Automated Classification of Masses on Mammography

LUCIANA MARINHO SOARES[1]
AURA CONCI[1]
ALBERTO D. VIANNA[2]

[1]CAA- Computação Aplicada e Automação
[2]IRSA and HUAP - Departamento de Radiologia – MRD
UFF – Universidade Federal Fluminense
Rua Passo de Pátria, 156, CEP 240210-240, Niterói, RJ, Brasil
{marinho, aconci}@caa.uff.br

Abstract. A scheme for identification of breast cancer as benign or malignant based on pattern recognition is presented. A database for use by the mammographic image analysis research community has been establish (http://www.caa.uff.br/~aconci/mam/framex1.htm). From these images, fifty-two cases with undoubted diagnostic have been used as input pattern for feature extraction and classification training . After extensive experimentation a set of features is extracted using shape and contour characterization. Two classes of classifier are used: discriminant functions and nearest neighbor classifier. We implemented an automatic computer diagnosis system that performs analysis capable of correct classification (with zero rate of false-positive and false-negative) on all tested cases until now. The original contributions of this work are: its project using a database, that expands with easiness whenever a new image with proven diagnosis be introduced in the bank, and the two used forms of classifiers that result in five independent approaches for evaluation of the mammography.

Keywords: Biomedical Image Processing, mammography image database, breast cancer, classification of nodules, mass classifier in mammography, digitized mammograms.

1 Introduction

Mammograms can depict most of the significant changes of breast tissue. The primary clinical signs of cancer are masses or tumors. The presence of spicular lesions or more diffuse stellate appearance in mammogram masses characterizes malignant breast cancer (Kopans, 1989). The medical diagnosis by mammogram is based on these identification (Tavassoly, 1992). Therefore, in the development of computer algorithms to aid in the diagnosis these patterns can also be used on tumor identification as benign or malignant cases. Unfortunately, they are very difficult to include on automated image analysis algorithms. A stellate tumor has an irregular shape with borders radiating spicules that may extent from few millimeters to many centimeters in size (Jiang et. al., 1997). Moreover, breast tissue around the masses may vary from grease to dense, the former presents tumors with well-defined border, and the latter represents poorly contrasted gray-level regions and images with ill-defined borders (Liu--Delp,1997). That is the x-ray image can presents very different sizes and grades of definition (De Paredes, 1989). Computer aided algorithms concentrate mainly on enhancement of mammograms to radiologist (Anguh—Silva,1997 ; Chang—Laine,1997 ; Crestana et al. ,1992). Many works consists of feature extraction followed by classification (Lorey et al. (1995), Méndez et al.(1996), Wodds--Bowyers, 1996). In this paper, we present a scheme to classifications of lesions in mammograms based on pattern discriminant functions and nearest neighbor classifier. In the next section, we describe the databases used for pattern identification and the classification approach. Then we present experimental results and conclusions.

2 Diagnostic Approach

A project to establish a database for use by the mammographic image analysis community is a collaborative effort involving the Antonio Pedro

University Hospital-HUAP, the Radiology Department of the Faculty of Medicine, the postgraduate course on Computer Application and Automation-CAA (Medical Images Research Program) of the Federal Fluminense University–UFF and the IRSA-Institute of Radiology S.A. The primary purpose of the database is to facilitate research in the development of computer algorithms to aid in screening and diagnosis. Secondary purpose of the database may include teaching or training aids. The database contains cases collected along 3 decades by Prof. A. D. Vianna. All images have been made available by IRSA and can be seen at http://www.caa.uff.br/~aconci /mam/framex1.htm . Each case includes breast image with diagnostic information from expert radiologists of the Department of Radiology. Both benign and malignant cases are included. From these 52 original images from different patients of proven diagnostic (all these have had either a biopsy proven or at least 3 years of subsequent follow-up without change) were identified by expert radiologists as the most representative cases and have been used to train the algorithm. These 52 images of the internal system database are composed of 27 benign cases and 25 malignant cases and have been used for feature extraction and pattern classifier (fig. 1).

