ELSEVIER

Clinical Imaging 37 (2013) 283–288

CLINICAL IMAGING

Malignant lesions on mammography: accuracy of two different computer-aided detection systems

Marc Lobbes [a,*], Marjolein Smidt [b], Kristien Keymeulen [b], Rossano Girometti [c], Chiara Zuiani [c], Regina Beets-Tan [a], Joachim Wildberger [a], Carla Boetes [a]

[a] *Maastricht University Medical Center, GROW School for Oncology and Developmental Biology, Department of Radiology, Maastricht, The Netherlands*
[b] *Maastricht University Medical Center, Department of Surgery, Maastricht, The Netherlands*
[c] *University of Udine, Institute of Diagnostic Radiology, Udine, Italy*

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 *Accu*Detect Galileo). True-positive fractions per image and case for soft densities, microcalcifications, and total cancers were assessed. Study results showed better overall performance of *Accu*Detect Galileo (when compared to Second Look) in detecting masses, microcalcifications, and all cancer types, especially in extremely dense breast parenchyma.

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 *Accu*Detect 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 *Accu*Detect, 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.

* 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).

0899-7071/$ – see front matter
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 lymphatic node, phyllodes tumor, and soft tissue tumor.

Two CAD systems were used: Second Look (version 7.2, iCAD Inc., Nashua, NH, USA), which is part of our standard analysis tool of the mammograms, and *Accu*Detect Galileo (version 4.0.1., Parascript LLC, Longmont, CO, USA). We opted for these two CAD systems since Second Look is a widely used, FDA-approved CAD system, whereas *Accu*Detect Galileo differs from traditional CAD systems in that it uses a so-called voting methodology (see also Discussion section). All cases were analyzed by both systems. The cases were analyzed per image using the mediolateral oblique (MLO) and craniocaudal (CC) projection separately and per case using both the MLO and CC projection. True-positive fractions (TPF) per image and per case were assessed for masses, microcalcifications, and all cancers. TPF per image is the fraction of images where the breast cancer was correctly identified. TPF per case is a fraction of cases where the breast cancer was correctly identified in at least one image projection. The image was correctly identified if at least one hypothesis focal point of the CAD system was located within the truth region of the breast cancer. This truth region was outlined by a blinded and independent expert radiologist (who had >10 years of experience and who was familiar with the case, but who was blinded for the study aim) by drawing contours of the mammographic abnormality on the mammogram based on pathology finding. The operating points for both systems were set at approximately the same false-positive rates per image (FPI) and per case (FPC). However, we did not have access to the internal settings of Second Look. Therefore, we computed the false-positive findings per image on the entire data set for Second Look first. We then defined the operating point of *Accu*Detect Galileo to be approximately equal to that of Second Look.

In a subanalysis of the data, the performance of both systems in extremely dense breasts was also evaluated. Breast density was visually assessed and categorized into one of the four quantitative breast density categories defined in the American College of Radiology (ACR) Breast Imaging Reporting and Data System: ACR 1, almost entirely fatty breasts (<24%); ACR 2, scattered fibroglandular densities (25%–49%); ACR 3, heterogeneously dense (50%–74%); and ACR 4, extremely dense (>75%) [8]. In total, 41 cases of extremely dense breasts (ACR 4) were observed, showing 39 masses and 13 calcifications. For the study group with ACR categories 1–3, there were 70 cases with masses and 18 cases with calcifications. All masses and calcifications were biopsy proven and were therefore part of the cancer cases. Most of the cases had both masses and calcifications. We felt that the number of cases per ACR classification was too small to compare each class separately, so it was decided not to break the number of masses and calcifications for each class.

To evaluate differences in accuracy of the two systems, a one-sided, exact McNemar's test was conducted to assess statistical significance of the results acquired. All P values <.05 were considered to be statistically significant.

3. Results

Final results of the comparison between Second Look and *Accu*Detect performances are presented in Tables 1 to 3. The percentages in the tables include the numbers for both CAD systems. For example, a true-positive percentage of 80% per case would mean that the CAD system detected cancer in 80 out of 100 patients.

When compared to Second Look, *Accu*Detect Galileo achieved higher TPF per image for masses (difference of 10.6%, with FPI set at 0.42, P=.0001) and calcifications (difference of 12.8%, with FPI at 0.2, P=.03) (Table 1). Per case, *Accu*Detect Galileo achieved slightly higher TPF for masses (difference of 2.7%, with FPC at 1.68) and calcifications (difference of 16.1%, with FPC at 0.8) (Table 2). However, these differences were not statistically significant for the detection of masses (P=.27). The detection improvement for calcifications was not significant either (P=.06). *Accu*Detect Galileo achieved significantly higher TPF for all cancers: per image difference of 6.9%, with FPI