



**HAL**  
open science

## Computer-aided detection in full-field digital mammography in a clinical population: performance of radiologist and technologists

Frank J. H. M. Biggelaar, Alphons G. H. Kessels, Jos M. A. Engelshoven,  
Carla Boetes, Karin Flobbe

► **To cite this version:**

Frank J. H. M. Biggelaar, Alphons G. H. Kessels, Jos M. A. Engelshoven, Carla Boetes, Karin Flobbe. Computer-aided detection in full-field digital mammography in a clinical population: performance of radiologist and technologists. *Breast Cancer Research and Treatment*, 2009, 120 (2), pp.499-506. 10.1007/s10549-009-0409-y . hal-00535364

**HAL Id: hal-00535364**

**<https://hal.science/hal-00535364>**

Submitted on 11 Nov 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Computer-aided detection in full-field digital mammography in a clinical population: performance of radiologist and technologists

Frank J. H. M. van den Biggelaar ·  
Alphons G. H. Kessels · Jos M. A. van Engelshoven ·  
Carla Boetes · Karin Flobbe

Received: 17 January 2009 / Accepted: 21 April 2009 / Published online: 6 May 2009  
© Springer Science+Business Media, LLC. 2009

**Abstract** The purpose of the study was to evaluate the impact of a computer-aided detection (CAD) system on the performance of mammogram readers in interpreting digital mammograms in a clinical population. Furthermore, the ability of a CAD system to detect breast cancer in digital mammography was studied in comparison to the performance of radiologists and technologists as mammogram readers. Digital mammograms of 1,048 consecutive patients were evaluated by a radiologist and three technologists. Abnormalities were recorded and an imaging conclusion was given as a BI-RADS score before and after CAD analysis. Pathology results during 12 months follow up were used as a reference standard for breast cancer. Fifty-one malignancies were found in 50 patients. Sensitivity and specificity were computed before and after CAD analysis and provided with 95% CIs. In order to assess the detection rate of malignancies by CAD and the observers, the pathological locations of these 51 breast cancers were matched with the locations of the CAD marks and the mammographic locations that were considered to be suspicious by the observers. For all observers, the sensitivity rates did not change after application of CAD. A mean sensitivity of 92% was found for all technologists and 84% for the radiologist. For two technologists, the specificity decreased (from 84 to 83% and from 77 to 75%). For the

radiologist and one technologist, the application of CAD did not have any impact on the specificity rates (95 and 83%, respectively). CAD detected 78% of all malignancies. Five malignancies were indicated by CAD without being noticed as suspicious by the observers. In conclusion, the results show that systematic application of CAD in a clinical patient population failed to improve the overall sensitivity of mammogram interpretation by the readers and was associated with an increase in false-positive results. However, CAD marked five malignancies that were missed by the different readers.

**Keywords** Digital mammography · Computer-aided detection · Performance · Radiologist · Technologist · Sensitivity · Specificity

## Introduction

Mammography is the most widely used method for early detection and diagnosis of breast cancer and has shown to result in a significant reduction in mortality. However, despite its effectiveness, a number of mammographically detectable breast cancers may still be missed [1].

Therefore, interest is growing in the use of computer-aided detection (CAD) systems, which mark suspicious regions that may otherwise be overlooked by the mammogram reader. Many studies have validated CAD techniques in screening populations and found a sensitivity ranging from 73 [2] to 96% [3]. Moreover, the impact of a CAD system on the performance of mammogram observers was evaluated in several studies [2, 4–14], showing that there is insufficient evidence to claim that CAD improves cancer detection rates but that it does increase recall rates in screening programs for breast cancer [15].

F. J. H. M. van den Biggelaar · J. M. A. van Engelshoven ·  
C. Boetes · K. Flobbe (✉)  
Department of Radiology, Maastricht University Medical  
Center, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands  
e-mail: k.flobbe@mumc.nl

A. G. H. Kessels  
Department of Clinical Epidemiology and Medical Technology  
Assessment, Maastricht University Medical Center, P.O. Box  
5800, 6202 AZ Maastricht, The Netherlands

Little is known on the feasibility of CAD software in mammography in a clinical population. In the Netherlands, hospital radiology departments perform both diagnostic and screening mammography. Diagnostic examinations are carried out in women with clinical signs and symptoms suggestive for breast cancer, women referred through the national breast cancer screening program and in women with a personal history of breast cancer. Furthermore, screening examinations are performed in asymptomatic women referred for a family history of breast cancer, a genetic predisposition or for reassurance. The application of CAD in such a clinical population is as a screening tool, indicating areas that need special interest in screening mammograms and indicating additional abnormalities in the ipsilateral and the contralateral breast in diagnostic mammograms.

