

基于双侧乳腺图像“共用”阈值分割的乳腺癌近期发病预测

张红军, 闫士举

上海理工大学医疗器械与食品学院, 上海 200093

【摘要】目的:基于“共用”阈值分割双侧乳腺图像的高密度区域并采用决策树分类算法对乳腺筛查者是否患癌进行预测分类,以提高女性乳腺癌近期发病风险预测精度。**方法:**采用185例女性筛查者的全数字化乳腺钼靶X线摄影(FFDM)Craniocaudal位图像。首先对双侧乳腺图像整体的灰度分布进行统计,采用一种“共用”阈值分割双侧乳腺图像的高密度区域,然后分别从乳腺区及其高密度区提取不对称特征与最大值特征。基于两类特征分别构建2个决策树弱分类器,再将两个弱分类器组合构建一个强分类器。乳腺癌近期发病风险预测性能采用留一法进行验证。**结果:**基于“共用”阈值法,所提取的图像特征中有7个特征的曲线下面积(AUC)>0.6,基于单独阈值法,所提取的图像特征中只有2个特征的AUC>0.6。采用强分类器,基于“共用”阈值法得到的AUC=0.970 1±0.014 8,预测准确率为95.68%;基于单独阈值法得到的AUC=0.982 4±0.008 9,预测准确率为93.51%。**结论:**采用“共用”阈值分割双侧乳腺图像的高密度区域更有利于从高密度区提取更多具有高预测力的图像特征;采用强分类器可显著提高乳腺癌近期发病预测精度。

【关键词】乳腺癌;近期发病预测;高密度区;“共用”阈值;决策树分类;医学图像处理

【中图分类号】TP751; R737.9

【文献标志码】A

【文章编号】1005-202X(2017)08-0820-05

Near-term breast cancer risk prediction based on bilateral breast image segmented with "shared" threshold method

ZHANG Hongjun, YAN Shiju

School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

Abstract: Objective To improve the accuracy of near-term breast cancer risk prediction by segmenting the bilateral breast image dense region with "shared" threshold method, and predicting whether women have breast cancers based on decision tree classification algorithm. **Methods** The full-field digital mammography images of Craniocaudal position obtained from 185 women were selected in this study. After the whole gray distribution of bilateral breast image was calculated and bilateral breast image dense region was segmented with "shared" threshold segmentation method, the asymmetric features and maximum features were extracted from breast region and dense region, separately. Two weak decision tree classifiers were built based on two types of features, and then were combined into a strong classifier. Leave-one-case-out method was used to validate the performance of near-term breast cancer risk prediction. **Results** The number of extracted features with the area under the curve (AUC) larger than 0.6 was 7 for "shared" threshold method, and 2 for single threshold method. With strong classifier, "shared" threshold method achieved an AUC of 0.970 1±0.014 8 and a prediction accuracy of 95.68%; the single threshold method achieved an AUC of 0.982 4±0.008 9 and a prediction accuracy of 93.51%. **Conclusion** Using "shared" threshold to segment bilateral breast image dense region is more advantageous to extract image features with high predictive power from dense region. Using strong classifier can significantly improve the accuracy of near-term breast cancer risk prediction.

Keywords: breast cancer; near-term risk prediction; dense region; "shared" threshold; decision tree classification; medical image processing

【收稿日期】2017-04-25

【作者简介】张红军, 硕士在读, 研究方向: 医学图像处理, E-mail: 15836086606@163.com

【通信作者】闫士举, 博士, 副教授, 研究方向: 医学图像处理, E-mail: yanshj99@aliyun.com

前言

乳腺癌是女性常见的恶性肿瘤^[1-2], 严重危害女性身心健康。科学研究表明, 构建科学的乳腺癌风险评估模型是一种有效的乳腺癌预防途径。自1989年Gail报告第一个乳腺癌风险预测模型以来, 国内外

