ORIGINAL REPORT

Impact of mammographic breast density on computer-assisted detection (CAD) in a breast imaging department

C. Romero Castellano a, *, C. Varela Nuñez b, R. Cuena Boy c, A. Almenar Gil a, J.M. Pinto Varela a, M. Botella Lopez d

a Unidad de Mama, Servicio de Radiología, Hospital Virgen de la Salud, Toledo, Spain
b Área de Información Tecnológica, Servicio de Salud de Castilla La Mancha, Toledo, Spain
c Área de Investigación, Hospital Virgen de la Salud, Toledo, Spain
d Departamento de Antropología Física, Universidad de Granada, Granada, Spain

Received 2 May 2010; accepted 25 August 2010

Abstract

Objective: To evaluate whether breast density influences the sensitivity of a computer-assisted detection (CAD) system for the detection of breast cancer.

Material and methods: We prospectively studied 8750 digital mammograms with an associated CAD system. We used BI-RADS criteria to classify breast density. We calculated the overall sensitivity of the radiologist and of the CAD system, as well as the sensitivity for each projection and type of finding in relation to the mammographic density of the breast. Finally, we analyzed the interval carcinomas. We used SPSS 11 for all statistical analyses.

Results: The overall sensitivity of the CAD system was 88.5% (95% CI: 83.2–92.7%), and the overall sensitivity of the radiologist was 93.5% (95% CI: 84.4–95.5%). The sensitivity of the craniocaudal view was 81.6% (95% CI: 76.5–90.7%) vs 76.5% (95% CI: 69.3–89.3%) for the mediolateral oblique view. The sensitivity for microcalcifications was 98.6% (95% CI: 96.5–99.7%), and the sensitivity for masses was 83.4% (95% CI: 81.2–91.7%). We detected discrepancies smaller than 20% both for microcalcifications present in the four types of densities and for masses with densities 1 and 2. In masses with density 3 the discrepancy was 20.8% and in those with density 4 it was 55%. The CAD system failed to mark only 9.1% (9/94) of the cancers presenting as masses. Half of the interval carcinomas were found in type 4 density and 75% manifested as masses, asymmetries, and distortions. The CAD system had marked 35.7% of the carcinomas.

Conclusions: The craniocaudal view was more sensitive, although this difference was not statistically significant. The sensitivity of CAD was high for microcalcifications in all four density types; however, CAD’s sensitivity for masses was low in density types 3 and 4. The CAD system only failed to mark 9.1% of the cancers presenting as masses but was not sensitive for the other two radiological findings included in this marking. Half of the interval carcinomas occurred in type 4 densities and 35.7% had been marked by the CAD system.

© 2010 SERAM. Published by Elsevier España, S.L. All rights reserved.
Impacto de la densidad mamaria mamográfica en el funcionamiento de un sistema de detección asistido por ordenador en una unidad de patología mamaria

Resumen

Objetivo: Evaluar si la densidad mamaria influye en la sensibilidad (global y por marcas) para la detección del cáncer de mama de un sistema de detección asistido por ordenador (CAD).

Materiales y métodos: Estudio prospectivo de 8.750 mammografías digitales con un sistema CAD asociado. Se clasificaron las densidades mamarias según los criterios BI-RADS. Calculamos la sensibilidad global del radiólogo y del CAD, la sensibilidad por proyección, por hallazgo, en relación con la densidad mamográfica y analizamos los carcinomas de intervalo. Para el análisis estadístico utilizamos el programa SPSS vs 11.

Resultados: Sensibilidad global del CAD 88,5% IC del 95% (IC95% 83,2–92,7%), sensibilidad del radiólogo 93,5% IC95% (84,4–95,5%), sensibilidad de la proyección craneocaudal 81,6% IC95% (76,5–90,7%) vs 76,5% IC95% (69,3–89,3%) para oblicuomedialateral, sensibilidad para microcalcificaciones 98,6% IC95% (96,5–99,7%), sensibilidad de marca masa 83,4% IC95% (81,2–91,7%). Detectamos discrepancias menores del 20% tanto para las microcalcificaciones presentes en los 4 tipos de densidades como para las masas con densidades 1 y 2; mientras en las masas con densidad 3 la discrepancia fue 20,8% y en la 4 fue 55%. El CAD solo dejó de marcar el 9,1% (9/94) de los cánceres diagnosticados como masas propiamente dichas. El 50% de los carcinomas de intervalo se produjo en densidad tipo 4 y el 75% se manifestaron como masas, asimetrías y distorsiones. El 35,7% de los carcinomas de intervalo fueron marcados previamente por el CAD.

