Radiation Physics and Chemistry 95 (2014) 261–263

Contents lists available at ScienceDirect

Radiation Physics and Chemistry

journal homepage: www.elsevier.com/locate/radphyschem

# Implementation of several mathematical algorithms to breast tissue density classification

C. Quintana [a], M. Redondo [b], G. Tirao [a,*]

[a] IFEG-CONICET and Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba, Medina Allende y Haya de la Torre, 5000 Córdoba, Argentina

[b] Facultad de Ciencias Médicas, Universidad Nacional de Córdoba, Pabellón Perú, Ciudad Universitaria, 5000 Córdoba, Argentina

## HIGHLIGHTS

- Breast density classification can be obtained by suitable mathematical algorithms.
- Mathematical processing help radiologists to obtain the BI-RADS classification.
- The entropy and joint entropy show high performance for density classification.

## ARTICLE INFO

*Article history:*
Received 1 October 2012
Accepted 10 October 2013
Available online 17 October 2013

*Keywords:*
Breast density classification
Mathematical processing
Computer-aided diagnostic systems
Mammography

## ABSTRACT

The accuracy of mammographic abnormality detection methods is strongly dependent on breast tissue characteristics, where a dense breast tissue can hide lesions causing cancer to be detected at later stages. In addition, breast tissue density is widely accepted to be an important risk indicator for the development of breast cancer. This paper presents the implementation and the performance of different mathematical algorithms designed to standardize the categorization of mammographic images, according to the American College of Radiology classifications. These mathematical techniques are based on intrinsic properties calculations and on comparison with an ideal homogeneous image (joint entropy, mutual information, normalized cross correlation and index Q) as categorization parameters. The algorithms evaluation was performed on 100 cases of the mammographic data sets provided by the *Ministerio de Salud de la Provincia de Córdoba, Argentina—Programa de Prevención del Cáncer de Mama* (Department of Public Health, Córdoba, Argentina, Breast Cancer Prevention Program). The obtained breast classifications were compared with the expert medical diagnostics, showing a good performance. The implemented algorithms revealed a high potentiality to classify breasts into tissue density categories.


## 1. Introduction

The breast tissue density is widely accepted to be an important risk indicator for the development of breast cancer. The BI-RADS (Breast Imaging Reporting Data System) density scale, developed by the American College of Radiology (ACR) (Reston, 1995), informs radiologists about the decline in sensitivity of mammography with increasing breast density. ACR defines four breast density categories according to the percentage of fat and fibroglandular tissue and radiologists evaluate and report it on the basis of visual analysis of mammography. The breast density categories are: density I, almost entirely fatty; density II, scattered fibroglandular tissue; density III, heterogeneously dense tissue and density IV as extremely dense tissue.

The various distribution of the parenchyma tissue makes automatic classification a difficult task, so in this scenario, as auxiliary tools, computer-aided diagnosis (CAD) and content-based image retrieval (CBIR) systems appear as real possibilities to help radiologists to reduce the variability of their analysis and also to improve the accuracy of mammography interpretation. CBIR systems (del Bimbo, 1999; Müller et al., 2004) use visual information extracted from images of database to retrieve similar images to a specific query. In CBIR systems, images are described as feature vectors and similarity is determined using measures of distance. The choice of a set of features that are able to capture pictorial content in a way closer to human perception is still a challenge. Some works have explored the use of CBIR and CAD systems to improve knowledge and provide facilities on these modalities. All the reported works agree with the importance of

* Corresponding author. Tel.: +54 3514 334 051; fax: +54 3514 334 054.
*E-mail address:* gtirao@famaf.unc.edu.ar (G. Tirao).

0969-806X/$ - see front matter 
http://dx.doi.org/10.1016/j.radphyschem.2013.10.006

the proper characterization of breast density, since this is critical to the retrieval process. The most effective features used for characterization were extracted from the gray level histogram (Wang et al., 2003) and texture patterns (Kinoshita et al., 2007), using CAD and CBIR systems, respectively. Wang et al. (2003) used histograms to automatically evaluate breast density according to ACR categories. A 71% of correct classification was obtained with the use of a neural network classifier. Kinoshita et al. (2007) used breast density as a pattern to retrieve images from a 1080 mammographies database. Shape, size and texture features, as well as less-explored features based on the Radon domain and granulometric measures, were applied to breast density characterization, employing the Kohonen self-organizing map neural network for the retrieval task. For breast density classification, Oliver et al. (2008) developed an automatic breast tissue classification methodology, using morphological and texture features. Castella et al. (2007) developed a semi-automatic method through texture features, gray level histograms, primitives, neighborhood gray tone difference matrix and fractal analysis.

