Semiautomatic estimation of breast density with DM-Scan software

I. Martínez Gómez a,*, M. Casals el Busto b, J. Antón Guirao c, F. Ruiz Perales d, R. Llobet Azpitarte e

a Programa de Cribado de Cáncer de Mama de la Comunidad Valenciana, Unidad de Alzira, Alzira, Valencia, Spain
b Programa de Cribado de Cáncer de Mama de la Comunidad Valenciana, Unidad de Valencia, Valencia, Spain
c Instituto Tecnológico de Informática, Universidad Politécnica de Valencia, Valencia, Spain
d Programa de Cribado de Cáncer de Mama, Centro Superior de Investigación Salud Pública (CSISP), Valencia, Spain

Received 27 July 2012; accepted 22 November 2012
Available online 3 November 2014

KEYWORDS
Mammography; Breast cancer; Breast density; Digital processing of mammograms; Computer-assisted diagnosis; Intraobserver variability; Interobserver variability

Abstract
Objective: To evaluate the reproducibility of the calculation of breast density with DM-Scan software, which is based on the semiautomatic segmentation of fibroglandular tissue, and to compare it with the reproducibility of estimation by visual inspection.
Materials and methods: The study included 655 direct digital mammograms acquired using craniocaudal projections. Three experienced radiologists analyzed the density of the mammograms using DM-Scan, and the inter- and intra-observer agreements between pairs of radiologists for the Boyd and BI-RADS® scales were calculated using the intraclass correlation coefficient. The Kappa index was used to compare the inter- and intra-observer agreements with those obtained previously for visual inspection in the same set of images.
Results: For visual inspection, the mean inter-observer agreement was 0.876 (95% CI: 0.873–0.879) on the Boyd scale and 0.823 (95% CI: 0.818–0.829) on the BI-RADS® scale. The mean intra-observer agreement was 0.813 (95% CI: 0.796–0.829) on the Boyd scale and 0.770 (95% CI: 0.742–0.797) on the BI-RADS® scale. For DM-Scan, the mean inter- and intra-observer agreement was 0.92, considerably higher than the agreement for visual inspection.
Conclusion: The semiautomatic calculation of breast density using DM-Scan software is more reliable and reproducible than visual estimation and reduces the subjectivity and variability in determining breast density.

© 2012 SERAM. Published by Elsevier España, S.L.U. All rights reserved.

* Please cite this article as: Martínez Gómez I, Casals el Busto M, Antón Guirao J, Ruiz Perales F, Llobet Azpitarte R. Estimación semiautomática de la densidad mamaria con DM-Scan. Radiología. 2014;56:429–434
* Corresponding author.
E-mail address: martinez_inm@gva.es (I. Martínez Gómez).

2173-5107/$ - see front matter © 2012 SERAM. Published by Elsevier España, S.L.U. All rights reserved.
Introduction

Screening mammography programs have reduced the rates of mortality of breast cancer up to 22% in women >50 years old and 15% in women between 40 and 49 years old.\(^1,2\) However, the larger the breast density is, the lower the sensitivity of the mammography – also associated with higher rates of interval cancers with poor prognosis when detected clinically.\(^3,4\) A dense breast tissue is per se a risk factor for breast cancer that is 4–6 times bigger in high dense breasts (breast density >75%) than in fat breasts (breast density <10%).\(^5,6\)

Breast density shows the amount of fibro glandular tissue with respect to breast fat. Since 1976 when Wolfe\(^7\) associated breast density with breast cancer and defined a 4 grade-categorization to grade it different categorizations have been used–all of them based on qualitative\(^8\) or quantitative\(^9,10\) criteria from visual or semiautomatic analysis of mammographies. BI-RADS\(^9,10\) 4th edition published by the American College of Radiology (ACR) standardizes breast density in the visual inspection by using a discrete quantitative scale categorizing density into four different categories. Both this and Boyd’s categorization establishing 6 different quantitative categories\(^7\) are the most widely accepted scales today. There is proof that the quantitative measure in the mammographic percentage of dense area is a better predictor of risk than the categorization in qualitative categories.\(^5,8\)

