Comparing the performance of a mass detection CAD system when using manual and automatic breast density information

Oliver A.[1], Freixenet J.[1], Lladó X.[1], Martí R.[1], Pont J.[2], Pérez E.[2], Zwiggelaar R.[3]
[1] University of Girona, IIiA - IdiBGi, Girona, SPAIN
[2] Hospital Josep Trueta, Dept of Radiology - IdiBGi, Girona, SPAIN

[3] University of Wales, Dept of Computer Science, Aberystwyth, UNITED KINGDOM

Keywords CAD system, Mass detection, Breast Density

Purpose

In a previous work we presented a CAD system for mass detection [1]. This system aims to detect masses in two different steps. First of all, the system learns the shapes of the real masses and uses this information to create a set of mass templates. In the second step, these templates are used in a probabilistic matching scheme which allows us to provide a set of suspicious regions. Afterwards, a false positive reduction algorithm is applied in order to distinguish between regions being masses and those being normal tissue.

Comparing to other CADs [2,3], a distinctive feature of our system is the inclusion of the breast density information [4]. In this sense, we create different templates for each of the four BIRADS density categories. Therefore, in the detection step only those templates belonging to the same BIRADS as the analysed mammogram are used. Note that in the detection of a new mammogram the breast density classification has to be done automatically. However in the training step this breast density classification can be performed using automatic estimations or manual annotations obtained by the radiologist. The goal of this paper is to analyse and compare the performance of the CAD system when using manual or automatic breast density annotations in the CAD training step.

Materials and methods

In this paper we use 89 MLO mammograms and 87 CC mammograms obtained by a Siemens Mammomat Novation digital mammographic unit. Each image of the database is 70 micron pixel edge and is stored according to the DICOM protocol, with a bit resolution of 12-bit per pixel. Two different image sizes were possible: 2560x3328 or 3328x4096 pixels, although in both cases the images are resized by a factor of four for time-computing reasons.

In this experiment we ensure that all the mammograms contain at least one mass. Moreover, all the mammograms of the dataset have their own annotation which includes the breast density classified according the BIRADS categories, the number of lesions, and their contours.

The evaluation of our system is done using Receiver Operating Characteristic (ROC) analysis [5]. A ROC curve indicates the true positive rate as a function of the false positive rate. When no useful discrimination is achieved the true positive rate is similar to the false positive rate. As the accuracy increases, the ROC curve moves closer to the upper-left-hand corner, where a higher sensitivity corresponds to a lower false positive rate. A measure commonly derived from a ROC curve is the area under the curve Az, which is an indication for the overall performance of the observer. For an ideal classifier the Az value is equal to one. Note that in this experiment the ROC analysis is performed at pixel-level: each pixel of the image is treated as an instance of the classification process, and thus it can be a pixel belonging to a mass or a pixel being normal tissue.

Results

Table 1 shows the confusion matrices according to the mammogram internal density for manual and automatic classification. The algorithm clearly obtained better performance for MLO mammograms (84%) than for CC ones (74%). Examining each class alone, almost of all the mammograms belonging to BIRADS I are correctly classified for MLO mammograms, while for CC views the performance is reduced. Moreover, the two mammograms belonging to BIRADS IV are misclassified in both confusion matrices. Table 2 shows the obtained results when training the CAD according to both manual and automatic annotations. Using the manual annotations provided by the radiologists, the Az value obtained by the CAD is 0.78 for CC views and 0.82 for MLO mammograms. In contrast, when training the CAD using the automatic breast density estimation we obtained Az = 0.79 for CC images and Az = 0.84 for MLO images. Observe that the results obtained when training the system using the automatic estimation are higher than those obtained when training the system using the expert annotations.

Table 1 Confusion matrix for breast density estimation. (a) MLO views and (b) CC views

a Automatic (84%) — MLO

Truth	B-I	B-II	B-III	B-IV
B-I	42	0	2	0
B-II	5	12	2	1
B-III	0	2	18	0
B-IV	0	1	1	0

b Automatic (74%) — CC

Truth	B-I	B-II	B-III	B-IV
B-I	36	4	3	2
B-II	6	11	2	0
B-III	2	1	17	0
B-IV	1	0	1	0

Table 2 Comparison of the algorithms performance (in terms of A_z) using manual or automatic breast density estimation

View	Expert Annotations	Automatic Estimation
CC	0.78 ± 0.15	0.79 ± 0.14
MLO	0.82 ± 0.13	0.84 ± 0.13

Conclusion

This paper presents a comparison when training a mass detection CAD system using manual annotations or automatic estimation of the breast density. Results obtained when using the automatic estimations outperform the ones obtained using the annotations provided by the radiologists. In this sense, the results of our experiments show that automatic methods are able to capture the mammogram appearance with more objectivity than a human expert. Moreover, the results of our CAD show the importance of developing an initial classification of the mammograms according to their breast density information.

Acknowledgment

This work was supported by the Ministerio de Educación y Ciencia of Spain under Grant TIN2007-60553 and by the UdG under Grant Id-iBGi-UdG91060080.

References

[1] J. Freixenet, A. Oliver, X. Lladó, R. Martí, and J. Martí. Mass Eigendetection and the Benefits of Introducing Breast Density Information, Proc. CARS, vol 2 suppl 1, pp 332–334, 2007.
[2] R2 ImageChecker, http://www.r2tech.com, accessed 01/01/2008.
[3] iCAD Second Look, http://www.icadmed.com, accessed 01/01/2008.
[4] J. N. Wolfe, Risk for breast cancer development determined by mammographic parenchymal pattern, Cancer, vol. 37, pp. 2486–2492, 1976.
[5] C. E. Metz, Evaluation of digital mammography by ROC analysis, in Int. Work. Dig. Mammography, 1996, pp. 61–68.
[6] J. Cohen. A coefficient of agreement for nominal scales, Educational and Psychological Measurement, vol 20, pp 27–46, 1960.