

Maurice Samulski
Rianne Hupse
Carla Boetes
Roel D. M. Mus
Gerard J. den Heeten
Nico Karssemeijer

Using computer-aided detection in mammography as a decision support

Received: 5 February 2010

Revised: 23 March 2010

Accepted: 20 April 2010

© The Author(s) 2010

This article is published with open access at Springerlink.com

Supported by grant from the Dutch Cancer Society, KUN 2006-3655.

M. Samulski (✉) · R. Hupse ·
R. D. M. Mus · N. Karssemeijer
Department of Radiology,
Radboud University Nijmegen Medical Centre,
Geert Grootplein 10, 6500 HB Nijmegen,
The Netherlands
e-mail: m.samulski@rad.umcn.nl
Tel.: +31-24-3614766
Fax: +31-24-3540866

C. Boetes
Department of Radiology,
Maastricht University Medical Centre,
Maastricht, The Netherlands

G. J. den Heeten · N. Karssemeijer
National Expert and Training Centre for
Breast Cancer Screening, Nijmegen,
The Netherlands

Abstract *Objective:* To evaluate an interactive computer-aided detection (CAD) system for reading mammograms to improve decision making. *Methods:* A dedicated mammographic workstation has been developed in which readers can probe image locations for the presence of CAD information. If present, CAD findings are displayed with the computed malignancy rating. A reader study was conducted in which four screening radiologists and five non-radiologists participated to study the effect of this system on detection performance. The participants read 120 cases of which 40 cases had a malignant mass that was missed at the original screening. The readers read each mammogram both with and without CAD in separate sessions. Each reader reported localized findings and assigned a malignancy score per finding. Mean sensitivity

was computed in an interval of false-positive fractions less than 10%.

Results: Mean sensitivity was 25.1% in the sessions without CAD and 34.8% in the CAD-assisted sessions. The increase in detection performance was significant ($p=0.012$). Average reading time was 84.7 ± 61.5 s/case in the unaided sessions and was not significantly higher when interactive CAD was used (85.9 ± 57.8 s/case).

Conclusion: Interactive use of CAD in mammography may be more effective than traditional CAD for improving mass detection without affecting reading time.

Keywords Mammography · Breast · Early detection of cancer · Decision making · Computer-assisted · Radiographic image interpretation

Introduction

Computer-aided detection (CAD) was introduced in breast cancer screening as a technology to avoid perceptual oversights and its effectiveness has been demonstrated in many studies [1–3]. Nevertheless, there is a continuing debate regarding the usefulness of CAD [4, 5]. While most radiologists agree that CAD systems have value because of their high performance in detecting microcalcifications, many believe that current CAD algorithms for masses and architectural distortions have too many false positives to allow effective use [6–8]. Evidently, more research is needed to improve CAD algorithms. However, the lack of confidence some radiologists have

in CAD may also be another reason. In previous research strong evidence was found that the performance of CAD algorithms may not be a problem, but that the concept of CAD may need to be revised [9]. The assumption on which CAD is currently based is that significant lesions initially missed by radiologists will be acted upon when CAD marks them. In practice, however, many lesions are not missed by perceptual oversight but due to incorrect interpretation [10–12]. Therefore, it is not surprising that studies reveal that many significant lesions are still missed even when CAD marks them [13–16]. To prevent such interpretation errors CAD needs to be designed to help radiologists with decision making.

The purpose of this study was to investigate a novel way of using CAD algorithms. In the traditional prompting approach [17, 18], CAD results are displayed after the reading is completed, offering the reader a possibility to check if no perceptual failures occurred related to search. In current practice, readers are strongly discouraged to downgrade their findings on the basis of CAD. Compared with the traditional approach, we investigated a method in which CAD marks are only displayed on request during the reading. This novel approach means that when the reader is inspecting a certain region in a mammogram, that particular region can be probed for the presence of any CAD information using a pointer and, if present, only the CAD information about this location is shown. In addition to the CAD mark also the level of suspicion computed by the CAD system is displayed. However image regions deemed normal by the reader are not probed for CAD and thus no other CAD marks elsewhere on the image would be shown. Obviously, this approach will not aid in avoiding perceptual oversights. However, this method has the potential to aid readers in making decisions when they inspect potential lesions, without being distracted by false positives of CAD.