已经建立了近30个类似的风险评估模型^[3-8],其中包括Claus模型、Myriad模型、BRCAPRO模型及Tyrer-Cuzick模型等。这些模型多需配合采用基因检测技术来获取相关数据,成本较高且无法对个体女性近期患癌的风险进行评估;对个体女性而言,很多情况下这些评估模型并不具备临床可接受的预测精度^[9]。而基于图像特征的乳腺癌近期发病风险预测模型无需采用基因检测技术且可对个体女性近期患癌的风险进行预测。根据现有文献,研究者们都是对左、右乳腺图像分别确定不同的阈值来分割乳腺高密度区域^[10-13]。此单独阈值法分割乳腺高密度区不利于提取的图像特征反映左、右乳腺的不对称特性,而不对称性对乳腺癌的发病预测具有重要参考价值^[14-15]。本研究拟对左、右乳腺图像灰度的整体分布进行统计分析,采用一种“共用”阈值法来分割双侧乳腺高密度区域并采用LADTree分类算法对乳腺筛查者是否患癌进行预测分类。

1 资料与方法

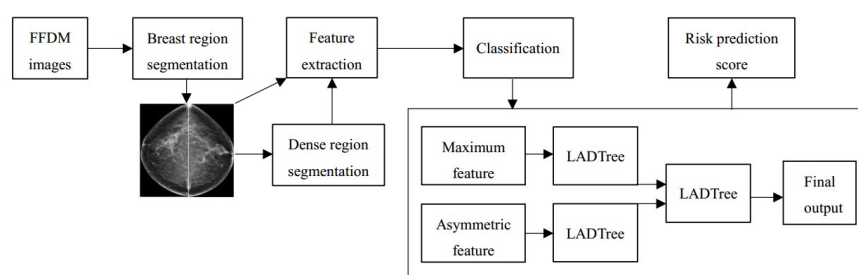
1.1 图像资料

本研究采用185例匿名女性筛查者的全数字化乳

腺钼靶X线摄影(Full-Field Digital Mammography, FFDM)图像。图像来自匹兹堡大学医学中心的临床数据库。所有图像均采用Hologic Selenia FFDM系统(Hologic Inc., Bedford, MA, USA)采集。每位筛查者都分别做了两次FFDM检查。在第一次检查中,185例筛查者都被放射科医生诊断为阴性(未患癌),而在第二次检查中,有89例筛查者被确诊为阳性(患癌),96例被确诊为阴性。对于每一位筛查者,做一次FFDM检查可以得到4幅乳腺图像,包括两幅左、右乳腺CC(Craniocaudal)位图像和两幅MLO(Midolateral Oblique)位图像。由于对MLO位图像进行乳腺区域分割时,胸壁肌肉与乳腺区域边缘检测效果不理想,本研究只采用了每位筛查者第一次检查的乳腺CC位图像。

1.2 方法

本文首先对FFDM图像进行乳腺区域分割,并采用两种取阈方法分割双侧乳腺高密度区域,其次分别从FFDM图像中的全乳腺区及高密度区提取不对称特征和最大值特征,然后构建LADTree分类模型对乳腺筛查者是否患癌进行预测分类,最后依据模型输出的风险预测得分评估模型的预测性能,本文的研究流程如图1所示。



FFDM: Full-field digital mammography; LADTree: Logistic alternating decision tree

图1 乳腺癌近期发病预测研究流程图

Fig.1 Flow diagram of near-term breast cancer risk prediction

1.2.1 乳腺区域分割 乳腺区域分割的步骤可简述为:①绘制灰度直方图并采用迭代阈值法检测出乳腺组织与图像背景之间的最优边界曲线,之后移除图像背景区域;②在分割后的乳腺图像上执行形态学腐蚀操作,移除皮肤区域。乳腺区域分割的详细步骤可参考文献^[11-12]。

1.2.2 乳腺高密度区域分割 乳腺灰度图像中灰度值大于设定的分割阈值区域称为乳腺高密度区域。乳腺图像的高密度区域能够反映乳腺组织的致密度,而乳腺组织的致密度对乳腺癌的诊断具有重要参考意义。传统的单独阈值法是对左、右乳腺图像分别确定不同的阈值来分割乳腺高密度区域,本研究拟

