Malignant Lesions on Mammography: Accuracy of Two Different Computer-Aided Detection Systems

Marc Lobbes, Marjolein Smidt, Kristien Keymeulen, Rossano Girometti, Chiara Zuiani, Regina Beets-Tan, Joachim Wildberger and Carla Boetes
No responsibility is assumed by Elsevier, its licensors or associates for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made.
Malignant lesions on mammography: accuracy of two different computer-aided detection systems

Marc Lobbes,⁎ Marjolein Smid, Kristien Keymeulen, Rossano Girometti, Chiara Zuiani, Regina Beets-Tan, Joachim Wildberger, Carla Boetes

⁎Corresponding author. Maastricht University Medical Center, Department of Radiology, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands. Tel.: +31 43 387 6910; fax: +31 43 387 6909.
E-mail address: marc.lobbes@mumc.nl (M. Lobbes).

Received 17 March 2012; accepted 6 April 2012

Abstract

We retrospectively compared the accuracy of two computer-aided detection (CAD) systems for the detection of malignant breast lesions on full-field digital mammograms. Mammograms of 326 patients were analyzed (117 patients with breast cancer, 209 negative cases), and each set of cases was read by two CAD systems (Second Look versus AccuDetect Galileo). True-positive fractions per image and case for soft densities, microcalcifications, and total cancers were assessed. Study results showed better overall performance of AccuDetect Galileo (when compared to Second Look) in detecting masses, microcalcifications, and all cancer types, especially in extremely dense breast parenchyma.

© 2013 Elsevier Inc. All rights reserved.

Keywords: Breast imaging; Breast neoplasms; Mammography; Computer-aided detection; Early detection of cancer

1. Introduction

During mammographic evaluation of the breasts, computer-aided detection (CAD) systems can be used to increase diagnostic accuracy. Using complex computerized algorithms, CAD systems mark culprit regions on mammographic images to attract the reader’s attention to certain suspicious features that might be overlooked otherwise.

Nonetheless, the added value of CAD systems for reading mammograms remains controversial. There are multiple studies showing beneficial results, for example, in screening mammography [1–5]. In contrast to these favorable results, there are also many studies to be found that question the value of CAD systems [6,7]. These conflicting results show that the added value of CAD for evaluation of mammograms is still under debate. With this respect, the algorithms of CAD systems are constantly improving to acquire higher accuracies. Our study aim was to retrospectively compare the accuracy of two computer-aided detection (CAD) systems for the detection of malignant breast lesions on full-field digital mammograms: the widely used and Food and Drug Administration (FDA)-approved Second Look versus AccuDetect Galileo, which differs from traditional CAD systems in that it uses a newly developed voting methodology approach for detecting culprit features on mammograms. In this voting methodology, there are two recognizers in its external voting scheme: one recognizer is Galileo; the other is AccuDetect, which harbors seven proprietary recognizers inside that are used for internal voting. Although this study does not evaluate the voting methodology itself, favorable results of this approach compared to widely accepted CAD systems might open the gate for new generations of CAD systems that will deliver better performance than other approaches.

0899-7071/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.clinimag.2012.04.017
2. Materials and methods

In a period of 2 years, a total of 326 screening mammograms were acquired from two identical full-field digital mammography units (Giotto Image, IMS Internazionale Medico Scientifica, Bologna, Italy). Acquisition of informed consent was waived by a certified medical ethics committee. Inclusion criteria were female sex, any ethnic origin, and the availability of bilateral two-view mammogram. Excluded were patients with significant existing breast trauma, breast implants, pregnancy, lactation, and prior surgical biopsy, breast cancer, and breast marker placement. This data set encompassed a total of 117 cancer cases and 209 negative cases. Candidate positive cases were the exams with a mammographically visible abnormality that proved to be a malignancy. All positive cancer cases were therefore histopathologically proven malignancies, whereas all negative cases were confirmed by benign findings at biopsy or follow-up imaging, or by a minimum of 12 months of follow-up. Negative cases were used for the calculation of the false positives per image and per case, in which false-positive rates are the average number of false positives per image or per case.