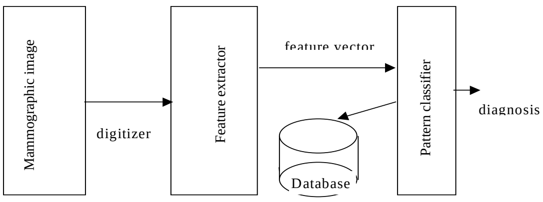


Figure 1 – Simplified diagram of the implementation

All images on the database (used to train the algorithm) are scanned on 8-bits per pixel (256 gray-level as figure 2 shows). Features are extracted within a certain neighborhood. In this work a surface of 200x200 pixels are analyzed (figure 3). If G(i,j) represents its gray-level (from 0 to 255) for each pixel (i,j), then the most important gray-level is the thresholding between the nodule gray-level, g_n , and the gray-level of its neighborhood. This has to be defined by histogram identification (Hussain, 1991). The parameters used on classification combine 6 features: the nodule boundary length, the nodule area, its inertial tensor of order two (2 features) and three (2 features). These features are extracted from the following parameters of the digitized image, G(i,j):

1. The number of pixels on the tumor edge (top images on fig. 3) : $E = \sum B\partial$, where ∂B represents the edge pixels of the nodule.

2. The nodule area (middle images on fig. 3) is defined as $A = \sum i \sum j B(i,j)$ where,

   $B(i,j) = 1$ if $G(i,j) \geq g_n$

   and $B(i,j) = 0$ if $G(i,j) < g_n$.

3. Considering the area's centroid (i_0 , j_0), where $i_0 = m_{10} / A$ and $j_0 = m_{01} / A$. The (p, q) central moment of $B(i,j)$ is defined as $m_{pq} = \sum \sum (i - i_0) p (j - j_0) q$. The second order central moments of $B(i,j)$, defines its inertial tensor of order two:

$$\begin{vmatrix} m_{20} & m_{11} \\ m_{11} & m_{02} \end{vmatrix}$$

This tensor has 2 invariant, its trace and its determinant. The former is the polar moment of inertia around the centroid: $I_1 = m_{20} + m_{02}$. The eccentricity is also an invariant: $I_2 = (m_{20} - m_{02})2 + 4m_{11}^2$. The thirty order central moments of $B(i,j)$, defines a tensor of order three. This tensor presents several invariant. In this work we have used:

$$I_3 = (m_{30} - 3m_{12})^2 + (3m_{21} - m_{03})^2$$

$$I_4 = (m_{30} + m_{12})^2 + (m_{21} + m_{03})^2$$

4. For classification accuracy of small and large tumors, we divide each feature by area powers: E/A ; I_1/A^2; I_2/A^4; I_3/A^5; and I_4/A^5.

These features are then classified. The classifier identifies the feature vector into one of the two classes: benignant-B or malignant-M. The decision boundary between the classes is expressed by discriminant functions and minimum distance classifier. In the minimum distance classifier, we use all benignant and malignant images in the database. We decide a new feature vector to be in class B (or M) if its distance to B (or M) is the minimum. For explicitly specifying the discriminant functions, we first have been made great number of analyses and experiments to specify the decision boundaries. This learning procedure show us that the two classes are separable by a plane in the 3-dimensional space defined by the features E/A, I_1/A^2 and I_4/A^5. Then the classification is realized using threshold logic (Li et al.,1997). Therefore, all these decision rules have been used in the implementation, which block diagram is illustrated in figure 1.

3 Experiments and Conclusions

The following figure 2 and 3 illustrated the two first steps of the program. The above mentioned 52 selected images of the training process are not used for performance evaluation. Experiments are done with completely different mammograms from Nijmegen Database (HTTP://www.mammography/Nijmegen). The images of this public database were scanned at a resolution of 50µm x50µm and 8-bits deep. All forty (40) Nijmegen mammograms were tested: seven (7) are benign cases (images 1,2,7,8,9,10 and 14 are benign cases) and thirty-three (33) are malignant cases. Each new image to be classified is compared with all benign (27 images) and malignant (25) images of the system database on the same time. These 40 tested images have been correctly classified with zero false-negative rate and zero false positives rate. That is the recognition rate for tumors with spicules was 100% on low-density breast. Subsequent extension of these databases and algorithm will include microcalcifications and surrounding dense regions (Byng et al.,1997).