Our study was motivated by previous research, which demonstrated a significant improvement in detection performance when CAD mass marks were independently combined with reader scores [10]. In that study, CAD marks on regions not reported by the reader were not used, which is similar to the approach investigated here. As independent combination of reader results with CAD would not be easily accepted in clinical practice, we designed a screening workstation in which readers themselves can combine their interpretation with CAD in an interactive way. To investigate the proposed CAD concept, we conducted a reader study in which nine readers participated.

Materials and methods

The institutional review board approved this retrospective study and waived the need for informed consent. For the purpose of this study, a dedicated mammographic workstation was developed that has the basic functionality that screening radiologists expect when they read digital mammograms on electronic displays, including dedicated hanging protocols, zooming, image manipulation, and local contrast enhancement tools. Brightness and contrast were easily adjustable and were set in advance for optimal efficiency. The workstation was equipped with a 30-inch color LCD panel (model FlexScan SX3031W; Eizo Nanao Technologies Inc., Hakui, Ishikawa, Japan) with a native resolution of $2,560 \times 1,600$. CAD processing was performed on a separate server and results were submitted to the workstation with the image data before a reading session started. CAD results were obtained from the R2 ImageChecker v8.0 (Hologic, Bedford, MA, USA).

On the workstation (Fig. 1) the presence of CAD marks can be queried interactively by clicking on suspect regions in the mammogram using a pointing device by the readers.

It is not possible to display all available CAD marks at once as in traditional CAD prompting devices. For each queried location, the workstation checks if a CAD mark is available at that location. If a CAD mark is available, it is presented to the reader by displaying the contour of the region detected by CAD along with a computer-estimated malignancy score. The contour of the region is colored based on the malignancy score using a continuous color scale ranging from red to yellow, for respectively high to low malignancy ratings. Previous studies show that giving readers additional information on the likelihood of CAD marks might be helpful in decision making [19–22].

The average number of CAD regions that could be activated was adjustable. Only CAD regions with malignancy ratings exceeding some threshold were included. In the observer study, we adjusted this threshold such that in normal cases the average number of false-positive regions was two per image.

Image database

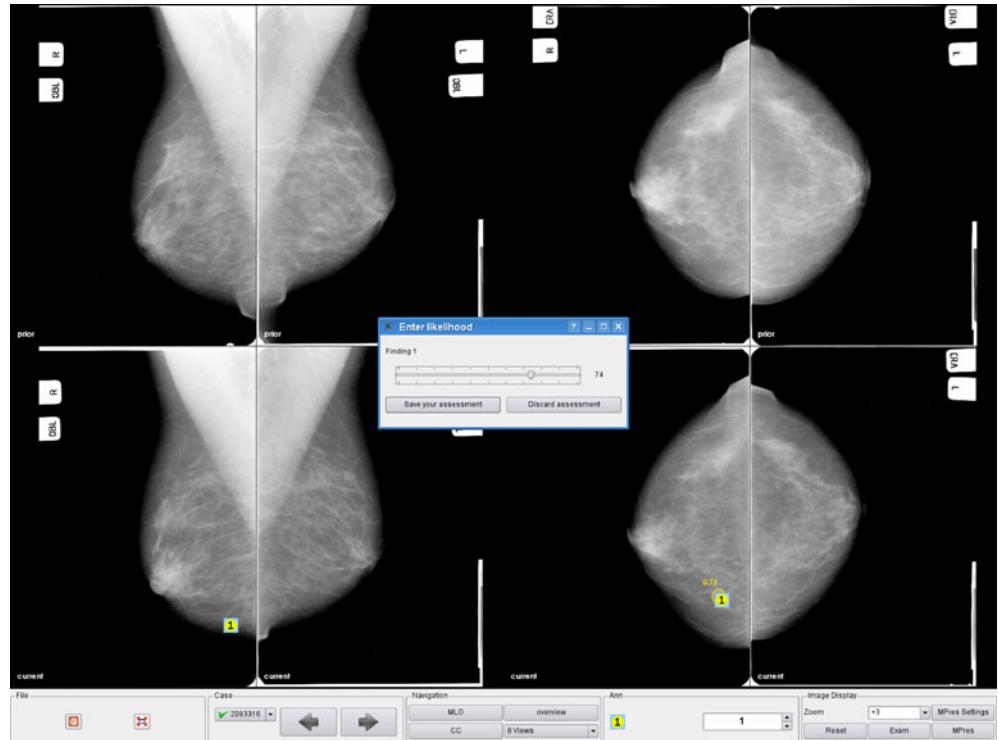
A total of 120 screening mammograms were selected from the Dutch breast cancer screening program and were digitized by using a laser digitizer suitable for medical applications (Lumiscan 85, Lumisys, Sunnyvale, CA, USA) at a pixel resolution of $50 \mu\text{m}$. The mammograms were averaged down to a resolution of $100 \mu\text{m}$, maintaining a gray-level resolution of 12 bits. From these cases, 40 had a biopsy-proven malignant mass, and 80 were cancer-free. As a result of the Dutch screening protocol, the majority of the cases had only mediolateral oblique (MLO) views available. Of the 120 cases only 25 had additional CC views. All cancer cases selected were subtle cancers that were missed at the original screening and were retrospectively identified as visible. We chose to use cases with missed cancers to maximize the power of our observer experiment. Cases with only microcalcifications were excluded. Each mammogram was presented with the corresponding prior screening mammogram, as is common in screening practice to allow detection of temporal changes. The study is summarized in Table 1.