All images were from female subjects, with age ranging from 30 to 96 years old. The set of positive cases consisted of 85 cases of only masses, 6 cases of only microcalcifications, and 26 mixed cases (consisting of both masses and microcalcifications). They consisted of invasive ductal carcinoma (59.4%), invasive lobular carcinoma (17.2%), ductal carcinoma in situ (7.8%), invasive ductal–lobular carcinoma (a pattern of tumoral growth originating from both lactiferous ducts and breast lobules, 4.7%), lymphoma (2.3%), lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and other malignancies (7.0%), such as papillary cancer, metastatic carcinoma or lobular carcinoma in situ (1.6%), and o
at 0.6 (P = .006), and per case difference of 4.3%, with FPC at 2.5 (P = .016) (Table 3).

In a subanalysis, performances of both CAD systems in extremely dense breasts (ACR 4) versus fatty to moderately dense breasts (i.e., ACR 1 to 3) were compared. For masses, Second Look’s TPF per case was 53.8% versus 69.2% for Detect Galileo (FPI for both systems was set at 0.42, P = .016). For calcifications, TPF for Second Look was 61.5% versus 76.9% for Detect Galileo, with similar operating points for both systems (FPI set at 0.2). However, this increase was not statistically significant, probably due to the limited number of cases with calcifications. Most importantly, for total cancer cases in extremely dense breasts, Detect Galileo’s TPF was 70.7% with the FPI set at 0.62 versus only 56.1% for Second Look with the FPI set at 0.64 (P = .016) (Fig. 1).

4. Discussion

Computer-aided detection systems, or CAD systems, use complex computerized algorithms to evaluate mammograms and assist the reader in their evaluation. CAD systems mark culprit regions such as masses, architectural distortions, or microcalcifications on mammographic images to attract the reader’s attention to certain suspicious features that might be overlooked. In the past, the positive value of CAD has been demonstrated in several studies, but there are also studies that question the added value of CAD in general [1–7]. This emphasizes the need for CAD systems to evolve and increase their accuracy. In this retrospective comparison study, we evaluated the accuracy of two CAD systems (Detect Galileo and Second Look) and demonstrated that Detect Galileo showed better overall performance than Second Look in detecting masses, microcalcifications, and all cancer types, especially in extremely dense breasts.

The first CAD system for mammography was approved by the United States FDA in 1998 (R2 Imagechecker, Hologic, Bedford, MA, USA). To the best of our knowledge, two more CAD systems were approved by the FDA since that year: Kodak Mammography CAD ENGINE (Carestream Health Inc., Rochester, NY, USA) for film-based CAD assessment and iCAD’s Second Look for both film and digital CAD assessment. However, there are more CAD systems commercially available, and the results of our study demonstrate that not all CAD products are alike. In our retrospective comparison study, Detect Galileo showed improved performance over Second Look in detecting masses, microcalcifications, and all cancer types, especially in extremely dense breasts.

In general, to improve CAD systems, the number of false-positive marks needs to be significantly decreased. Traditionally, CAD systems have focussed on the early detection of cancer rather than the elimination of the false-positive markers. Second Look detects culprit lesions on mammograms using image processing, pattern recognition, and artificial intelligence techniques based on the knowledge from thousands of mammograms. The improved performance of the Detect Galileo CAD system over the well-validated Second Look program is mostly due to the software algorithms on which Detect Galileo program is based. In contrast to Second Look, it uses a unique voting system methodology to deliver higher accuracy of the CAD results. This methodology finds its origin in postal service automation (reading information from imaged mail pieces), but can also be applied in medical imaging solutions. In short, the voting system methodology reduces the false-positive rates by combining the results of multiple algorithms and employing a voting mechanism that considers each of these results. Each image recognition process identifies areas of interest on the mammogram independently, without sharing information with the other recognition process(es). Next, the different areas of interest can be compared to determine a confidence value related to the accuracy of the identifications. Comparing the results of the multiple image recognition processes allows for the mitigation of the inherent faults of the imaging recognition
process, thus potentially leading to reduced false-positive and false-negative rates [9].