对双侧乳腺图像灰度的整体分布进行统计分析,采用一种“共用”阈值法来分割双侧乳腺高密度区域。

(1)“共用”阈值法分割双侧乳腺高密度区的原理。合并统计双侧乳腺图像整体的灰度分布,取双侧乳腺图像整体的灰度中值作为分割阈值,双侧乳腺图像中灰度值大于该分割阈值的区域称为乳腺高密度区域。

(2)“共用”阈值法的优势。传统的单独阈值法是对左、右乳腺图像分别采用不同的阈值进行高密度区分割。与之相比,“共用”阈值法则采用同一个标准提取左、右乳腺图像的高密度区,更有利于反映左、右乳腺图像的灰度分布差异,从而更有利于使提

取的图像特征反映左、右乳腺的不对称特性,而不对称性对于乳腺癌的发病预测具有重要参考价值^[14]。此外,采用“共用”阈值法分割双侧乳腺高密度区可能更有利于从乳腺高密度区提取更多有价值的图像特征,从而有益于提高乳腺癌近期发病风险预测精度。

1.2.3 特征提取 本研究分别采用单独阈值法和“共用”阈值法来分割双侧乳腺高密度区,并采用计算公式分别从FFDM图像中的乳腺区及其高密度区提取图像特征。提取的图像特征包括:①图像的灰度统计特征:包括均值、标准差、偏度、峰度等^[10,14,16];②图像的纹理特征:包括分别从灰度共生矩阵、灰度游程矩阵、灰度区域大小矩阵及局部灰度差分矩阵中计算的纹理特征^[17];③基于分形维数的特征^[10,14,16];④从分割后的乳腺区域的灰度直方图中提取的特征^[14,16];⑤将分割后的乳腺区域变换成局部像素值波动图像提取的特征等^[14,16]。分别采用两种取阈方法对乳腺图像进行高密度区提取,然后采用计算公式从图像中的全乳腺区及提取的高密度区分别提取了152个图像特征。

同一例筛查者左、右乳腺图像对应特征值差的绝对值称为不对称特征,对应特征值中较大者称为最大值特征。本研究分别采用不对称特征^[14-15]及最大值特征^[11]作为筛查者的乳腺图像特征值并采用min-max方法^[14]将两类特征分别标准化到0和1

之间。

1.2.4 分类 本文采用WEKA软件集成的决策树(Logistic Alternating Decision Tree, LADTree)分类算法对乳腺筛查者是否患癌进行预测分类。LADTree是使用LogitBoost策略的多类别交替决策树。LogitBoost是一种改进的Boost算法,其优势在于能将若干个弱分类器(Weak Classifier)组合构建一个分类准确率较高的强分类器(Strong Classifier)。本文先基于不对称特征(Asymmetric Feature)和最大值特征(Maximum Feature)构建两个LADTree弱分类器,然后再将这两个弱分类器组合构建一个强分类器对乳腺筛查者是否患癌进行预测分类。为了对分类方法的预测性能进行客观比较,本研究采用留一法对分类器进行验证。本文用受试者工作特征曲线(Receiver Operating Characteristic, ROC)、曲线下面积(Area Under the Curve, AUC)、预测准确率、敏感度及特异度指标评估分类器的预测性能。

2 结果

如表1所示,基于“共用”阈值法共有7个特征的AUC>0.6,最大AUC=0.639 6±0.040 3。而基于单独阈值法只有2个特征的AUC>0.6,最大AUC=0.620 9±0.040 7,不难发现采用“共用”阈值分割双侧乳腺高密度区更有利于从高密度区提取更多具有高预测力的图像特征。

表1 AUC>0.6的图像特征
Tab.1 Image features with area under the curve (AUC) larger than 0.6