In addition to the use of CAD systems to increase the cancer detection rate, another performance improving development in diagnostic breast imaging includes the deployment of radiographic technologists in double reading or pre-reading mammograms. In the setting of an increased workload for radiologists, this alternative working practice, known as skill mixing, has been explored in many studies [16–21]. A systematic review showed that technologists could be as sensitive as radiologists in detecting breast malignancies, but with higher false-positive rates [22]. This raises the question whether the application of CAD software could also enhance the reading performance of technologists.

The aim of the current study was twofold. First, the impact of a CAD system on the performance of mammogram readers in interpreting digital mammograms in a clinical population was evaluated. Second, the ability of a CAD system to detect breast cancer in digital mammography was studied in comparison to the performance of radiologists and technologists as mammogram readers.

## Methods

### Patient selection

Digital mammography examinations of 1,050 consecutive women referred to the radiology department of Maastricht University Medical Centre between January 2007 and July 2007 were selected for this study. Two patients with a proven breast malignancy at the time of mammography examination were excluded. Consequently, mammograms of 1,048 patients with a mean age of 51 years (range 20–90) were included in the study. In 829 patients (79%), diagnostic mammography was performed, whereas 219 women (21%) were referred for screening mammography. Indications for referral for diagnostic breast imaging were:

follow up of prior breast malignancy ( $n = 285$ , 27%), including 164 examinations after lumpectomy and 121 examinations after mastectomy, the occurrence of a palpable breast lump ( $n = 255$ , 24%), other symptomatic complaints like pain or nipple abnormalities ( $n = 189$ , 18%), follow up of a prior benign abnormality ( $n = 62$ , 6%) and referral through the national breast cancer screening program ( $n = 38$ , 4%). Indications for screening mammography were: family history of breast cancer, including BRCA gene mutation ( $n = 174$ , 17%), and other asymptomatic reasons for referral ( $n = 45$ , 4%).

In all patients, a standard two-view unilateral or bilateral mammography examination was performed, using a full-field digital mammography system (Giotto Image FFDM, IMS, Bologna, Italy).

The institutional medical ethics committee approved the study.

### Reference standard

The reference standard for the presence or absence of breast cancer was determined by the pathologic results from core needle biopsies and surgical excisions within a follow up of 12 months. Breast cancer status was presumed to be negative when no pathologic condition was reported after 1 year. Lobular carcinomas in situ were excluded as malignancies.

Pathology data were retrieved from PALGA, a nationwide network and registry of histopathology and cytopathology in the Netherlands, to which all Dutch hospital pathology departments are linked.

In the population described earlier, 51 breast cancers were found in 50 patients, leading to a prevalence of 4.8% (50/1,048). One patient had bilateral breast cancer. Forty-six patients with a malignancy (92%) were referred for diagnostic mammography; four were referred for screening mammography.

The histopathologic classification of breast cancer included 6 ductal carcinomas in situ, 35 invasive ductal malignancies, 8 invasive lobular malignancies and 2 other invasive malignancies.

Breast density categories were scored for all patients with breast cancer by a well-experienced radiologist (CB). Twenty-one mammography examinations were classified as <25% dense, 17 examinations as 25–50%, 12 as 50–75% and 1 as >75%.

### CAD system

All digital images were analyzed using the Second Look Digital CAD system (iCAD Inc.; Beavercreek, OH, USA; version 7.2-H; date of installation: December 4th 2006). The system uses artificial intelligence and pattern

recognition technology to highlight suspicious regions on a mammogram with size-varying marks [10]. Possible masses on the mammogram are marked by ellipses; possible microcalcifications are indicated by squares. The software uses binary thresholds in order to decide whether a mark is displayed.

Instruction of the observers in the use of the Second Look CAD system was provided by the manufacturer by onsite training after installation. Furthermore, 9 months of clinical experience with the software was realized before the onset of this study. During this period, all readers used the software according to their own judgment in each specific patient.

#### Impact of CAD system on reader performance

In order to evaluate the impact of the CAD system on the performance of the observers, the set of 1,048 consecutive mammography examinations was evaluated before and after CAD analysis. Each mammogram was interpreted by four observers, consisting of the radiologist on duty and three technologists.

Two radiologists were involved in the study and, according to their work schedule in daily practice, they both evaluated half of the set of mammograms. They have, respectively, 5 and 20 years of experience in reading over 1,000 mammograms a year in the department. The technologists involved had 1 year experience in mammogram interpretation, as part of a project on radiology skill mixing in which they were trained as mammogram readers [23]. The observers had full information on patients' age and reason for referral, and had access to prior screening and diagnostic mammographic images. All were blinded to the evaluations of the other observers.