Observer study design

Nine readers, of which four were certified screening radiologists and five were non-radiologists with mammogram reading skills, participated in the study. Before the actual observer study, 60 training cases were presented to the non-radiologists. The expert radiologists were presented with fewer training cases due to time constraints. The number of training cases presented to the radiologists ranged from 10 to 30. The training cases served to familiarize the observers with the system, including the reporting functionalities, the interactive CAD functionality, and the controls for adjusting the brightness and contrast.

The observers read the case set in two batches of 60 cases each. Each batch consisted of two sessions. In the

Fig. 1 The graphical user interface of the CAD workstation used in the observer experiments. The *upper row* shows prior mammograms and the *lower row* displays the current screening mammograms that have to be reported. In the case shown here, a reader reported a localized finding in both projections and is asked to assign a malignancy score between 0 and 100 to that finding. In the craniocaudal (CC) view, a CAD region was present at the reported location



first session, 30 mammograms were read with CAD and 30 without. In the second session, CAD was made available for the cases initially read without CAD and vice versa. Each session had a balanced mix of normal and abnormal cases. The order of the cases within each subset was randomized in the two sessions to minimize reading order effects.

The observers were instructed to search for malignant masses and architectural distortions only, and were informed that the study set did not contain microcalcification cases. They were also informed what the approximate proportion of the abnormal cases was. To report abnormalities, readers were asked to mark the finding in the MLO and CC view, and assign a malignancy score on a continuous scale ranging from 0 to 100. Readers were

also instructed to mark at least one finding per case, unless a case was so obviously normal that no reasonable finding could be marked. In the with-CAD session, the readers could query the CAD system by clicking on regions in the mammogram that they were inspecting. Otherwise the reading and reporting was the same as in the non-CAD sessions. They were free to report any finding, regardless if it was marked by CAD or not. There was no limit on the reading time.

Independent combination of readers and CAD

In a previous study the potential contribution of CAD in improvement of mammographic interpretation was investigated by independently combining findings of the readers with detection results of the CAD software [10]. We applied the same method to the experimental data obtained in this study. In that way we could compare the effect of interactive use of CAD during reading with the effect of combining reader reports with CAD independently after the reading is completed. In summary, independent combination was implemented as follows: only locations in the mammogram that the observers reported were considered. For every finding it was checked whether the location of the finding was marked by CAD and its level of malignancy was determined. If two views were available and the finding was marked in both views, the highest level of malignancy assigned to either of the CAD regions was taken. If the finding was

Table 1 Study overview

Total cases	120
Normal cases	80
Cancer cases	40
Cancer cases detected by CAD ^a	33
Available CAD regions ^b	587
Available true-positive CAD regions	41
Available false-positive CAD regions	546

^a Cancers hit in at least one view by the CAD system at an operating level of 2.0 false-positive markings per image

^b Regions that could be queried at the operating level of 2.0 false-positives markings per image

not marked at all by CAD a zero level was assigned. The combined malignancy score of a finding was computed by taking a weighted average of the reader score with the CAD-estimated malignancy score.