It is remarkable that a highly significant increase in CAD performance for AccuDetect was observed for masses or microcalcifications per image, but per case, this improvement was not statistically significant. We hypothesize that Second Look might suppress some hypotheses if their confidences are significantly lower than the best hypothesis for the case. This leads to reduction of false positives per case without a significant decrease of recognition per case and also to possible significant decrease of recognition per image. To put it more simply, it means that, in many cases, Second Look recognizes masses on only one projection. This strategy might be adopted by Second Look for screening programs, but we are not sure since we are not able to look into the software algorithms of Second Look. However, the improvement in accuracy when using AccuDetect Galileo was statistically significant for all cancers. In addition, the strength of the accuracy of AccuDetect lies in the breasts with extremely dense fibroglandular tissue. In these cases, AccuDetect Galileo shows a significant improvement in CAD accuracy over Second Look.

Mammographic breast density is an important issue in breast imaging. McCormack showed in a large meta-analysis of 42 studies that the risk of developing breast cancer is increased in (extreme) dense breasts. This risk can be as high as 4.6-fold when compared with fatty breasts [10]. These results were later confirmed by Boyd et al., showing that the risk of developing breast cancer was increased 4.7-fold in women with extremely dense breasts (>75%) versus women with less than 10% breast density [11]. In addition, Carney et al. demonstrated that the sensitivity and specificity for mammograms of fatty breast were 88% and 97%, respectively. They also showed that this accuracy decreases with increase in breast density. For extremely dense breasts, mammographic sensitivity and specificity decreased to 62% and 90%, respectively [12]. Thus, the use of CAD systems to detect breast cancer in extremely dense breasts is advisable. In literature, only a handful of papers have been published over the years regarding the accuracy of CAD systems in different ACR Breast Imaging Reporting and Data System categories of breast density. Comparison of these studies remains difficult because of the variation of systems and software versions used and due to the variation in patient population, breast density classifications, mammography findings (soft tissue densities, calcifications, architectural distortions, etc.), and histologic results.

In 2001, Ho and Lam were the first to evaluate the influence of breast density on CAD system accuracy, using an early version of Second Look (version 1.1) in a data set of 264 screening mammograms. They showed that CAD system sensitivity significantly decreased as breast density increased, while its specificity remained relatively constant [13]. Probably due to the improvements of later software versions, these initial findings could not be reproduced. For example, Brem et al. evaluated the performance of the CAD system Second Look (version 4) in dense (ACR categories 3 and 4) and nondense (ACR categories 1 and 2) breasts. They showed that the overall CAD detection of breast cancer was

<table>
<thead>
<tr>
<th>CAD system</th>
<th>Cancer FP per image</th>
<th>Cancer TP per image</th>
<th>Cancer FP per case</th>
<th>Cancer TP per case</th>
</tr>
</thead>
<tbody>
<tr>
<td>AccuDetect Galileo</td>
<td>0.62</td>
<td>72.3%</td>
<td>2.48</td>
<td>84.6%</td>
</tr>
<tr>
<td>Second Look</td>
<td>0.64</td>
<td>65.4%</td>
<td>2.56</td>
<td>80.3%</td>
</tr>
<tr>
<td>Difference</td>
<td>0.02</td>
<td>6.9%</td>
<td>0.08</td>
<td>4.3%</td>
</tr>
</tbody>
</table>

*P=.006; **P=.016.