All mammographic findings were registered on case record forms. Abnormalities were sketched in a representation of each breast in craniocaudal and mediolateral oblique views. A BI-RADS score was given, which is based on a grading reporting scale for mammography with an increasing degree of suspicion for malignancy [24]. Furthermore, it was indicated whether the observer advised additional diagnostic workup. The impact of CAD on the performance of the observers was evaluated in special reading sessions, in which the mammograms were read again with the help of the CAD software.

All observers indicated whether the CAD marks added valuable diagnostic information to their own original evaluation, sketched in the relevant CAD marks on the representation of the breasts and assigned a BI-RADS score for that abnormality. Furthermore, it was indicated whether additional diagnostic workup was advised by the readers, based on the new findings of the CAD system.

To evaluate the diagnostic performance of the observers, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were computed before and after CAD analysis and provided with 95% CIs. Furthermore, a diagnostic odds ratio (DOR) with 95% CIs was calculated before and after application of the CAD system. The DOR is a measure for the diagnostic performance of a test as it combines sensitivity and specificity in one measure. The DOR can be derived by dividing the odds of a positive test result among diseased persons by the odds of a positive test result among non-diseased persons  $[(\text{sensitivity}/(1 - \text{sensitivity})) \times (\text{specificity}/(1 - \text{specificity}))]$ . A DOR of 1 implies that the test has no discriminative power at all; the larger the DOR, the better the test discriminates between patients with and without the disease of interest [25].

The statistical measures were calculated on patient level with a single overall BI-RADS assessment per patient that was based on the most suspicious score of the two breasts. BI-RADS 1 (normal) and 2 (benign finding) assessments were interpreted as negative for analysis. In addition, mammograms given a BI-RADS 3 (probably benign finding) assessment with a recommendation of short-term or normal follow up were recoded as BI-RADS 3– and were considered as negative for analysis.

Mammograms with a BI-RADS 3 assessment with a recommendation of immediate work-up were recoded as BI-RADS 3+ and were considered as positive for analysis. This sub-classification of BI-RADS 3 assessments was adopted from the Breast Cancer Surveillance Consortium [26], and has been used in several mammography studies [5, 26–28]. Furthermore, BI-RADS assessments 0 (need additional imaging evaluation), 4 (suspicious abnormality) and 5 (highly suggestive of malignancy) were considered as positive for analysis.

Statistical analyses were performed using SPSS 16.0 for Windows (SPSS Inc., Chicago, USA) and Stata 9.0 statistical software package (StataCorp LP, Texas, USA). McNemar's test was used to test differences between sensitivity and specificity before and after CAD analysis. Furthermore, in order to test differences between the DORs before and after CAD application, the bootstrap technique was used, which is a resampling method with replacement from the original sample [29]. *P* value less than 0.05 was considered to be statistically significant.

#### Validity of CAD software

In order to determine the ability of the CAD system to detect breast cancer in mammograms, all images with proven malignancies ( $n = 51$ ) were evaluated by a well-experienced breast radiologist (CB), who was not involved in the primary data collection of this study. First, the breast

quadrant with a proven malignancy was obtained from the pathology reports from core needle biopsies and surgical excisions. Then, the pathological location of the breast cancers were matched visually with the corresponding location on the mammogram images. Consequently, it was defined whether the CAD system had correctly indicated these concerning locations. Similarly, the detection rates of the observers in the study were defined and were compared to the detection rate of the CAD system.

All cases with locations of CAD marks corresponding with the center of the malignancy were considered to be true-positive. Furthermore, when the location of the tumor was confirmed by CAD in at least one mammographic view of the examination, the result was regarded as true-positive.

Consequently, the sensitivity of the CAD software was computed as the number of malignancies correctly marked by the system divided by the total number of malignancies.

Furthermore, in order to determine the rate of false-positive results, CAD was applied to mammography examinations of 100 consecutive patients with normal mammograms (BI-RADS 1,  $n = 64$ ) or mammograms with clearly benign findings (BI-RADS 2,  $n = 36$ ). All CAD marks in these images were counted and considered to be false-positive. Subsequently, the average number of false-positive marks of the CAD system per patient was calculated with 95% CIs.

## Results

### Impact of CAD on reader performance

In Table 1, the performance of mammogram interpretation by the observers is presented before and after CAD analysis, thus with and without knowledge of the findings of the CAD software. For all observers, the sensitivity rates did not change after application of the CAD software. A mean sensitivity of 92% was found for all technologists, as compared to the sensitivity of 84% for the radiologist.