Statistical analysis

We used localization receiver operating characteristic (LROC) to analyze the data for differences in reader performance between reading with and without using interactive CAD, for individual readers, as well as for the average reader. To determine a LROC, the decision threshold is varied and the correct localization fraction is plotted as a function of the false-positive fraction. The false-positive fraction is defined as the fraction of normal cases recalled as a function of the decision threshold.

For every reader, we determined the cutoff point at which the false-positive recall rate was 10%, by thresholding the scores the observer had given to the findings. The primary metric of detection performance was the mean correct localization fraction in the false-positive fraction interval ranging from 0 to 0.1. This interval is chosen because in screening programs radiologists usually have recall rates below 10%.

The location of each finding was indicated in the MLO view and CC view. A finding was considered a true positive (TP), if it had a correct location in at least one of the views. We defined a location to be correct if the distance between the observers' marked location and the true cancer location was less than 2 cm. The false-positive fraction was estimated from the observers' marked locations in the normal cases. We computed significance of differences between sessions with and without CAD for the average reader by using the Wilcoxon signed rank test. Differences with a P value of less than 0.05 were considered significant. The statistical analysis was performed by using R data analysis software (version 2.9.0; R Foundation for Statistical Computing, Vienna, Austria). The number of times reported and unreported TP and false-positive (FP) CAD regions were queried was computed for every reader. A CAD region was considered queried if the distance between the observers' query location and the center point of the CAD region was less than 0.5 cm, or if the query location was within the CAD region.

Reading times

Reading times per case were automatically recorded in the reading sessions. Mean reading time per case and its standard deviation was computed for every reader in both reading modes. Reading times exceeding 5 min were excluded from the analyses on the basis of the assumption that these excessively long reading times were the result of interruptions during the session. As a result, approximately 3% of all cases were excluded from the time analysis. Average reading times for the unaided session

and the session with CAD were calculated. Paired reading times were compared by Wilcoxon signed rank testing. A P value of less than 0.05 was considered to indicate a statistically significant difference.

Results

The results of the nine individual readers are shown in Table 2. It also shows results obtained by independently combining reader scores with CAD. The mean correct localization fraction of a reader in the false-positive fraction interval ranging from 0 to 0.1 (TPF10) is used as the performance measure. Results show that radiologists did not perform better in this study than the non-radiologists. We computed average LROC curves from all the readers, the non-radiologists, and the radiologists. These are shown in Figs. 2, 3, and 4, respectively.

The performance of the average reader increased with CAD at low false-positive rates from 25.1% to 34.8%. Every reader improved their performance using CAD with the exception of reader 8. The difference between reading with and without CAD for the average reader, measured by the performance metric defined above, was statistically significant ($p=0.012$). Results confirm that performance may also be increased by independent combination with CAD scores, with a smaller increase, however, than obtained with interactive use of CAD. The difference we found between interactive use of CAD and independent combination is not statistically significant.

As an example, a mammogram of a woman with an invasive ductal carcinoma is shown in Fig. 5. In this case, seven of the nine readers correctly localized the cancer in both sessions, but rated their finding substantially more suspicious in the session with interactive CAD enabled, one reader only located the cancer correctly in the session where CAD was enabled, and one reader did assign a slightly lower rating to the cancer in the session with

Table 2 Reader detection performance in the false-positive fraction interval ranging from 0 to 0.1

	Without CAD TPF10 (%)	With CAD TPF10 (%)	Independent combination TPF10 (%)
Non-radiologists			
1	41.1	51.3	43.3
2	35.3	51.5	41.7
3	16.0	25.9	26.3
4	15.4	25.2	27.4
5	18.3	41.9	26.7
Average	25.2	39.2	33.0
Radiologists			
6	24.3	32.3	33.6
7	24.8	28.8	30.2
8	30.2	25.7	37.0
9	20.2	30.4	30.0
Average	24.9	29.3	32.7
Reader average	25.1	34.8	32.9