Parameter	Single threshold method-based features		"Shared" threshold method-based features						
	F86	F112	F90	F3	F86	F4	F89	F_105	F_4
Region of interest	Dense	Dense	Dense	Entire	Dense	Entire	Dense	Dense	Entire
Feature class	Asymmetric	Asymmetric	Asymmetric	Asymmetric	Asymmetric	Asymmetric	Asymmetric	Maximum	Maximum
AUC	0.620 9±	0.603 8±	0.639 6±	0.626 8±	0.624 6±	0.620 1±	0.605 8±	0.619 3±	0.606 0±
	0.040 7	0.041 2	0.040 3	0.040 5	0.040 6	0.040 8	0.041 1	0.040 9	0.041 1

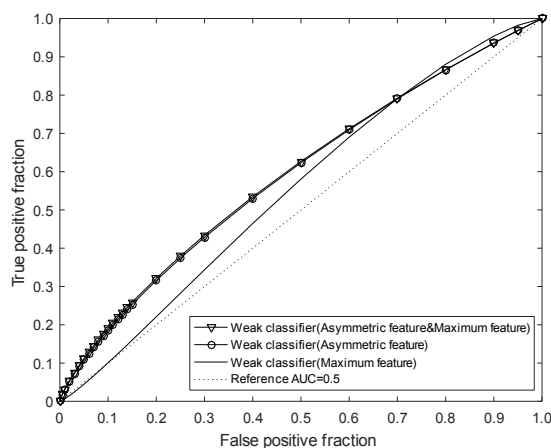
如表2所示,采用弱分类器,基于“共用”阈值法所得的最大AUC=0.679 4±0.040 1,基于单独阈值法所得的最大AUC=0.593 4±0.041 1。对于两种取阈方法来说,由弱分类器组合构建的强分类器的预测性能显著高于所有弱分类器的预测性能。图2、图3分别是两种取阈方法采用弱分类器、强分类器所得ROC曲线对比图。

表3所示,由弱分类器组合构建的强分类器的预测精度显著高于所有弱分类器的预测精度。采用强分类器,基于“共用”阈值法所得预测准确率为95.68%(177/185),敏感度为98.88%(88/89),特异度为92.71%(89/96)。基于单独阈值法所得的预测准确率为93.51%(173/185),敏感度为91.01%(81/89),特异度为95.83%(92/96)。

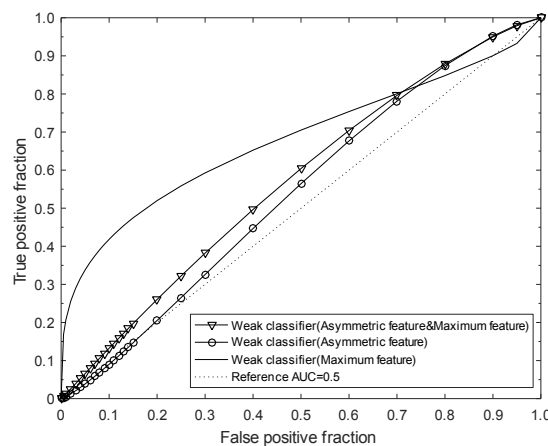
表2 两种取阈方法所得AUC对比

Tab.2 Comparison of AUC obtained with two threshold methods

AUC	Weak classifier based on asymmetric feature and maximum feature	Weak classifier based on asymmetric feature	Weak classifier based on maximum feature	Strong classifier
AUC obtained with single threshold method	0.593 4±0.041 1	0.590 3±0.041 2	0.553 0±0.041 8	0.982 4±0.008 9
AUC obtained with "shared" threshold method	0.571 3±0.041 4	0.541 9±0.041 7	0.679 4±0.040 1	0.970 1±0.014 8



a: ROC curve obtained with single threshold method



b: ROC curve obtained with "shared" threshold method

图2 两种取阈方法采用弱分类器所得ROC曲线对比

Fig.2 Comparison of receiver operating characteristic (ROC) curve obtained with two threshold methods using weak classifier