Acknowledgment

This work was supported, in part, by project FINEP/RECOPE SAGE #0626/96, CNPq (302649/87-5) and CAPES

The authors acknowledge the kind help from the referees, especially for pointing out relevant references, in particular the http://www.rose.brandeis.edu/users/mammo/digital.html

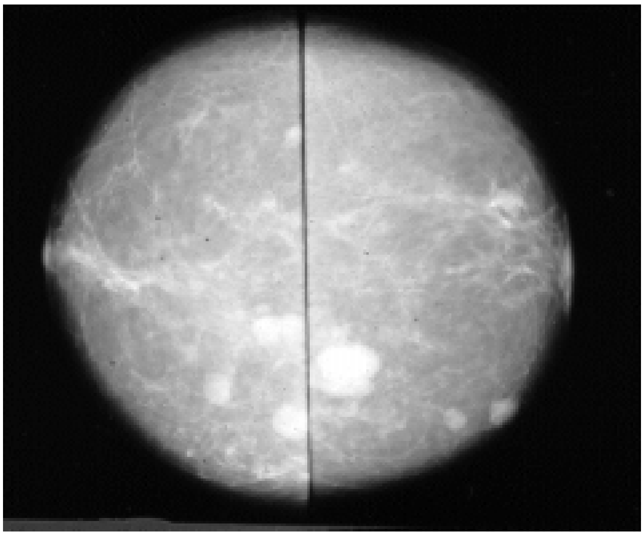


Fig 2 – Original image on the database

References

Anguh, M. M. e Silva, A. C., "Multiscale Segmentation and Enhancement in Mammograms", *Anais do X SIBGRAPI*, 1997, 136-139. <dpi.inpe.br/banon/1998/03.25.16.20>.

Braga, Neto, U. M. e Siqueira, Neto, W. A., "A Technique for the Automatic Detection and Classification of Mammographic Calcifications using the Watershed Method", *Anais do VIII SIBGRAPI*, 1995, 273-280.

Byng, J. W., Critten, J. P., Yaffe, M. J., Little, L., Lockwood, G., Jong, R. A., Fishell, E., Tritchler, D. e Boyd, N. F.,1997, "Analysis of Digitised Mammograms for the predition of Breast Cancer Risk", *http://www.radonc.uchicago.edu/IWDM/IWDMa27.html*.

Chang, C.-M., Laine, A., 1997, "Enhancement of Mammograms from Oriented Information", *ICIP'97- IEEE Proceedings of International Conference on Image Processing*, October 26-29, Santa Barbara, Vol. 3, pp. 524-527, No 607

Crestana, R. H. S., Guillardi Neto, T. e Jorge, L. A. C., "Realce de Mamografias Através de Processamento de Imagens", *SIBGRAPI V*, 1992, 9-12.

De Paredes, E. S., 1989, *Atlas of Film-Screen Mammography.*

Jiang, H., Tiu, W., Yamamoto, S., Iisaku,S-I., 1997, "Detection of Spicules in Mammograms", *ICIP'97- Proceedings of IEEE International Conference on Image Processing*, October 26-29, Santa Barbara, Vol. 3, pp. 520-523, No 380

Hussain, Z. , 1991, *Digital Image Processing: practical applications of parallel processing techniques, Ellis Horwood.*

Kopans, D.B., 1989, *Breast Imaging*, J. B. Lippincott Co.

Li, H. , Liu, K.J.R., Lo, S-C B., Wang, Y., 1997, "Stochastic Model and Probabilistic Decision-Based Classifier for Mass Detection in Digital Mammograms", *ICIP'97- Proceedings of IEEE International Conference on Image Processing*, October 26-29, Santa Barbara, Vol. 3, pp. 539-542, No 460

Liu, S., Delp, E.J., 1997, "Multiresolution Detection of Stellate Lesions in Mammograms", *ICIP'97- Proceedings of IEEE International Conference on Image Processing*, October 26-29, Santa Barbara, Vol. 2, pp. 109-112, No 614