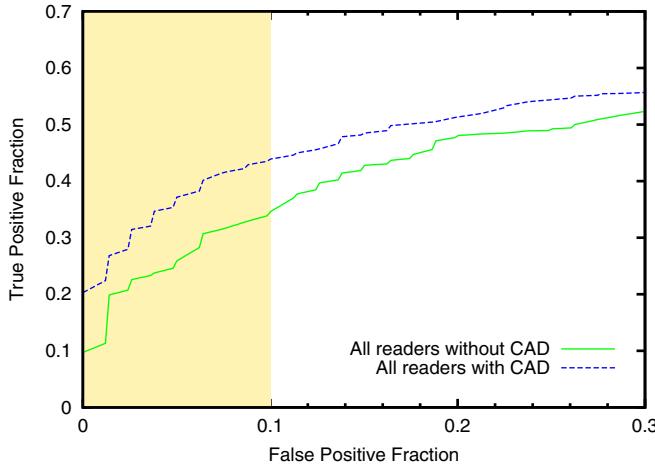


Fig. 2 Average LROC curves obtained from the nine readers for the detection of cancers with and without using CAD. The false-positive fraction interval ranging from 0 to 0.1, where the mean correct localization fraction is computed, is highlighted in *light yellow*

CAD. In Fig. 6, the same case is shown with the activated CAD region. The average time to read a case without CAD was 84.7 ± 61.5 s. The radiologists read the cases much faster than the non-radiologists. Average reading time in the session with CAD was 85.9 ± 57.8 s/case (Table 3). There were no significant differences in reading times for the session with CAD and the session without CAD ($p=0.13$) (Table 3).

The CAD system had a lesion-based sensitivity of 80.4% (41/51) at the operating level of 2.0 false-positive markings per image used in the study. The number of available CAD regions was 587. Table 4 shows that on average 274.2 of the 546 false-positive CAD regions (50.2%) were not queried. It also shows that on average 5 of the 41 true-positive CAD regions (12.2%) were not queried. The radiologists queried far fewer false-positive CAD regions than the non-radiologists.

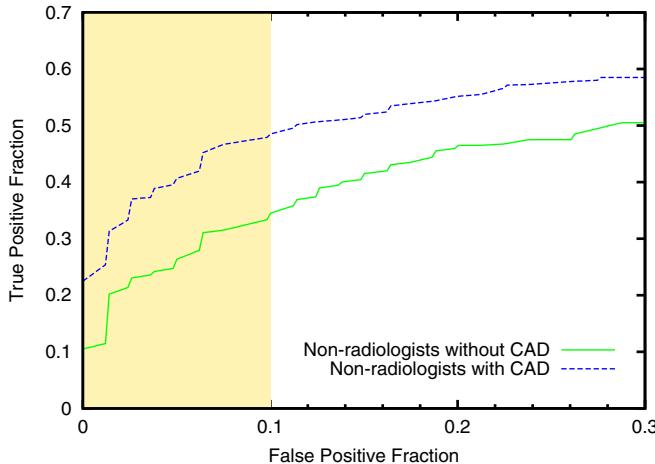


Fig. 3 Average LROC curves obtained from the five non-radiologists

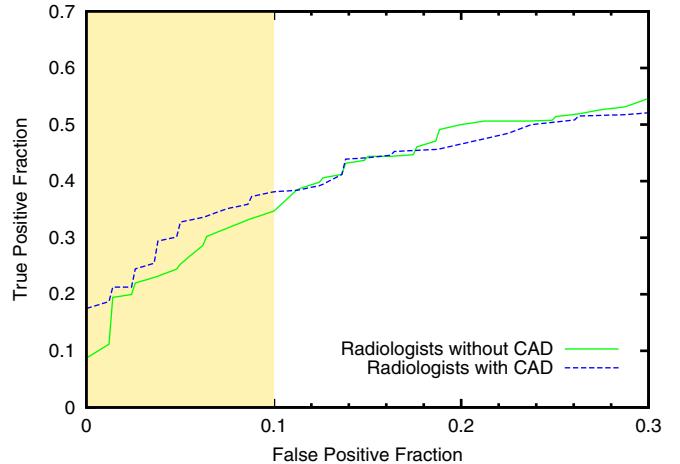


Fig. 4 Average LROC curves obtained from the four radiologists

Discussion

Results of this study show that readers are able to improve detection performance when they use CAD for interpretation of mass lesions in an interactive way. The beneficial effect of CAD can be attributed fully to improvement of interpretation, because traditional CAD prompts to avoid perceptual oversights were not shown. The effectiveness was remarkable given that the readers in this study used the interactive system for the first time and had limited training. It is noted that in a previous experiment using a similar observer study design and dataset no significant



