数据维度、要求样本量小和分类思想简单等特点^[10]。本研究中，LDA 和 Logistic 回归模型的表现逊于 RF 和 SVM 模型。LDA 模型本质是将初始样本映射到维度更低的样本空间中进行类别输出。LDA 模型对非高斯分布的样本及在样本分类信息依赖方差时表现较差，且容易出现过拟合。Logistic 回归模型作为被广泛应用的一种广义线性回归模型，存在对连续变量异常值较敏感的问题，其预测结果严重依赖数据质量^[4]。但相比属于黑箱算法可解释性差的 RF 和 SVM 模型，Logistic 回归模型的结果具有较强的可解释性。因此，在疾病诊断中仍具有重要价值。近年来，利用机器学习模型预测疾病类型以及预后的研究越来越多。彭涛等^[5]发现 RF 模型在多参数 MRI 诊断前列腺癌中表现最好，SVM 模型次之；李琳等^[11]研究发现 RF 与 SVM 模型对原发性肝癌的预后预测中优于 Logistic 回归模型。本研究结果与之相似。

RF 模型支持对参数重要性的排序。结果显示重要性排在前 5 位的纹理参数分别为长游程补偿、梯度峰度、偏斜度、高高频小波转换系数 S6 以及峰度。游程矩阵反映了具有特定灰度值的像素在既定方向上出

现的频数，长游程补偿值越小表明图像纹理越不均匀。峰度表示的是直方图的锐度。偏斜度为直方图的不对称性的度量方法。高高频小波转换系数反应了图像的细节及清晰度。以上纹理特征均被研究证明在对鉴别肿瘤良恶性中存在价值^[12-13]。

本研究的局限性：①本研究为回顾性单中心研究，且样本量相对较少，研究结果可能存在一定偏倚；② ROI 为人工勾画，因此纹理特征的提取结果可能受到人为因素的影响；③本研究仅讨论了 4 种模型算法，需对更多的模型算法做进一步的对比研究。

综上所述，基于乳腺 X 线图像纹理特征的机器学习模型在鉴别乳腺肿块的良恶性中具有一定优势，其中 RF 模型表现出较好的分类性能，SVM 模型的表现仅次于 RF 模型。

参考文献：

- [1] Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015 [J]. CA Cancer J Clin, 2016, 66(2): 115-132.
- [2] 钟秋子, 荣庆林, 唐玉, 等. I - II 期乳腺癌保乳术后放疗的临床疗效及预后分析[J]. 中华放射肿瘤学杂志, 2018, 27(2): 165-169.
- [3] Andersen MB, Harders SW, Ganeshan B, et al. CT texture analysis can help differentiate between malignant and benign lymph nodes in the mediastinum in patients suspected for lung cancer[J]. Acta Radiol, 2016, 3(3): 669-676.
- [4] 刘哲, 刘小彤, 尹畅畅, 等. T₂WI 机器学习在鉴别高级别胶质瘤和脑单发转移瘤中的应用[J]. 西安交通大学学报(医学版), 2019, 40(5): 794-799.
- [5] 彭涛, 肖建明, 张仕慧, 等. 基于多参数 MRI 及影像组学建立机器学习模型诊断临床显著性前列腺癌[J]. 中国医学影像技术, 2019, 35(10): 1526-1530.
- [6] 邱露, 方向明, 陈宏伟. 人工智能辅助 CT 肺结节良恶性鉴别的研究进展[J]. 临床放射学杂志, 2019, 38(12): 2453-2456.
- [7] 彭文静, 徐凯, 刘正立, 等. 乳腺 X 线图像纹理分析鉴别乳腺小节节良恶性的价值[J]. 中国医学影像学杂志, 2018, 26(12): 890-893, 898.
- [8] 韩英, 何生, 姜增誉, 等. 人工智能在乳腺影像领域的应用现状[J]. 放射学实践, 2019, 34(7): 813-816.
- [9] Qian C, Wang L, Gao Y, et al. In vivo MRI based prostate cancer localization with random forests and auto-context model[J]. Comput Med Imaging Graph, 2016, 52(1): 44-57.
- [10] 李梅, 张伟, 李永忠, 等. 支持向量机神经网络在判别前列腺癌中的应用研究[J]. 四川大学学报(医学版), 2013, 44(4): 666-668.
- [11] 李琳, 杨日东, 王哲, 等. 基于机器学习方法的原发性肝癌患者预后预测研究[J]. 中国数字医学, 2019, 14(3): 34-37.
- [12] 谢元亮, 牡丹, 谢伟, 等. DCE-MRI 纹理分析鉴别宫颈鳞癌与腺癌及预测分级的价值[J]. 放射学实践, 2019, 34(8): 835-840.
- [13] 徐圆, 段钰, 吴晶涛, 等. 基于 CT 增强扫描的纹理分析技术鉴别肾脏透明细胞癌恶性程度的可行性研究[J]. 临床放射学杂志, 2019, 38(9): 1693-1697.

(收稿日期: 2020-06-29 修回日期: 2020-09-22)