Under the Breast Cancer Prevention Program, projected and funding by the Department of Public Health, Córdoba, Argentina, about 100 mammographic studies are performed weekly, where four mammographic images are acquired in each study. The breast density classification was performed by two radiologists for being this classification relevant and subjective, in order to balance possible errors and different criteria. Therefore, some mathematical feature of the mammographic image could be very useful to aid and facilitate the radiologist tasks. In this paper, the performance of different mathematical algorithms designed was studied to standardize the classification of breast density, according to ACR categories. These mathematical algorithms were developed to find image parameters calculated in an easy way, with low computational cost, without requiring a database and designed to be implemented in Public Health Services. In this sense, several parameters, widely used for image quality evaluation as mean value, standard deviation, signal-to-noise ratio (SNR) and contrast ratio to noise ratio (CNR), were implemented for breast density characterization. It was also used the entropy ($H$) (Pluim et al., 2003), parameter derived from information theory. Finally, a new method was proposed to classify the breast according the BI-RADS density scale, using a similarity parameter, such as the normalized cross-correlation (NCC) (Lewis, 1995), parameters derived from information theory, such as join entropy (JH) and mutual information (MI) (Pluim et al., 2003), and the index $Q$ (Wang and Bovik, 2002), commonly used for assessment of image quality.

## 2. Experimental

The mammographic image dataset used in this study correspond to patients of the Breast Cancer Prevention Program of the Department of Public Health, Córdoba, Argentina, containing 100 mammographic images, 50 craniocaudal (CC) and 50 Medio-Lateral Oblique (MLO) views of breast.

The mathematical algorithms were designed to reveal the dependence on breast density with diverse parameters. As a first step, a recognition procedure was applied to identify the region where the breast is located. This procedure is based on generating a segmented image ($I_{segm}$) where the pixels corresponding to the direct beam (background pixels) are set to zero and the rest of the pixels hold their original value. The $I_{segm}$ was obtained from a binary image ($I_{bin}$) calculated by segmentation of the original image, with a threshold defined as $M_{back}+3\sigma_{back}$, where $M_{back}$ is the mean value of background pixels, and $\sigma_{back}$ is their standard deviation. Note that the $I_{bin}$ is an image where the background pixels have a value of zero and the rest of the pixels have the value one.

For breast density classification, the SNR and CNR parameters were calculated from the $I_{segm}$ and the total number of pixels that have non-zero values was considered as breast size. Another parameter tested was the information entropy $H$ (Pluim et al., 2003), which is evaluated from the probability distribution of the intensity, calculated from the histogram of the segmented image $I_{segm}$. Besides, it was proposed to apply a hybrid method, between CAD and CBIR techniques, where similarity parameters were used to quantify the similarity between a mammographic image and the binary image $I_{bin}$, considered as a patron image. They were used as similarity parameters, the NCC (Lewis, 1995), the JH and the MI (Pluim et al., 2003), and the index $Q$ (Wang and Bovik, 2002). All the obtained results of density classification were analyzed and checked by expert medical radiologists.

## 3. Results and discussions

The mammographic image dataset was divided into two groups of 50 images each (CC and MLO), corresponding to the views of normal breast. Each of these two groups was divided into 8 sub-groups according to the parameter value $t$, defined as

$$t=8\frac{P-P_{min}}{P_{max}-P_{min}}$$

where $P$ is output parameter value of the considered mathematical algorithm, and $P_{max}$ and $P_{min}$ are the maximum and minimum values obtained of $P$ for all images of the each sub-group. The $t$ value fluctuates between 0 and 8. From the parameter $t$ value, the 8 sub-groups were formed, which are not necessarily equivalent, since a priori a linear relationship between breast density and the obtained parameters by the different methods cannot be assumed. A trained radiologist evaluated the images belonging to each sub-group studying their intensities and classifying them according to ACR categories. Table 1 indicates the ACR categories of mammographic images of each sub-group, according to the different methods used. The '–' denotes sub-groups with disparate breast density, and two numbers mean that they are between two categories.