Fig. 5 Mediolateral oblique mammographic views of a woman with an invasive ductal carcinoma indicated by the *arrow*. Seven of the nine readers correctly localized the cancer in both sessions, but rated their finding substantially more suspicious in the session with interactive CAD enabled, one reader only located the cancer correctly in the session where CAD was enabled, and one reader did assign a slightly lower rating to the cancer in the session with CAD

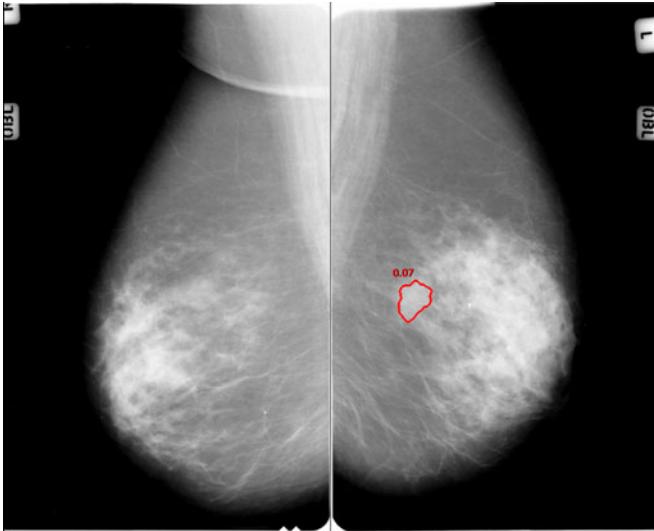


Fig. 6 The same case as in Fig. 5 with the activated CAD region. The red contour and a CAD score close to zero indicate a high probability that this is a cancer

improvement with traditional CAD prompting was found when readers had limited training [23]. This suggest that for mass detection interactive CAD may be more effective than traditional CAD. This is in accordance with studies suggesting that interpretation errors are more common than perception errors [10, 11].

Results obtained in this study show that readers are able to exploit the predictive power of CAD to improve their decisions. This may come as a surprise, because due to the large number of false positives it is often believed that the performance of CAD for masses is much less than that of an experienced reader. It is noted, however, that in a previous study it was shown that the performance of the CAD system was comparable to that of experienced readers when analysis was restricted to locations identified by the radiologists [9]. This is what counts in this study, because CAD results were only shown on regions probed by the readers. Interestingly, malignancy ratings of CAD were also used previously in the large CADET II trial [1] conducted in the UK, where the size of the CAD marks was used to represent the computed likelihood of cancer. Positive results of this trial could also be related to using CAD as a decision support.

The potential gain of using CAD for decision making was also demonstrated in a previous study, in which CAD information was independently combined with reader scores [10]. Results in this study confirm that by independent combination of reader scores with CAD, performance can be improved (Table 2). On average, we found that the improvement in performance was larger when readers used CAD themselves than when CAD was independently combined with their scores. However, the difference was not significant. Interestingly, for one of the radiologists (number 8) detection performance decreased when using interactive CAD, whereas performance increased with independent combination. This may well

be due to insufficient training. Readers need to learn how to weight CAD information in their decisions.

Table 3 shows the average reading times per reader for the sessions with and without CAD. We found that for the non-radiologists the average reading time was slightly reduced when they used CAD. For the radiologists the reading time increased less than 3 s on average with CAD. It seems that interactive use of CAD does not cost much extra time, because the information is presented at the moment the reader asks for it.

In the experiments we used a threshold to adjust the average number of CAD regions per image that could be activated. On average, there were two false positives per normal image. In clinical practice the operating point of prompting systems for masses in mammography are often set to a level near 0.5 false positives per image. We used more regions, because it was thought that in the interactive system more false positives would be tolerable. Many of them are never activated, and if they are activated they are perceived very differently than traditional prompts. The radiologists queried far fewer false-positive CAD regions than the non-radiologists which may indicate they are more confident in their reading.