For technologist 1 and 3, the specificity decreased from 84 to 83% and from 77 to 75%, respectively. For the radiologist and technologist 2, the application of CAD did not have any impact on the specificity rates.

Furthermore, the DOR rates decreased after the use of CAD software. For the radiologist, the DOR decreased from 96 to 90, whereas the mean DOR for the three technologists decreased from 52 to 49.

In Table 2, the effect of CAD is shown in all cases that were found to be positive for analysis, with a BI-RADS score of 0, 3+, 4 and 5 given by each observer. Before CAD analysis, the radiologist classified 49 cases as BI-RADS 3+ (5% of all mammograms), whereas the technologists scored

**Table 1** Reader performance (50 patients with breast cancer in 1,048 patients)

		Before CAD analysis	After CAD analysis
Radiologist	Sens	84 (71–93)	84 (71–93)
	Spec	95 (93–96)	95 (93–96)
	PPV	45 (34–55)	44 (34–54)
	NPV	99 (98–100)	99 (98–100)
	DOR	96 (43–210)	90 (42–193)
Technologist 1	Sens	92 (81–98)	92 (81–98)
	Spec	84 (82–86)	83 (81–86)*
	PPV	23 (17–29)	22 (16–28)
	NPV	99 (99–100)	99 (99–100)
	DOR	62 (23–167)	57 (21–155)*
Technologist 2	Sens	92 (81–98)	92 (81–98)
	Spec	83 (81–85)	83 (81–85)
	PPV	22 (16–28)	21 (16–27)
	NPV	99 (99–100)	99 (99–100)
	DOR	57 (21–154)	56 (21–152)
Technologist 3	Sens	92 (81–98)	92 (81–98)
	Spec	77 (74–79)	75 (72–77)**
	PPV	17 (12–22)	15 (11–20)
	NPV	99 (99–100)	99 (99–100)
	DOR	38 (14–103)	34 (13–91)**

Sens, spec, NPV, PPV are shown as percentages. Data in parentheses are 95% CIs

\* Difference between before and after data ( $P = 0.002$ )

\*\* Difference between before and after data ( $P < 0.001$ )

a mean number of 165 patients as BI-RADS 3+ (16% of all mammograms). Furthermore, it was shown that application of CAD did lead to an increase in the number of positive cases, but mainly in the BI-RADS 3+ category ( $n = 30$ ). The highest increase was found for technologist 3, from 178 to 200 BI-RADS 3+ cases. However, no additional malignancies were detected after CAD analysis, so these cases can be regarded as false-positive results.

Further analysis of these 30 cases showed that before CAD, these mammograms had a BI-RADS classification of either benign ( $n = 6$ ) or normal ( $n = 24$ ).

The two other changes in BI-RADS category after the application of CAD included a shift of BI-RADS 2 into 0 for technologist 1 and BI-RADS 2 into 4 for technologist 3.

### Validity of CAD software

Analyzing all CAD marks on the 51 mammograms with proven breast cancer showed that the CAD system correctly marked 40 malignancies. Consequently, the sensitivity of the detection of malignancies by the CAD software in this specific population was 78%. In 11 of the

**Table 2** Effect of CAD in all cases regarded to be positive for analysis

	Radiologist		Technologist 1		Technologist 2		Technologist 3	
	Before	After	Before	After	Before	After	Before	After
BI-RADS 0								
Total ( <i>n</i> )	7	7	2	<b>3</b>	3	3	41	41
Malignancies ( <i>n</i> )	1	1	0	0	0	0	0	0
BI-RADS 3+								
Total ( <i>n</i> )	49	<b>51</b>	155	<b>160</b>	161	<b>162</b>	178	<b>200</b>
Malignancies ( <i>n</i> )	9	9	6	6	4	4	7	7
BI-RADS 4–5								
Total ( <i>n</i> )	38	38	46	46	50	50	57	<b>58</b>
Malignancies ( <i>n</i> )	32	32	40	40	42	42	39	39
Total positive cases								
Total ( <i>n</i> )	94	<b>96</b>	203	<b>209</b>	214	<b>215</b>	276	<b>299</b>
Malignancies ( <i>n</i> )	42	42	46	46	46	46	46	46

The values given in bold are the only data that changed after CAD analysis

51 mammograms (22%), however, CAD analysis missed the malignancy (false-negative result). In three mammograms, no CAD marks were given at all, whereas marks were placed outside the location of the malignancy on eight mammograms.

Furthermore, in the 100 consecutive patients with normal ( $n = 64$ ) and benign findings ( $n = 36$ ), CAD software indicated 225 marks in 79 patients, leading to an average number of 2.9 false-positive marks per patient (95% CI 2.4–3.3). Of these 79 patients, 60% were scored as BI-RADS 1 and 40% as BI-RADS 2. In all 100 patients, the CAD system displayed an average number of 2.3 false-positive marks per patient (95% CI 1.8–2.7) with 0.7 calcification marks (95% CI 0.4–1) and 1.6 mass marks (95% CI 1.3–1.8).