Conclusones: La sensibilidad fue mayor en la proyección craneocaudal pero no significativamente. La sensibilidad del CAD fue alta para microcalcificaciones presentes en los 4 tipos de densidades, sin embargo, para la marca masa fue baja en densidades 3 y 4. El CAD sólo dejó de marcar el 9,1% de los cánceres diagnosticados como masas propiamente dichas pero fue muy poco sensible para los otros 2 hallazgos radiológicos incluidos en esta marca. El 50% de los carcinomas de intervalo se produjeron en densidades tipo 4 y el 35,7% fueron marcados previamente por el sistema de detección asistido por ordenador.

© 2010 SERAM. Publicado por Elsevier España, S.L. Todos los derechos reservados.

Introduction

Breast cancer is the most common cancer type in women under 50 and is responsible for 17.5% of cancer deaths in women.1,2

The prognostic factors established by the TNM classification system show that the 5-year survival rate drops from 90% in stage I to 57% in stage II (A or B). This means that early breast cancer diagnosis is as important as diagnosis itself since the sooner the tumor is detected, the smaller it will be.3,4

For this reason, breast cancer screening campaigns have been launched in many countries and new technological advances, such as digital mammography, have been implemented.5 Digital mammography has been shown to have a higher diagnostic sensitivity in young women where the breast normally has a higher concentration of glandular tissue (BI-RADS 3 and 4).5,6 These dense breasts entail not only complex diagnosis because of poor visualization of the radiological features, but also a higher breast cancer risk.8

Furthermore, the use of digital mammograms has allowed for the development of Computer-Aided Detection (CAD) systems.

CAD is designed to identify and mark a potentially suspicious region, alerting the reader to areas missed or misinterpreted on the mammogram. CAD is therefore not intended to replace the radiologist, but acts as a second-opinion tool that intends to solve problems related to fatigue and lack of concentration and to avoid diagnostic errors.

However, the literature does not clearly explain how breast density affects the performance of the new digital mammography CAD systems or whether breast density affects the two mark types equally. It must not be forgotten that young women will especially benefit from CAD since these breasts have higher proportion of glandular tissue. Therefore, our goal is to evaluate whether breast density influences (global and mark-based) sensitivity of a CAD system for breast cancer detection.

Material and methods

Our breast disease unit is a multidisciplinary team made up of three radiologists, three pathologists, four breast surgeons, two plastic surgeons, one gynecologist, two oncologists, one radiotherapist, one physical therapist, one psychiatrist, and two psychologists.

Women aged 35–70 years were prospectively studied. Data were collected between January and December 2007. Patients were then followed for 2 years in order to detect the interval carcinomas. The study was approved by the ethics committee of our institution.

A total of 8750 digital mammograms were performed with a mammograph (MAMMOMAT® NovationDR Siemens, Germany) with breast-specific workstations. We used a D R2...
Technology ImageChecker CAD system (version 5.4) installed in August 2006. This system provides direct digital reading (raw data), and detects suspicious mammographic findings signalled by two mark types, by means of mathematical algorithms:

1. A micro mark for microcalcifications as a radiological finding (a triangle in our version).
2. A mark for masses, which includes three radiological findings: distortion, asymmetry, and actual masses (an asterisk in our study).

The readers, who performed single reading, were two radiologists specializing in breast with 8- and 4-year experience, respectively. A 6-month training period with the CAD system was completed before engaging in the data analysis (performed between August and December, with a mean reading volume of 2740 mammograms per radiologist).

For all cases, a team of technicians specially trained in mammographic imaging obtained two craniocaudal (CC) and mediolateral oblique (MLO) views of each breast. Only patients with tumors visible on the mammograms on at least one of the views were included in the study. Two male patients with breast cancer were excluded because of lack of the four views and two pregnant women with breast cancer were excluded because the lesions were not mammographically visible. Similarly, none of the cases diagnosed in other institutions but that were referred to us were included in the study.

For all cases, breast density was classified according to the American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS)\(^8\): type 1 density (glandular tissue is <25%), type 2 (between 25% and 50% of the breast), type 3 density (between 50% and 75% of the breast), and type 4 density (>75% of glandular tissue).

Data were collected initially on a sheet that divided the CC and MLO views of both sides into 1 cm in wide grids. These grids were numbered 1–30. The suspicious areas detected by the radiologist were marked on the grids and BI-RADS categories were assigned on the mammogram without CAD.