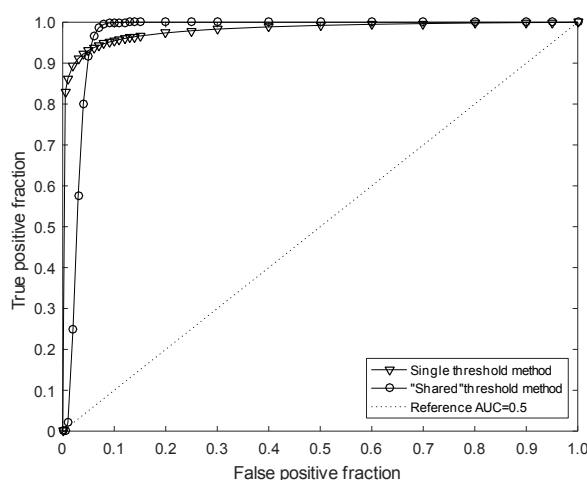


图3 两种取阈方法采用强分类器所得ROC曲线对比

Fig.3 Comparison of ROC curve obtained with two threshold methods using strong classifier

3 讨论

本研究的创新点: ① 与传统乳腺癌风险评估模型相比, 本文构建的乳腺癌风险预测模型无需采用基因检测技术, 且能够对个体女性近期患乳腺癌的风险进行预测; ② 根据现有文献, 目前研究者们

都是采用单独阈值法分割双侧乳腺高密度区, 本文提出一种新的“共用”阈值法来分割双侧乳腺高密度区。

综合对比分析两种取阈方法的实验结果不难发现, 从面向乳腺癌风险预测的应用来看, “共用”阈值法在以下3个方面要优于传统的单独阈值法: ① 与传统的单独阈值法相比, 采用“共用”阈值法分割双侧乳腺高密度区更有利于从高密度区提取更多具有高预测力的图像特征; ② 采用弱分类器, 基于“共用”阈值法所得的最高预测准确率(59.46%)及特异度(86.46%)显著高于基于单独阈值法所得的最高预测准确率(55.14%)及特异度(47.92%); ③ 采用强分类器, 基于“共用”阈值法所得预测准确率(95.68%)及敏感度(98.88%)显著高于基于单独阈值法所得的预测准确率(93.51%)及敏感度(91.01%)。

本研究的局限性: ① 所使用的图像数据库较小; ② 目前只采用了乳腺CC位图像。接下来的研究中, 我们拟与医院加强合作获取更多的图像数据; 进一步研究胸壁肌肉识别与分割算法; 基于更大的图像数据库, 结合使用CC位图像和MLO位图像进一步验证分类器性能的鲁棒性。

表3 预测准确率、敏感度、特异度对比(%)

Tab.3 Comparison of prediction accuracy, sensitivity, specificity (%)

Classification method	Single threshold method			"Shared" threshold method		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Weak classifier based on asymmetric feature and maximum feature	54.59	69.66	40.63	54.59	60.67	48.96
Weak classifier based on asymmetric feature	55.14	70.79	40.63	56.76	65.17	48.96
Weak classifier based on maximum feature	55.14	62.92	47.92	59.46	30.34	86.46
Strong classifier	93.51	91.01	95.83	95.68	98.88	92.71

4 结 论

通过本研究可以得出以下结论: ① 与现有左、右乳腺高密区分别采用各自阈值划分相比, 采用“共用”阈值法分割双侧乳腺高密度区更有利于从高密度区提取更多具有高预测力的图像特征; ② 采用由弱分类器组合构建的强分类器可显著提高个体女性乳腺癌近期发病预测精度。本研究构建的预测模型可显著提高个体女性乳腺癌近期发病风险预测精度, 有望降低临床诊断的假阳性率及假阴性率, 为乳腺癌的早期诊断提供潜在的有价值参考信息。