#### Breast cancer detection by CAD versus observers

In Table 3, detailed information of the performance of the observers and the CAD system in detecting the 51 malignancies in this study is presented. The mean detection rate of the observers in diagnostic mammograms was 87% (range 79–89).

In the screening mammograms, the three technologists detected all four malignancies (100%), whereas the radiologist detected three.

Distinguishing the performance by breast density levels, the mean detection rate of the observers was 94% in non-dense breasts (<50% density) compared with 67% in dense breasts (>50% density). The CAD system detected 84% of the malignancies in the non-dense breasts, compared with 62% in dense breasts.

When specifying the histopathologic type of breast cancer, it is shown that all patients with DCIS were correctly identified by all observers and the CAD system.

Furthermore, the observers and CAD detected about 85% of all patients with invasive ductal carcinoma. Women with invasive lobular carcinomas were found by the radiologist and technologists in 6 to 8 out of 12 cases, whereas CAD identified only 4.

Analyzing the CAD marks given by the system in the 51 mammograms with a histology-proven breast cancer, showed that five malignancies were indicated by a CAD mark without being noticed as suspicious by the observers. The radiologist failed to identify four of these malignancies, one technologist failed in two cases and two technologists failed to detect three.

Four of these malignancies were in diagnostic mammograms; one malignancy was detected in screening mammograms. Three patients had non-dense breasts (<50% density) and two patients had dense breasts (>50% density). All malignancies were invasive ductal carcinomas.

#### Discussion

This study demonstrated that the application of CAD was associated with an increased number of false-positive cases in all observers and failed to improve the sensitivity of mammogram interpretation. Furthermore, the CAD system correctly indicated 78% of the 51 cancers which is consistent with results in other studies with consecutive patient series in screening mammography, reporting sensitivities ranging from 73 to 96% [2, 3, 13, 30].

The use of a consecutive patient series is a strength of this study, as it represents the patient population referred for diagnostic and screening mammography in daily clinical practice. However, the use of special reading sessions in order to evaluate the mammograms with and without the

**Table 3** Detection of malignancies ( $n = 51$ )

	<i>n</i> malignancies	<i>n</i> detected by					
		Radiologist	Technologist 1	Technologist 2	Technologist 3	CAD	CAD alone <sup>a</sup>
Total	51 (100%)	42 (82%)	45 (88%)	46 (90%)	45 (88%)	40 (78%)	5 (10%)
Setting							
Screening	4 (8%)	3 (75%)	4 (100%)	4 (100%)	4 (100%)	3 (75%)	1
Diagnostic	47 (92%)	39 (83%)	41 (87%)	42 (89%)	41 (87%)	37 (79%)	4
Density (%)							
<25	21 (41%)	19 (90%)	20 (95%)	20 (95%)	20 (95%)	19 (90%)	2
25–50	17 (33%)	16 (94%)	16 (94%)	16 (94%)	16 (94%)	13 (76%)	1
51–75	12 (24%)	7 (58%)	9 (75%)	10 (83%)	9 (75%)	8 (67%)	2
>75	1 (2%)	0	0	0	0	0	–
Type							
DCIS	6 (12%)	6 (100%)	6 (100%)	6 (100%)	6 (100%)	6 (100%)	–
Invasive ductal malignancies	35 (69%)	29 (83%)	30 (86%)	30 (86%)	29 (83%)	30 (86%)	5
Invasive lobular malignancies	8 (16%)	6 (75%)	7 (88%)	8 (100%)	8 (100%)	4 (50%)	–
Other invasive malignancies	2 (4%)	1 (50%)	2 (100%)	2 (100%)	2 (100%)	0	–
CAD analysis							
Detected	40 (78%)	36 (90%)	38 (95%)	37 (93%)	37 (93%)	–	–
Missed	11 (12%)	6 (55%)	7 (64%)	9 (82%)	8 (73%)	–	–

<sup>a</sup> Malignancies detected by CAD, but missed by at least one of the observers

application of CAD software can be regarded as a potential limitation, as this does not represent daily practice. The lack of direct clinical consequences of the BIRADS assessments given, combined with the large volume of mammograms in these reading sessions, may possibly have led to a loss of focus and a potential decrease in performance. However, in order to provide equal evaluation circumstances for all observers and to monitor the data collection closely, this study design was thought to be the most appropriate.