Today there are several automation modalities\(^12\) that improve the estimation of breast density and the categorization of parenchymal patterns like, for instance, the modalities introduced by Cumulus from the University of Toronto\(^1\) or Madena from the University of Southern California.\(^13\) These modalities use analog mammographies that need to be scanned and digitized. Nevertheless, digital mammography is a widely used and accepted modality. The goal of this study is to develop a computer program that will allow us to analyze and estimate an objective and reproducible breast density in direct digital mammographies without further processing.\(^14\)

Materials and methods

Patients

Six hundred and fifty-five (655) direct digital mammographies in a former project Determinants of Density in Mammography in Spain (DDM-Spain) (project FIS PI060386) that investigated breast density as one of the risk factors of breast cancer in 3584 women from 7 Spanish screening centers. For our study the mammographies were randomized out of 3 screening centers equipped with digital mammograms: Barcelona – Hologic\(^\circ\) Lorad M-IV\(^\circ\) digital mammogram (Bedford, MA, USA); Palma de Mallorca–Siemens Novation\(^\circ\) digital mammogram (Muenchen, Germany); and Valencia–Senographe 2000\(^\circ\) GE Medical System S.A. (Buc Cedex, France). Women were between 45 and 69 years of age. Women diagnosed with ovarian or breast cancers, those who underwent a breast surgical procedure or those who were prosthetic carriers were precluded. Both approval and informed consent came from Carlos III Institute of Health Bioethical Committee (Madrid, Spain).

Study modality

In a 1st stage 3 highly experienced radiologists in screening mammographies visually categorized the breast density of 655 mammographies according to Boyd and BI-RADS\(^\circ\) scales (Fig. 1). Previously they had done a combined training
with 300 direct digital images from different screening programs—technically compatible though. To assess breast density the left craniocaudal view (CCV) was used because it has fewer technical issues when proceeding to segment the image since it includes less chest muscle than other projections. Since to categorize breast density a high resolution image is not necessary for the analysis one conventional 17’’ monitor and a non-diagnostic radiologic image visualization software were used (K-PACS V1.6.0; free software, http://www.k-pacs.de).

In a 2nd stage 3 months later we used a computer application (MD-Scan) specifically designed for this study in the Universidad Politécnica de Valencia (Valencia, Spain). It is a tool used for computed assisted design (CAD) in order to assess breast density and simultaneously reduce its subjectivity. This application was designed to work with PNG format images that use a compression algorithm to reduce the weight of the image without losing quality. This is why the original DIGICOM digital format original images were converted into PNG format images through the DIGICOM 2 free software. When loading the mammographic image DM-Scan first processes it for the sake of identifying the breast and then proceeds to segment it automatically—defining the breast contour while isolating the rest of the image. When segmentation is not precise like when chest muscle or other image components are included in the segmented area, it is possible to preclude them. Through this process we can obtain the total size of the breast measured in number of pixels. Then the radiologist defines the threshold of brightness to establish the separation between the dense and the fat tissue which in turn allows us to know the dense tissue total area also measured in number of pixels and thus its exact percentage with respect to the breast total size.

We need to remember that the attenuation of X-rays and therefore the level of brightness in the mammography is not only based on the tissue density but also on its thickness. This means that the breast side closer to the chest muscle is sometimes brighter than peripheral side and this is why the selection of dense tissue only from a threshold of brightness usually picks out regions closer to the chest muscle. To correct this defect a digital filter called “breast filter” is used to estimate breast thickness in each image point while proportionally darkening the corresponding pixel. Given that the real thickness of each breast is not known the application gives us manual controls that allow us to edit the parameters configuring this filter. Particularly the brightness of each breast pixel \( p_{ij} \) multiplies by a correction coefficient \( k_{ij} \) based on an alpha parameter that takes values between 0 and 1—defined by the user so that:

\[
\begin{align*}
  k_{ij} &= \alpha + (1 - \alpha)d_{ij}
\end{align*}
\]

where \( d_{ij} \) represents the distance relative to pixel \( p_{ij} \) at the rim of the breast. In Fig. 2 the graphic interface of the application is shown.