致谢: 感谢匹兹堡大学医学中心为本研究提供图像资料, 同时感谢所有为本研究做出贡献的人。

【参考文献】

- [1] YANG Q, LI L H, ZHANG J, et al. A new quantitative image analysis method for improving breast cancer diagnosis using DCE- MRI examinations[J]. Med Phys, 2015, 42(1): 103-109.
- [2] AGHAEI F, TAN M, HOLLINGSWORTH A B, et al. Computer-aided breast MR image feature analysis for prediction of tumor response to chemotherapy[J]. Med Phys, 2015, 42(11): 6520-6528.
- [3] BOGGS D A, ROSENBERG L, ADAMSCAMPBELL L L, et al. Prospective approach to breast cancer risk prediction in African American women: the black women's health study model[J]. J Clin Oncol, 2015, 33(9): 1038-1044.
- [4] PANKRATZ V S, DEGNIM A C, FRANK R D, et al. Model for individualized prediction of breast cancer risk after a benign breast biopsy[J]. J Clin Oncol, 2015, 33(8): 923-929.
- [5] DAI J C, HU Z B, JIANG Y, et al. Breast cancer risk assessment with five independent genetic variants and two risk factors in Chinese women[J]. Breast Cancer Res, 2012, 14(1): 1-12.
- [6] VARESCO L, VIASSOLO V, VIEL A, et al. Performance of BOADICEA and BRCAPRO genetic models and of empirical criteria based on cancer family history for predicting BRCA mutation carrier probabilities: a retrospective study in a sample of Italian cancer genetics clinics[J]. Breast, 2013, 22(6): 1130-1135.
- [7] MEADS C, AHMED I, RILEY R D. A systematic review of breast cancer incidence risk prediction models with Meta-analysis of their performance[J]. Breast Cancer Res Treat, 2012, 132(2): 365-377.
- [8] LINDOR N M, LINDOR R A, APICELLA C, et al. Predicting BRCA 1 and BRCA2 gene mutation carriers: comparison of LAMBDA, BRCAPRO, Myriad II, and modified Couch models[J]. Fam Cancer, 2007, 6(4): 473-482.
- [9] GAIL M H, MAI P L. Comparing breast cancer risk assessment models [J]. J Natl Cancer Inst, 2010, 102(10): 445-452.
- [10] CHANG Y H, WANG X H, HARDESTY L A, et al. Computerized assessment of tissue composition on digitized mammograms[J]. Acad Radiol, 2002, 9(8): 899-905.
- [11] TAN M, PU J T, CHENG S, et al. Assessment of a four-view mammographic image feature based fusion model to predict near-term breast cancer risk[J]. Ann Biomed Eng, 2015, 43(10): 2416-2428.
- [12] TAN M, PU J T, ZHENG B. Reduction of false-positive recalls using a computerized mammographic image feature analysis scheme[J]. Phys Med Biol, 2014, 59(15): 4357-4373.
- [13] TAN M, ZHENG B, RAMALINGAM P, et al. Prediction of near-term breast cancer risk based on bilateral mammographic feature asymmetry [J]. Acad Radiol, 2013, 20(12): 1542-1550.
- [14] WANG X W, LEDERMAN D, TAN J, et al. Computerized prediction of risk for developing breast cancer based on bilateral mammographic breast tissue asymmetry[J]. Med Eng Phys, 2011, 33(8): 934-942.
- [15] ZHENG B, TAN M, RAMALINGAM P, et al. Association between computed tissue density asymmetry in bilateral mammograms and near-term breast cancer risk[J]. Breast J, 2014, 20(3): 249-257.
- [16] WANG X W, LEDERMAN D, TAN J, et al. Computerized detection of breast tissue asymmetry depicted on bilateral mammograms: a preliminary study of breast risk stratification[J]. Acad Radiol, 2010, 17(10): 1234-1241.
- [17] VALLIERES M, FREEMAN C R, SKAMENE S R, et al. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities [J]. Phys Med Biol, 2015, 60(14): 5471-5496.

(编辑: 薛泽玲)