Another limitation of this study is the lack of information on a potential learning curve in using the CAD software. Although the CAD software was available to all observers for 9 months before the onset of the study, detailed information about the actual experience of the observers with the CAD software was missing.

In this study, the sensitivity of mammogram evaluation without including information on additional imaging tests was 84% for the radiologist and 92% for the three technologists (McNemar,  $P$  value 0.125). However, the higher sensitivity for the technologists comes at the cost of a lower specificity. The technologists' mean specificity was significantly lower than for the radiologist (81 and 95% respectively, McNemar,  $P$  value < 0.001). Apparently, technologists were more cautious and uncertain in their evaluation which did lead to a higher rate of false-positive results. These findings are in accordance with the results of a systematic literature review indicating that technologists

could be as sensitive as radiologists in detecting malignancies on mammograms, but with lower specificity [22].

It needs to be mentioned that the sensitivity of the radiologist for mammogram evaluation in our study is in concordance with the performance found in other studies on similar patient populations [31, 32]. In addition, the DOR which combines sensitivity and specificity into one measure was higher for the radiologist (96) as compared to the technologists (38–62). This means that the overall performance in mammogram evaluation was better for the radiologist as compared to the technologists.

In most published studies, CAD was applied to conventional mammographic images that were digitized. A few studies evaluated CAD in digital mammograms. In a study of Kim et al. [12], initial and follow up mammograms of 93 patients with a malignancy were read by a CAD system. An overall sensitivity of 90% was found, whereas the reproducibility of true-positive CAD marks was high. Yang et al. [3] reported a CAD sensitivity of 96% in 103 screening breast cancers. In a study of Wei et al. [33], CAD was evaluated in full-field digital mammograms and digitized screen-film mammograms. In a set with 131 masses, no significant difference was found between the performance of CAD in the digital images and the screen-film images.

It can be assumed that the availability of digital mammography, like in this study, will facilitate the applicability of CAD software in daily clinical practice. Evaluating

mammograms with CAD software in digital images avoids loss of image quality and reproducibility of the CAD system because of a mechanical scanning digitization process. Furthermore, in digital mammography CAD software can be applied immediately after image acquisition, whereas in conventional analog mammography, the digitization of mammograms will provide a time delay.

Most studies report the application of CAD software in screening mammograms. Only a few studies evaluated the feasibility of CAD software in a population including both symptomatic patients as well as women referred for screening mammography. Brancato et al. [34] included 3,425 women attending a self-referral breast centre for digital mammography. Of all women, 25% presented with symptoms and 75% presented for screening. In this population, 102 malignancies (88 in symptomatic women and 14 in women having screening) were detected by the radiologist without the use of CAD. After application of the CAD software, three additional malignancies were found in patients with symptoms and two in women presenting for screening.

Furthermore, in a prospective study, Dean and Ilvento [30] evaluated the impact of CAD on cancer detection in 9,520 film-screen mammograms. Of all mammograms, 60% were screening studies and 40% were diagnostic. 104 malignancies were detected and, after CAD application, a 10.8% increase in cancer detection was found.

In our study, only 21% of all women attended for screening mammography. The other 79% were referred for diagnostic mammography. In this population, no additional malignancies were found by the observers after CAD analysis. However, a close analysis of the CAD marks that were given in mammograms with histology-proven malignancies showed that CAD indicated five additional malignancies that were missed by one or more observers. This indicates that the readers in this study did not always respond to a correct mark of the CAD system. The large number of false-positive marks resulting in a low specificity of the CAD system, may be an explanation. On average, over two false-positive marks were displayed per patient, so more than 99% of the marks should be ignored to find one true-positive finding that was missed initially by the observer. The interpretation of a CAD mark is based on mammographic features of the finding and the experience of the observer. Furthermore, patient characteristics, reason for referral and clinical findings influence the observers' decision to act on CAD marks. Therefore, it can be argued that it is difficult to improve sensitivity while maintaining the high specificity of the observer.

The application of CAD decreased the performance of two technologists in reading mammograms in a clinical population because of higher false-positive rates. It can be assumed that the technologists were indecisive of the CAD

marks because of the large number of marks. Furthermore, it is likely that the degree of experience will play a role. Studies evaluating the association between radiologists' experience and accuracy in reading screening mammograms, demonstrate a higher specificity in radiologists who routinely evaluate large numbers of mammograms [35, 36]. Therefore, it may be assumed that the use of CAD software will influence inexperienced readers, like the technologists in this study, more than well-experienced readers, like the radiologist. The results of this study support this assumption, as two technologists produce much more additional false-positive results after CAD analysis, compared to the radiologist.