The mammogram was subsequently read by the CAD system, which recorded new suspicious areas or changed initial diagnosis of certain areas. A true positive CAD mark was considered when the location of the mark, on at least one-view of the mammogram, was over the lesion detected by the radiologist (CAD-marked lesions were within the same grid square on the CAD output and the suspicious area, therefore, always less than 1 cm away suspicious area). A false positive CAD was considered when the location of the mark was not over the grid marked by the radiologist and no interval carcinoma was detected after follow-up.

All of the carcinomas underwent Tru-cut or vacuum-assisted percutaneous biopsy, and were histopathologically confirmed. Final staging was completed by MRI in all cases. Diagnosis of exclusion of breast carcinoma was obtained from negative biopsies or from a negative mammographic follow-up after 2 years. CAD algorithms are not designed to mark lesions >2 cm as these lesions are expected to be detected by the radiologist. Thus, this study is only concerned with tumors <2 cm.

The follow-up of the cases reported lasted 26–30 months at the time this study was finished. The radiologist’s and CAD global sensitivities were calculated. Sensitivity of the CAD for each view as well as for each finding was also calculated. Finally, sensitivity was estimated in relation to mammographic density.

The results are given in percentages and their corresponding 95% confidence intervals (95% CI) based on the exact limits of the binomial distribution. Agreement between the radiologist and the CAD results regarding the anatomical location of the lesion is assessed the unweighted kappa index. Kappa significance is estimated with the chi-square test. Finally, to determine whether breast density and lesion type seen by the radiologist are factors affecting independently the percentage of agreement between the radiologist and the CAD system the odds ratios (OR) and their 95% CI are estimated with adjustment of an unconditional logistic regression model where the dependent variable is the agreement or lack of agreement regarding the lesion type, and the independent variables are density and lesion type diagnosed by the radiologist. To this aim, the software SPSS v 11 was used.

**Results**

A total of 8750 mammograms were analyzed in our hospital between January and December 2007. 74.5% of them were screening mammograms (6518) and 25.5% (2232) were diagnostic mammograms.

There were no statistically significant differences in the mammogram distribution for the four density types (p = 0.2) (Fig. 1).

A total of 205 carcinomas were diagnosed and 192 of them were mammographically visible. Out of these 192, two male carcinomas were excluded from the study because of the lack of the four views and two carcinomas were excluded because they were diagnosed in other center and referred to our hospital for a second opinion (n = 188).

CAD overall sensitivity was 88.5% 95% CI (83.2–92%) and the radiologist sensitivity was 93.5% 95% CI (84.4–95%). Sensitivity for the CC view was higher than that for the MLO view (81.6% [76.5–90.7%] vs 76.5% [69.3–89.3%]), though this difference was not significant (p = 0.8). It was shown that 12% (23/192) of our carcinomas were only marked on one view (mostly on CC view [14/23]), although they were visible on both views. The level of agreement between the CAD system and the radiologist regarding anatomic localization was very

![Figure 1](http://www.elsevier.es) Mammogram distribution according to density.
high for both views, reaching Kappa values of 0.82 and 0.75 for CC and MLO view, respectively \((p < 0.05\) in both cases).

Overall CAD sensitivity per mark was 98.6\% 95\% CI \((96.5–99.7\%\) for microcalcification marks and 83.4\% 95\% CI \((81.2–91.7\%\) for mass marks (Fig. 2).

Regarding the percentage of discrepancy between CAD detection and radiologist (i.e. carcinomas detected by the radiologist that CAD was not able to detect and vice versa), and using the algorithms separately, our analysis revealed that when the radiologist diagnosed a mass, CAD was in disagreement in 19.5\% of the cases. This discrepancy decreased to 1.8\% when the radiologist diagnosed microcalcifications \((p = 0.003)\). Consequently, the mathematical algorithm seems to be more sensitive for microcalcification detection than for mass detection.

The next step was to correlate this analysis to breast density. Our results showed that overall CAD discrepancy was 1.5\% for carcinomas detected on mammograms with type 1 density, and 2.5\% on type 2 density. For type 3 and type 4 densities, the discrepancy was 16.5\% and 31.3\%, respectively, and were statistically significant \((p = 0.003)\).

Next, the algorithms were used separately again for mass and microcalcification detection in order to estimate the association of CAD-radiologist discrepancy with breast density and lesion type. The logistic regression showed that adjusted ORs were significant for microcalcifications diagnosed by the radiologist \((OR = 0.07; 95\% CI [0.015–0.333])\) and for all density types 1, 2 and 3 with respect to type 4. For type 1, OR = 0.161 (95\% CI [0.043–0.604]). For type 2, OR = 0.029 (95\% CI [0.003–0.246]). For type 3, OR = 0.256 (95\% CI [0.097–0.674]) \((Table 1)\).