Interactive CAD is intended to aid the reader in decision making and will not help to avoid perceptual oversights. The success of the interactive approach may be explained by assuming that perceptual oversights do not occur frequently. In our study this appeared to be the case. On average only 5 (12.2%) of the true-positive CAD regions were not probed by the reader. Thus, in the reader study at most 12.2% of the cancers were overlooked, while none of them were reported in the original screening. Results also show that on average 274.2 (50.2%) false-positive CAD regions were not activated, limiting the number of false positives to which the readers are exposed. It is noted that the system can easily be extended by displaying the most suspicious, non-queried CAD regions as traditional prompts after the reading is completed.

In general, the response of the radiologists to the interactive CAD system was very positive and they

Table 3 Mammogram reading times

	Average reading time per case (s)		
	Without CAD	With CAD	P value
Non-radiologists			
1	83.6±47.0	111.5±70.3	0.001
2	84.3±59.2	67.7±42.1	0.03
3	131.1±65.1	129.5±56.9	0.51
4	158.8±68.1	146.0±62.3	0.23
5	33.4±29.6	35.2±29.0	0.45
Average	97.0±70.0	96.7±67.4	0.97
Radiologists			
6	63.1±45.6	58.9±37.8	0.57
7	57.8±31.7	70.8±44.6	0.002
8	73.1±44.1	73.1±31.4	0.42
9	86.7±52.1	88.6±39.1	0.12
Average	70.0±45.1	72.8±39.8	0.02
Reader average	84.7±61.5	85.9±57.8	0.13

Table 4 Number of CAD regions queried

	Queried CAD regions	Non-queried FP CAD regions	Non-queried, unreported TP CAD regions	Non-queried CAD regions but reported TP finding
Non-radiologists				
1	290	293	2	2
2	338	244	3	2
3	330	251	4	2
4	500	83	3	1
5	196	377	7	7
Average	330.8	249.6	3.8	2.8
Radiologists				
6	176	396	8	7
7	262	319	6	0
8	209	365	9	4
9	444	140	3	0
Average	272.75	305	6.5	2.75
Reader average	305	274.22	5	2.78

There were 587 CAD regions in total; 546 false-positive CAD regions and 41 true-positive CAD regions

preferred it to conventional CAD prompting systems. An advantage of the proposed system is that obvious false positives of the CAD system are rarely shown, as the readers do not probe these regions. This may increase confidence in CAD.

In our study the reading conditions were less optimal than in screening practice, because a 4-megapixel color display was, instead of two 5-megapixel grayscale monitors commonly used in mammography. This might have a negative effect on the detection performance, especially for detecting microcalcifications. As microcalcification cases were not included in our study we do not believe that image quality influenced our study outcome. This is supported by a study from Kamitani et al. [24] in which no significant differences were found between the observer performances for detecting breast cancer masses when performing soft-copy reading on 3- or 5-megapixel LCD monitors. Another limitation of our study is the absence of CC views in most cases. In the Dutch screening program, two-view mammography is not always performed at subsequent screens. Obviously, absence of additional CC views might affect the radiologists' detection performance. However, readers in our study are used to interpreting single-view mammography. We would like to note that both limitations did not affect the difference in detection performance described in this paper, because the conditions were similar in the sessions with CAD and the sessions without CAD.

Participants in this study were not reading under normal screening conditions. It may be that their alertness, concentration, and decision thresholds were affected by the knowledge that this study was a controlled laboratory experiment in which their decisions would be recorded and used in a study, and that the balance between cancer and normal cases was artificial. Because their assessments of the mammographic cases in this retrospective observer study would not affect patient care, their decisions could be different from those in an actual clinical setting. This effect has been described, among others, by Gur et al. [25]. However, the reading conditions in the with-CAD and without-CAD sessions were similar, and therefore the

observed effect on detection performance can be attributed solely to the use of the interactive CAD system. Because we performed LROC analysis, decision thresholds did not affect study results.

As in many other studies, the sample was heavily weighted towards cancer cases. Not doing so would make this form of research extremely expensive. The effect on sensitivity and recall rates of radiologists using this interactive CAD system for real-life screening can only be determined by a large randomized controlled trial in which radiologists use this system during routine use and for a substantial period [17]. Nevertheless, a laboratory study is generally a first step to demonstrate the usefulness of a CAD concept before a large trial is performed.