Concluding, this study has disclosed valuable information about the application of CAD software in mammography in a daily clinical population. Although the results show that systematic application of CAD in a clinical patient population failed to improve the overall sensitivity of mammogram interpretation by the readers and was associated with an increase in false-positive results, CAD marked five malignancies that were missed by the different readers. However, the usefulness of CAD software in a clinical population remains limited.

**Acknowledgments** The authors are grateful to the radiologists Dr. E. van der Linden and Dr. E. Mussen, and the breast technologists Mrs. J. Amory-Maes, Mrs. M. Beerts and Mrs. J. Meijers for their dedicated participation in this study. Pathology data were kindly provided by the nationwide network and registry of histo- and cytopathology in The Netherlands (PALGA). The study was financially supported by Maastricht University Medical Center.

## References

- Collins MJ, Hoffmeister J, Worrell SW (2006) Computer-aided detection and diagnosis of breast cancer. *Semin Ultrasound CT MR* 27:351–355. doi:10.1053/j.sult.2006.05.009
- Taylor PM, Champness J, Given-Wilson RM et al (2004) An evaluation of the impact of computer-based prompts on screen readers' interpretation of mammograms. *Br J Radiol* 77:21–27. doi:10.1259/bjr/34203805
- Yang SK, Moon WK, Cho N et al (2007) Screening mammography-detected cancers: sensitivity of a computer-aided detection system applied to full-field digital mammograms. *Radiology* 244:104–111. doi:10.1148/radiol.2441060756
- Gilbert FJ, Astley SM, Gillan MG et al (2008) Single reading with computer-aided detection for screening mammography. *N Engl J Med* 359:1675–1684. doi:10.1056/NEJMoa0803545
- Fenton JJ, Taplin SH, Carney PA et al (2007) Influence of computer-aided detection on performance of screening mammography. *N Engl J Med* 356:1399–1409. doi:10.1056/NEJMo066099
- Freer TW, Ulissey MJ (2001) Screening mammography with computer-aided detection: prospective study of 12, 860 patients in a community breast center. *Radiology* 220:781–786. doi:10.1148/radiol.2203001282
- Georgian-Smith D, Moore RH, Halpern E et al (2007) Blinded comparison of computer-aided detection with human second