Finally, the independent effect that both density and lesion type had on the probability of discrepancy between CAD and radiologist detection was shown as percentages lower than 20\% both for microcalcifications of any density type and for masses with density types 1 and 2. In contrast, discrepancy reached 20.8\% (95\% CI [10.8–34.1\%]) and 55\% (95\% CI [31.5–76.9\%]) for masses with density types 3 and 4, respectively.

For the analysis of the mass mark, the three findings involved (asymmetry, distortion, and actual masses) were separated. The authors calculated the percentage of carcinomas radiologically associated with each of these findings that had not been marked by CAD. Overall, CAD only failed to mark 9.1\% (9/94) of the cancers manifesting as masses, 17.6\% (14/21) as distortions, and 47.9\% (7/15) as asymmetries (Fig. 3).

Five carcinomas were marked exclusively by CAD. The analysis of the interval carcinomas showed that 3.58\% (1/28) of them occurred in type 1 density, 10.8\% (3/28) in type 2, 35.8\% (10/28) in type 3, and 50\% (14/28) in type 4. Despite the increased percentage of interval carcinomas detected as breast density increased, no statistically significant differences were found between the four density types \((p = 0.6;\) Fisher’s exact test). Nonetheless, we acknowledge that may be due to a sampling problem. As much as 75\% (21/28) of the interval carcinomas were manifested as masses, asymmetries and distortions.

### Table 1  Discrepancy between the Computer-Aided Detection (CAD) system and the radiologist according to lesion type and density.

<table>
<thead>
<tr>
<th>CAD marks</th>
<th>Breast density</th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
<th>Type 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Microcalcifications</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrepancy</td>
<td></td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>9.5%</td>
</tr>
<tr>
<td>No. = 57</td>
<td></td>
<td>8/8</td>
<td>8/8</td>
<td>20/20</td>
<td>19/21</td>
</tr>
<tr>
<td><strong>Masses</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrepancy</td>
<td></td>
<td>3.1%</td>
<td>15.4%</td>
<td>20.8%</td>
<td>55.0%</td>
</tr>
<tr>
<td>No. = 131</td>
<td></td>
<td>31/32</td>
<td>23/26</td>
<td>42/53</td>
<td>9/20</td>
</tr>
<tr>
<td>OR(^a)</td>
<td></td>
<td>0.061</td>
<td>0.029</td>
<td>0.134</td>
<td>0.256</td>
</tr>
</tbody>
</table>

\(^a\) Odds ratios that are significant for masses and all density types 1, 2 and 3 with respect to density type 4.
Discussion
The CAD system is based on two mathematical algorithms (one for microcalcification detection and the other for mass detection). Once the interval carcinomas were analyzed, overall sensitivity of our CAD system was 88.5% (95% CI [83.2–92.7%]). However, according to our experience and to previous studies, the sensitivity is higher—though not significant—on the CC view. It should be noted that 12% of the carcinomas in our series were only detected on one view in spite of the fact that they were mammographically visible on both views. For this reason, in our opinion, each of the marks provided by the CAD should be assessed separately. Moreover, the fact that a lesion is only marked on one view does not entail a lower probability of malignancy. Improvement of the algorithm is thus necessary to avoid false positives, which hinder and delay the radiologist reading.

Breast density may potentially influence the performance of CAD algorithms. In fact, this was the case for our data since probabilities for CAD and radiologist to provide discordant diagnoses of lesion type increased significantly for density 3 and 4 cases with respect to the rest of cases (discrepancy of 16.5% and 31.3%, respectively) (p = 0.003).

This seems reasonable, but the question is whether breast density is equally determinant for both mark types. Breast density should be expected to affect in lesser degree the detection of microcalcifications, which are much denser than the rest of the parenchyma, than the detection of masses, whose borders can be confounded by the adjacent glandular tissue.

In our study, the CAD-radiologist discrepancy rate decreased to 1.8% when the radiologist diagnosed microcalcifications, regardless of density (p = 0.003). However, this rate reached its peak value for masses in density 4 breast (discrepancy of 55% [95% CI [31.5–76.9%]) (Table 1).

There are few research studies that address this issue. One of the first studies to report the significant influence of breast density on cancer detection rates was performed by Ho and Lam in 2003. Since they used the first software applications available on the market, the sensitivity rates reported were very low.