**Statistical analysis**

To estimate the inter-observer concordance in both stages 10% of the images were randomized and analyzed twice with a 2-month separation between both readings. In both stages the inter-observer concordance was estimated among pairs of radiologists and then we compared the concordance averages of the visual and semiautomatic methods. For the results expressed as categories—visual categorization, the concordance was analyzed through the Kappa (\( k \)) coefficient with quadratic weights and 95% intervals of confidence (IC). For the results expressed in the continuous-scale score (DM-Scan) the interclass correlation coefficient (ICC) was used. To assess the Kappa coefficient a weighted equation with quadratic weights was used so that disagreements in the most distant categories have a greater penalty. Given two (2) categories \( i, j \), the weighting factor \( W_{ij} \) used was:

\[
W_{ij} = 1 - \frac{(i - j)^2}{(N - 1)^2}
\]

where \( N \) represents the number of categories. The Kappa coefficient with quadratic weights is statistically comparable to the ICC that in turn allows us to make comparisons between the two (2) methods studies.

**Figure 1**  Boyd/BI-RADS® visual classification.
The statistical study was done with a free software R (R Development Core Team, 2011). R is a free software environment for the analysis of statistical results. To estimate concordance the irr\textsuperscript{21} package was used.

**Results**

**Categorization through visual inspection**

In the inter-observer concordance study the average concordance was 0.876 (95% CI: 0.873–0.879) in the Boyd scale and 0.823 (95% CI: 0.818–0.829) in the BI-RADS\textsuperscript{®} scale. The average intra-observer concordance was 0.813 (95% CI: 0.796–0.829) in the Boyd scale and 0.770 (95% CI: 0.742–0.797) in the BI-RADS\textsuperscript{®} scale. When comparing both scales the average concordance seen is slightly higher with the Boyd scale and it is remarkable that the intra-observer concordance is lower than the inter-observer concordance in both scales.

**Semiautomatic categorization**

In the analysis done through DM-Scan the average inter- and intra-observer concordance was ICC = 0.92 (95% CI: 0.916–0.928)–remarkably better than the concordance with respect to the visual categorization. In this case the intra-observer concordance was the same as the inter-observer concordance.

**Discussion**

This study showed that a computer application automating the determination of breast density obtains excellent intra- and inter-observer concordances while improving visual inspection.

Today breast density can be estimated visually but there is variability among radiologists or in one radiologist only. The concordances reported in former studies go from moderate to really good concordances like \(k=0.54\) in Ciatto
et al.’s study, k = 0.77 in Ooms et al.’s study, k = 0.85 in Perez-Gómez et al.’s study and k > 0.90 in Garrido-Estapa’s study where only one radiologist categorized the same group of mammographies into 4 grades. The variability is not easy to explain but it can be due to the various methods used since the studies are different in the number of mammographies analyzed, the number of radiologists involved (between 1 and 12), the radiologist’s own experience, the pre-study training, the type of mammography (analogical or digital) and the probability of not coinciding–variability that is when assigning a quantitative category to the images found in the barrier that separates 2 categories. Also various studies do not accurately specify the statistical method used to estimate the Kappa coefficient that makes it hard to compare the results found. We can say that the visual estimate of breast density-subject to a high level of subjectivity is not easy to do. Yet despite the intrinsic difficulty of this task our visual concordances were included between the k coefficients of 0.61 and 0.80 so they are good.

Unlike other studies it is surprising that our average intra-observer concordance in the visual analysis is lower than the inter-observer concordance. It is remarkable that both studies have not been done on the same sample in the intra-observer study only 10% of the images were used. Also, the inter-observer study was done immediately after the previous training while the intra-observer study was done months later—something that could justify a minor adjustment to the criteria of categorization agreed during the previous training. However, the difference vanished when the DM-Scan screening was done that other than obtaining excellent concordance rates it showed no discordances in the inter- and intra-observer variability regardless of the passing of time. So the estimation of breast density can be more accurate and objective with computer applications capable of determining breast density automatic or semi-automatically.

One of the first authors to suggest the use of computers as tools to determine breast density was Boyd who introduced a semi-automatic method (Cumulus) to estimate breast density based on the manual selection of 2 thresholds for the segmentation of breast and dense tissue respectively. Unlike other applications the application of DM-Scan makes an initial estimate of the percentage of breast tissue that can be accepted or modified which speeds up and facilitates its treatment while reducing the subjectivity of manual manipulation of the measure. It also corrects bright differences due to breast thickness and not density.