- reading in screening mammography. *AJR Am J Roentgenol* 189:1135–1141. doi:[10.2214/AJR.07.2393](https://doi.org/10.2214/AJR.07.2393)
8. Gromet M (2008) Comparison of computer-aided detection to double reading of screening mammograms: review of 231, 221 mammograms. *AJR Am J Roentgenol* 190:854–859. doi:[10.2214/AJR.07.2812](https://doi.org/10.2214/AJR.07.2812)
  9. Hadjiiski L, Sahiner B, Helvie MA et al (2006) Breast masses: computer-aided diagnosis with serial mammograms. *Radiology* 240:343–356. doi:[10.1148/radiol.2401042099](https://doi.org/10.1148/radiol.2401042099)
  10. Karssemeijer N, Otten JD, Verbeek AL et al (2003) Computer-aided detection versus independent double reading of masses on mammograms. *Radiology* 227:192–200. doi:[10.1148/radiol.2271011962](https://doi.org/10.1148/radiol.2271011962)
  11. Khoo LA, Taylor P, Given-Wilson RM (2005) Computer-aided detection in the United Kingdom National Breast Screening Programme: prospective study. *Radiology* 237:444–449. doi:[10.1148/radiol.2372041362](https://doi.org/10.1148/radiol.2372041362)
  12. Kim SJ, Moon WK, Cho N et al (2008) Computer-aided detection in full-field digital mammography: sensitivity and reproducibility in serial examinations. *Radiology* 246:71–80. doi:[10.1148/radiol.2461062072](https://doi.org/10.1148/radiol.2461062072)
  13. Ko JM, Nicholas MJ, Mendel JB et al (2006) Prospective assessment of computer-aided detection in interpretation of screening mammography. *AJR Am J Roentgenol* 187:1483–1491. doi:[10.2214/AJR.05.1582](https://doi.org/10.2214/AJR.05.1582)
  14. Malich A, Schmidt S, Fischer DR et al. (2009) The performance of computer-aided detection when analyzing prior mammograms of newly detected breast cancers with special focus on the time interval from initial imaging to detection. *Eur J Radiol* 69:574–578. doi:[10.1016/j.ejrad.2007.11.038](https://doi.org/10.1016/j.ejrad.2007.11.038)
  15. Taylor P, Potts HW (2008) Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate. *Eur J Cancer* 44:798–807. doi:[10.1016/j.ejca.2008.02.016](https://doi.org/10.1016/j.ejca.2008.02.016)
  16. Haiart DC, Henderson J (1991) A comparison of interpretation of screening mammograms by a radiographer, a doctor and a radiologist: results and implications. *Br J Clin Pract* 45:43–45
  17. Hillman BJ, Fajardo LL, Hunter TB et al (1987) Mammogram interpretation by physician assistants. *AJR Am J Roentgenol* 149:907–912
  18. Pauli R, Hammond S, Cooke J et al (1996) Radiographers as film readers in screening mammography: an assessment of competence under test and screening conditions. *Br J Radiol* 69:10–14
  19. Pauli R, Hammond S, Cooke J et al (1996) Comparison of radiographer/radiologist double film reading with single reading in breast cancer screening. *J Med Screen* 3:18–22
  20. Sumkin JH, Klaman HM, Graham M et al (2003) Prescreening mammography by technologists: a preliminary assessment. *AJR Am J Roentgenol* 180:253–256
  21. Wivell G, Denton ER, Eve CB et al (2003) Can radiographers read screening mammograms? *Clin Radiol* 58:63–67. doi:[10.1053/crad.2002.1087](https://doi.org/10.1053/crad.2002.1087)
  22. van den Biggelaar FJ, Nelemans PJ, Flobbe K (2008) Performance of radiographers in mammogram interpretation: a systematic review. *Breast* 17:85–90. doi:[10.1016/j.breast.2007.07.035](https://doi.org/10.1016/j.breast.2007.07.035)
  23. van den Biggelaar FJHM, Kessels AG, van Engelshoven JM et al (2008) Diagnostic performance of breast technologists in reading mammograms in a clinical patient population. (submitted for publication)
  24. D’Orsi CJ, Bassett LW, Berg WA et al (2003) Breast Imaging Reporting and Data System: ACR BI-RADS-Mammography. American College of Radiology (ACR), Reston
  25. Glas AS, Lijmer JG, Prins MH et al (2003) The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 56:1129–1135. doi:[10.1016/S0895-4356\(03\)00177-X](https://doi.org/10.1016/S0895-4356(03)00177-X)
  26. Barlow WE, Lehman CD, Zheng Y et al (2002) Performance of diagnostic mammography for women with signs or symptoms of breast cancer. *J Natl Cancer Inst* 94:1151–1159
  27. Poplack SP, Tosteson AN, Grove MR et al (2000) Mammography in 53, 803 women from the New Hampshire mammography network. *Radiology* 217:832–840
  28. Smith-Bindman R, Chu PW, Miglioretti DL et al (2003) Comparison of screening mammography in the United States and the United Kingdom. *JAMA* 290:2129–2137. doi:[10.1001/jama.290.16.2129](https://doi.org/10.1001/jama.290.16.2129)
  29. Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman & Hall, New York
  30. Dean JC, Ilvento CC (2006) Improved cancer detection using computer-aided detection with diagnostic and screening mammography: prospective study of 104 cancers. *AJR Am J Roentgenol* 187:20–28. doi:[10.2214/AJR.05.0111](https://doi.org/10.2214/AJR.05.0111)
  31. Flobbe K, Bosch AM, Kessels AG et al (2003) The additional diagnostic value of ultrasonography in the diagnosis of breast cancer. *Arch Intern Med* 163:1194–1199. doi:[10.1001/archinte.163.10.1194](https://doi.org/10.1001/archinte.163.10.1194)
  32. Zonderland HM, Coerkamp EG, Hermans J et al (1999) Diagnosis of breast cancer: contribution of US as an adjunct to mammography. *Radiology* 213:413–422
  33. Wei J, Hadjiiski LM, Sahiner B et al (2007) Computer-aided detection systems for breast masses: comparison of performances on full-field digital mammograms and digitized screen-film mammograms. *Acad Radiol* 14:659–669. doi:[10.1016/j.acra.2007.02.017](https://doi.org/10.1016/j.acra.2007.02.017)
  34. Brancato B, Houssami N, Francesca D et al (2008) Does computer-aided detection (CAD) contribute to the performance of digital mammography in a self-referred population? *Breast Cancer Res Treat* 111:373–376. doi:[10.1007/s10549-007-9786-2](https://doi.org/10.1007/s10549-007-9786-2)
  35. Barlow WE, Chi C, Carney PA et al (2004) Accuracy of screening mammography interpretation by characteristics of radiologists. *J Natl Cancer Inst* 96:1840–1850
  36. Moss SM, Blanks RG, Bennett RL (2005) Is radiologists’ volume of mammography reading related to accuracy? A critical review of the literature. *Clin Radiol* 60:623–626. doi:[10.1016/j.crad.2005.01.011](https://doi.org/10.1016/j.crad.2005.01.011)