More recent researches based on more advanced software report similar results. For instance, a study published by Malich et al. in 2005 evaluated 200 mammograms and 127 of them revealed histologically proven carcinomas. In this study, breast density was divided into four groups according to the BI-RADS density patterns. The study reported the following rates of detection of microcalcifications and masses: density 1, 93.7% and 94.7%; density 2: 97% and 96.1%; density 3: 66.3% and 79.1%; and density 4: 61% and 33.3%, respectively.

The study concluded that neoplasms manifesting as microcalcifications are not as significantly affected by breast density as masses are. However, the sensitivity rate for types 3 and 4 is not acceptable for either of the two marks. Also in 2005, Bret et al. reported different results from a sample of 809 CAD-detected breast carcinomas. They divided the mammograms into two groups according to the BI-RADS classification: non-dense (types 1 and 2) and dense (types 3 and 4) mammograms. These authors concluded that breast density does not affect overall CAD sensitivity (90% for non-dense and 88% for dense mammograms). It was also shown that breast density does not affect sensitivity either (95% in non-dense mammograms and 93% in dense mammograms) when carcinomas manifest as microcalcifications. Nevertheless, breast density does affect mass detection significantly (p = 0.03), showing a decrease in sensitivity from 88% in non-dense mammograms to 79% in dense mammograms.

In comparison with previous works, our study is prospective in nature, and is based on a CAD system with direct digital reading and that includes interval carcinomas, resulting thus in conclusive sensitivity results. In our analysis, sensitivity was much higher for microcalcification marks (96.6%, 95% CI [96.5–99.7%]) than for mass marks (93.4%, 95% CI [81.2–91.7%]). Moreover, it is argued that the probability of CAD-radiologist discrepancy (OR = 0.07; 95% CI [0.015–0.333]) is higher for mass-type lesions than for microcalcifications. The CAD system could therefore assist radiologists in the detection of breast cancer at very early stages, such as in situ carcinoma, since the presence of microcalcifications is the most common imaging finding and CAD is especially wellsuited for microcalcification detection.

In contrast, mass detection algorithms seem to yield much lower sensitivity rates. As a consequence, the mass mark was analyzed by first classifying results according to the three types of radiological finding involved (asymmetries, distortions, and masses themselves). This was done in order to find out whether CAD is actually less sensitive for mass detection or its sensitivity is affected by its low detection rate for asymmetries and distortions.

According to our results, the lower sensitivity of CAD for mass mark seems to be due to its low sensitivity for marking carcinomas that manifest as asymmetries and distortions. In fact, while CAD failed to mark 17.6% (4/21) of carcinomas diagnosed as distortions and 47.9% (7/15) of carcinomas diagnosed as asymmetries, it only failed to mark 9.1% of carcinomas diagnosed as actual masses.

Although our study is not based on a large sample size, the 28 interval carcinomas detected showed an upward trend as breast density increased. Nevertheless, no statistically significant differences were found between the four density types.

In conclusion, overall sensitivity of our CAD system was similar to that reported by the references provided in this paper. Sensitivity was higher—though not significantly—for the CC view and 12% of the carcinomas were only marked on one view, although they manifested radiologically on both views.

CAD sensitivity was very high for microcalcifications and for all four density types. However, breast density affects CAD sensitivity for mass lesions much more negatively. The cumulative effect of both factors (lesion type and sensitivity) led to a sensitivity rate lower than 50% for the mass mark on density 4 breast.

After separately analyzing the mass mark for each of the radiological findings involved, the CAD system only failed to mark 9.1% of the carcinomas manifesting as masses but had low sensitivity to mark the two other radiological findings (asymmetries and distortions), which affected the overall sensitivity rate of the mass mark. Fifty percent of the
interval carcinomas occurred in type 4 density and 35.7% were previously marked by the CAD system.

Authors

Cristina Romero, the main author, has contributed to drafting the manuscript, acquisition and analysis of results.

Celia Varela is responsible for the study design, together with Dr. Cristina Romero with intellectually relevant contributions.

Rafael Cuena is responsible for the statistical analysis.

Asunción Almenar and Dr. Romero took part in the acquisition of data.

J.M. Pinto, Dr. Romero and Celia Varela contributed to the analysis and interpretation of data.

Miguel Botella, director of the PhD thesis which this manuscript belongs to, was an active participant in the manuscript, providing a critical review and final approval.

All authors have read and approved the final version of the manuscript.

Conflicts of interests

The authors declare no conflict of interests.

References