Table 3
True-positive fractions per case and per image for total cancers

Fig. 1. Example of improved accuracy of AccuDetect Galileo (correct, green prompts) versus Second Look (false positive, red prompt) in extremely dense breast parenchyma. Truth data are presented as the dotted orange line and represent the true extension of the biopsy-proven malignancy.
not affected by breast density [14]. In another study of 350 cases, Malich et al. (using Second Look version 5) concluded that overall breast density decreased CAD system accuracy for masses, but not for microcalcifications [15]. Obenauer et al. used ImageChecker version 2.3 from R2 Technology to evaluate 226 mediolateral oblique and 186 craniocaudal mammography views of histologically proven cancers and concluded that there seemed to be only a tendency for breast density to affect the detection rate of breast cancers [16]. Although there seems to be no convincing evidence that breast density affects CAD system performance in more dense breasts, there are no prior studies available that compare the accuracy of two or more different CAD systems in dense breasts, as we did in our current study. This study showed that AccuDetect Galileo significantly outperformed Second Look in ACR category 4, i.e., extremely dense breasts.

Independent of the accuracy of the different CAD systems, many radiologists dislike the use of CAD systems due to the low specificity. They feel that the large amount of false-positives findings presented by CAD decreases their diagnostic confidence when evaluating mammograms. Houssami et al. estimated that there will be approximately 400 false-positive prompts for each true prompt on a cancer in screening mammography. In a review of the literature, Boyer et al. estimated that even 1000 false-positive prompts needed to be evaluated by the reader for each true cancer [17]. Therefore, the ratio of false prompts to true prompts on cancer which the reader had overlooked is high [18]. A possible solution for this problem is the development of so-called interactive prompting in CAD systems, in which CAD marks are only displayed for regions of interest. Using interactive CAD system, Samulski et al. showed a significant increase in detection performance for this system, while total reading times for the mammograms were not influenced. Reading radiologist preferred the use of this interactive approach over conventional CAD use [19]. However, these systems are not (yet) commercially available. Future developments might focus on reprogramming existing and accurate CAD systems to use an interactive approach.

Our study demonstrates that (European) institutions that are interested in purchasing CAD software should pay close attention to performance details when selecting a proper system, and not only focus on FDA approval status. There is no need for these institutions to limit their selection only to systems that were cleared by the FDA. Our current study is an example of the proof that there might be other CAD applications available that might deliver comparable or even significantly better performance results.

There are some limitations to this study. First, we compared AccuDetect Galileo with Second Look version 7.2. Although the latter is a widely used and FDA-approved CAD system, it has been updated by systems such as Second Look Digital and Second Look Premier. Use of these systems would most likely lead to smaller performance differences when compared with AccuDetect Galileo, but unavailability of these systems within our hospital prevented us from performing this comparison. Nonetheless, we feel that the comparison with Second Look version 7.2 is still valid since many institutions worldwide use this CAD system instead of Second Look Digital or Premier.

Second, the study population used is relatively small. However, we feel that, in this exploratory study, these numbers are more than sufficient to assess whether the new voting methodology would result in improved diagnostic performances. Nevertheless, these numbers are too small to directly translate to screening mammography populations. Despite this limitation, the current results warrant future studies that evaluate the benefits of the voting approach in screening mammography.

Third, breast density was assessed by visual assessment of digital mammograms. Recently, Lobbes et al. showed that visual assessment of breast density is less accurate than using a (semi-) automated computerized system, even for experienced breast radiologists [20]. Only recently did commercial software systems become available for breast density measurements, such as Volpara (Matakina International, Wellington, New Zealand) or Quantra (Hologic, Bedford, MA, USA). However, the use of these systems is still only limited to GE and Hologic mammography units.

5. Conclusion

AccuDetect Galileo showed better overall performance than Second Look in detecting masses, microcalcifications per image, and all cancer types per image and per case. The observed improvement was not statistically significant for masses and microcalcifications per case. Most interestingly, accuracy of AccuDetect Galileo was significantly higher in extremely dense breasts.

References


This reprint is provided with the support of PARASCIPT, LLC.