From Table 1, it can be observed that the $M_{im}$, $\sigma$, SNR and CNR parameters well recognized the low density breasts. While $M_{im}$ and $\sigma$ show an increasing tendency with breast density, SNR and CNR show a particular behavior, high density breasts take intermediate values and low density breasts extreme values. On the other hand, MI and NCC parameters distinguish with a good performance either breasts of high and low densities, but not in the intermediate regions. The MI parameter exhibit an increasing trend with breast density showing more sensibility to low density breasts. NCC presents a decreasing tendency with high sensibility

**Table 1**
Breast density classification of mammographic images of each sub-group (Sg), according to the different methods used. The classification was made taking into account the ACR categories. – denotes sub-groups with disparate breast density, and two numbers means breast density between two categories.

| | $M_{im}$ | $\sigma$ | SNR | CNR | $H$ | JH | MI | NCC | $Q$ |
|---|---|---|---|---|---|---|---|---|---|
| Sg 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | – |
| Sg 2 | 1 | 1 | – | – | 1 | 1 | 1 | 4 | – |
| Sg 3 | – | 1 | – | – | 1 | 1 | – | 4 | – |
| Sg 4 | – | – | – | – | 1–2 | 1–2 | – | 3–4 | – |
| Sg 5 | – | – | – | – | 2–3 | 2–3 | – | – | – |
| Sg 6 | – | – | – | – | 3 | 3 | – | – | – |
| Sg 7 | – | – | 1 | 1 | 3–4 | 3–4 | – | – | – |
| Sg 8 | – | – | 1 | 1 | 4 | 4 | 4 | 1 | – |

to high density breasts. The $H$ and JH parameters show a very good performance for whole range of densities, with a high correlation between ACR classification and breast images belonging to each sub-group. These parameters have greater sensitivity in the region of low density breasts. Finally, it can be clearly observed that index $Q$ shows no relation to breast density, grouping in each subgroups breast images with very different densities, without any obvious tendency.

## 4. Conclusion

With the aim to find simple parameters to specify the breast density, their dependence with different parameters was studied, such as commonly used parameters for image quality evaluation ($\sigma$, SNR and CNR), the mean value $M_{\mathrm{im}}$ and the information entropy $H$. Besides, a hybrid method was also proposed to classify breast density, where similarity parameters were used to quantify the similarity between a mammographic image and the binary image $I_{\mathrm{bin}}$, considered as a patron image. The NCC, parameters as JH and MI, used in the information theory, and the index $Q$ were used as similarity parameters.

All studied parameters, except the index $Q$, showed some relation to breast density. $M_{\mathrm{im}}$, $\sigma$, SNR and CNR efficiently recognize low density breasts, while MI and NCC properly identify breasts with both high and low densities. The $H$ and JH parameters show high correlation between obtained values and breast density for the whole density range, demonstrating a great potential to be consider as useful density classification parameters. The implemented algorithms revealed a high potentiality to breast density classification, which would be proven over an extended dataset to obtain a more general and reliable validation, according to breast density classification performed by several medical and radiologist experts.

## Acknowledgments

Financial supports from the *Consejo Nacional de Investigaciones Científicas y Técnicas* (CONICET) is gratefully acknowledged. CONICET partially supported this work with a PhD fellowship to C.Q. The authors thank the Breast Cancer Prevention Program of the Health Ministry of Córdoba for the mammographic image dataset.

## References

Castella, C., Kinkel, K., Eckenstein, M.P., Sottas, P.E., Verdun, F.R., Bochud, F.O., 2007. Semiautomatic mammographic parenchymal patterns classification using multiple statistical features. Acad. Radiol. 14, 1486–1499.

del Bimbo, A., 1999. Visual Information Retrieval, first ed. Morgan Kaufmann Publishers Inc, California.

Kinoshita, S.K., Marques, P.M.A., Pereira, R.R., Rodrigues, J.A.H., Rangayyan Jr., R.M., 2007. Content-based retrieval of mammograms using visual features related to breast density patterns. J. Dig. Imaging 20, 172–190.

Lewis, J.P., 1995. Fast template matching. In: Proceedings of the Canadian Image Processing, pp. 120–123.

Müller, H., Michoux, N., Bandon, D., Geissbuhler, A., 2004. A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. Int. J. Med. Inform. 73, 1–23.

Oliver, A., Freixenet, J., Martí, R., Pont, J., Pérez, E., Denton, E.R.E., Zwiggelaar, R., 2008. A novel breast tissue density classification methodology. IEEE Trans. Inf. Technol. Biomed. 12, 55–65.

Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A., 2003. Mutual-information-based registration of medical images: a survey. IEEE Trans. Med. Imaging 22, 986–1004.

Reston, V.A., 1995. Breast Imaging Reporting and Data System, 2nd ed. American College of Radiology, Reston, VA.

Wang, Z., Bovik, A.C., 2002. A universal image quality index. IEEE Signal Process. Lett. 9, 81–84.

Wang, X.H., Good, W.F., Chapman, B.E., Chang, Y.H., Poller, W.R., Chang, T.S., Hardesty, L.A., 2003. Automated assessment of the composition of breast tissue revealed on tissue-thickness-corrected mammography. Am. J. Roentgenol. 180, 257–262.