Among the limitations of this study we have to bring up that we did not estimate the extra time necessary to add the estimation of breast density to daily routine that other former studies have estimated between 18 and 40 per projection and 5–8 min per study. As it happens with other semi-automatic tools for the estimation of breast density and even if they can improve the results of visual inspection there is a certain degree of subjectivity since they need one operator to set the threshold to separate dense from fat tissue—something that can induce certain variability in the estimation. But this is not a real limitation given our goal was to estimate it for the semi-automatic measure too.

With a significantly higher number of mammographies than the rest of studies published our results confirm those obtained in other studies done through visual or semi-automatic categorization or when comparing both visual and semi-automatic categorizations. Yet despite both modalities, visual inspection and DM-Scan are valid to estimate breast density, the DM-Scan modality is more precise, reduces subjectivity significantly, is more reliable to establish breast density and consequently allows us to homogenize criteria and helps us to design more adequate screening protocols based on breast density.

In sum our study confirms that the estimation of breast density with the semi-automatic DM-Scan application is reliable and easily reproducible while reducing the subjectivity and variability of visual inspection.

Ethical responsibilities

Protection of people and animals. Authors confirm that for this investigation no experiments with human beings or animals have been carried out.

Data confidentiality. Authors confirm that there are no personal data from patients in this article.

Right to privacy and informed consent. Authors confirm that they have obtained the written informed prior consent from patients and/or subjects appearing in this article. This document is in the possession of the corresponding author.

Authors

1 Manager of the integrity of the study: IMG.
2 Original Idea of the Study: IMG and MCE.
3 Study Design: IMG and MCE.
4 Data Mining: IMG, MCE, FRP, R Li and JA.
5 Data Analysis and Interpretation: IMG, MCE, FRP, R Li and JA.
6 Statistical Analysis: R Li and JA.
7 Reference Search: MCE and IMG.
8 Writing: IMG, MCE, R Li, FRP and JA.
9 Manuscript critical review: MCE, FRP and R Li.
10 Final Version Approval: IMG, MCE, FRP and R Li.

Conflict of interests

Authors reported no conflicts of interests.

Acknowledgements

We wish to thank the following institutions: Centro Nacional de Salud: Instituto de Salud Carlos III de Madrid, Fundación Gent x Gent (EDEMAC Project), Fondo Español de Investigación Sanitaria and the Federación Española de Pacientes con Cáncer de Mama (FECMA) for financing the project Determinants of Density in Mammography in Spain (DDM-Spain)—origin of the mammographies presented in this study.
We also wish to thank other members of DDM-Spain: Pilar Moreo, Pilar Moreno y Soledad Abad (Aragón, Spain); Francisca Collado y Magdalena Moyá (Balears, Spain); Isabel González, Carmen Pedraz and Francisco Casanova (Castilla-León, Spain); Mercé Peris (Cataluña, Spain); Carmen Santamaría, José Antonio Vázquez Carrete, Montserrat Corujo and Ana Beién Fernández (Galicia, Spain); Nieves Ascunce, María Ederra, Milagros García and Ana Barcos (Navarra, Spain); Manuela Alcaraz, Jesús Vioque (C. Valenciana, Spain); Virginia Lope, Nuria Aragonès and Anna Cabanes (Madrid, Spain).

Also thanks to the Instituto Tecnológico de Informática de la Universidad Politécnica de Valencia for designing and producing the DM-Scan for automatic and semi-automatic readings and to doctors Beatriz Pérez-Gómez and Marina Pollán from the Enviromental Epidemiology and Cancer Unit of the Centro Nacional de Epidemiología, Instituto de Salud Carlos III, and doctors Dolores Salas and Josefa Miranda from the Dirección General Salud Pública, Programa de Cribado de Cáncer de Mama, Centro Superior de Investigación Salud Pública (CSISP) (Valencia, Spain), and doctor Carmen Palop (Unidad de Prevención de Cáncer de Mama de Sagunto, Valencia, Spain) for their collaboration and help during this study.

References