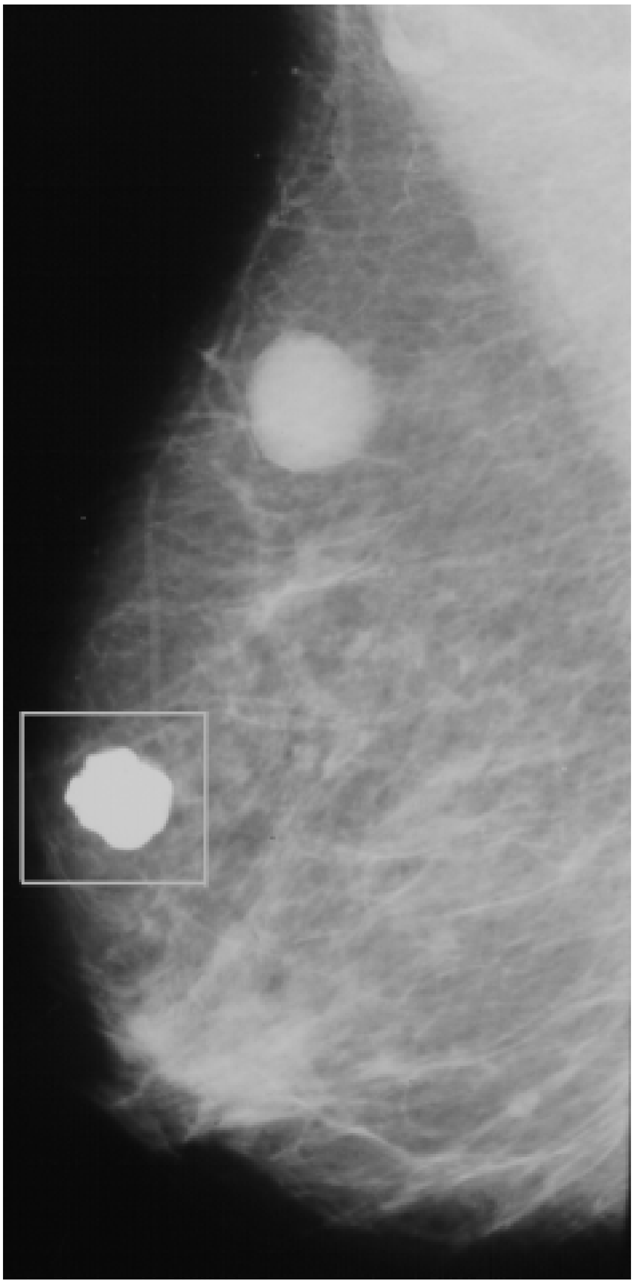
Lorey, R. A., Solka, J. L., Rogers, G. W., Marchette, D. J. e Priebe, C. E., 1995, "Mammographic Computer-Assisted Diagnosis using Computational Statistics

Pattern Recognition”, *Real-Time Imaging, Academic Press*, Vol.1,95-104.

Méndez, A. J., Tahoces, P. G., Lado, M. J., Souto, M. e Vidal, J. J., 1996, “Computer-Aided Diagnosis:Detection of Masses on Digital Mammograms”, *Proceedings IWISP’96*, 4-7 November, Manchester, U. K., 465-468

Tavassoly, F. A. , 1992, *Pathology of the Breast.*

Wodds K. and Bowyers K., 1996, “A general view of detection algorithms”, *Proceedings of 3rd International Workshop on Digital Mammography* , June, 9-12, Chicago, pp. 385-390.



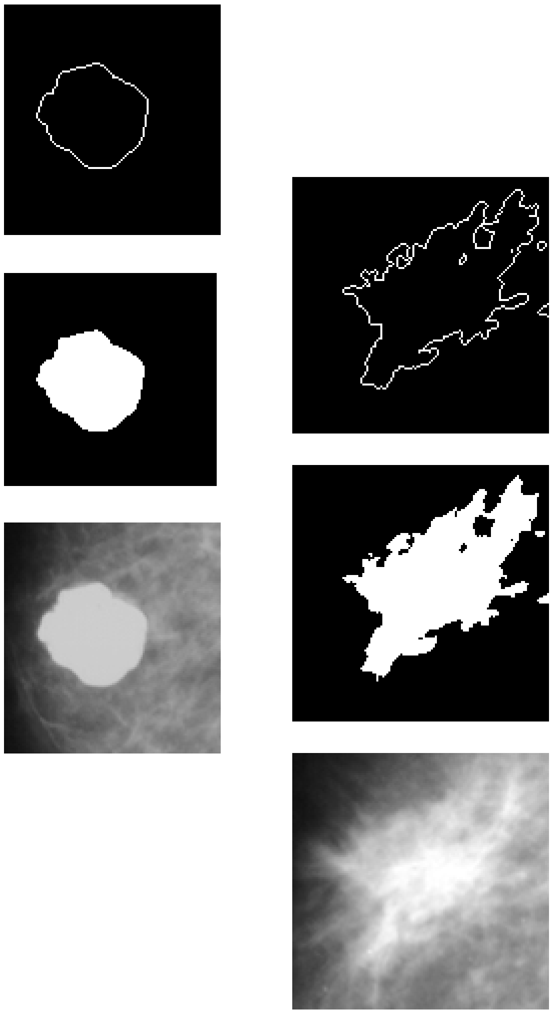


Fig 3 – Regions to be investigated on a mammogram (left). Benign tumor, its area and contour (center). Stellate malignant tumor, its area and contour (right).