The readers participating in this study had different backgrounds and experience. We expect that when readers gain more experience with the system they will learn how to optimize use of it. In addition, readers need to find out how to weight CAD information in their decisions, and we expect them to improve this when they gain more understanding of the strengths and weaknesses of the CAD software.

Conclusions

We found that in addition to using CAD in the traditional way to avoid perception errors, there is a large potential for using CAD as a decision aid to reduce interpretation failures. Results suggest that interactive CAD may be more effective than traditional CAD for improving mass detection without affecting reading time.

Acknowledgements This work was supported by a grant from the Dutch Cancer Society, KUN 2006-3655.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Gilbert FJ, Astley SM, Gillan MG, Agbaje OF, Wallis MG, James J, Boggis CR, Duffy SW, CADET II Group (2008) Single reading with computer-aided detection for screening mammography. *N Engl J Med* 359:1675–1684. doi:[10.1056/NEJMoa0803545](https://doi.org/10.1056/NEJMoa0803545)
2. Morton MJ, Whaley DH, Brandt KR, Amrami KK (2006) Screening mammograms: interpretation with computer-aided detection—prospective evaluation. *Radiology* 239:375–383. doi:[10.1148/radiol.2392042121](https://doi.org/10.1148/radiol.2392042121)
3. Dean JC, Ilvento CC (2006) Improved cancer detection using computer-aided detection with diagnostic and screening mammography: prospective study of 104 cancers. *Am J Roentgenol* 187:20–28. doi:[10.2214/AJR.05.0111](https://doi.org/10.2214/AJR.05.0111)
4. Fenton JJ, Taplin SH, Carney PA, Abraham L, Sickles EA, D’Orsi C, Berns EACG, Hendrick RE, Barlow WE, Elmore JG (2007) Influence of computer-aided detection on performance of screening mammography. *N Engl J Med* 356:1399–1409. doi:[10.1056/NEJMoa066099](https://doi.org/10.1056/NEJMoa066099)
5. Gur D, Stalder JS, Hardesty LA, Zheng B, Sumkin JH, Chough DM, Shindel BE, Rockette HE (2004) Computer-aided detection performance in mammographic examination of masses: assessment. *Radiology* 233:418–423. doi:[10.1148/radiol.2332040277](https://doi.org/10.1148/radiol.2332040277)
6. Houssami N, Given-Wilson R, Ciatto S (2009) Early detection of breast cancer: overview of the evidence on computer-aided detection in mammography screening. *J Med Imaging Radiat Oncol* 53:171–176. doi:[10.1111/j.1754-9485.2009.02062.x](https://doi.org/10.1111/j.1754-9485.2009.02062.x)
7. Alberdi E, Povvakalo AA, Strigini L, Ayton P, Given-Wilson R (2008) CAD in mammography: lesion-level versus case-level analysis of the effects of prompts on human decisions. *Int J Comput Assist Radiol Surg* 3:115–122. doi:[10.1007/s11548-008-0213-x](https://doi.org/10.1007/s11548-008-0213-x)
8. Taplin SH, Rutter CM, Lehman CD (2006) Testing the effect of computer-assisted detection on interpretive performance in screening mammography. *AJR Am J Roentgenol* 187:1475. doi:[10.2214/AJR.05.0940](https://doi.org/10.2214/AJR.05.0940)
9. Karssemeijer N, Otten JD, Rijken H, Holland R (2006) Computer aided detection of masses in mammograms as decision support. *Br J Radiol* 79:S123–S126. doi:[10.1259/bjr/37622515](https://doi.org/10.1259/bjr/37622515)
10. Karssemeijer N, Otten JD, Verbeek AL, Groenewoud JH, de Koning HJ, Hendriks JH, Holland R (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. Manning D, Ethell S, Donovan T (2004) Categories of observer error from eye tracking and AFROC data. In: Chakraborty DP, Eckstein MP (eds) *Proc SPIE: medical imaging 2004—image perception, observer performance, and technology assessment*. vol 5372. Bellingham, Wash, pp 90–99. doi:[10.1117/12.533489](https://doi.org/10.1117/12.533489)
12. Blanks RG, Wallis MG, Given-Wilson RM (1999) Observer variability in cancer detection during routine repeat (incident) mammographic screening in a study of two versus one view mammography. *J Med Screen* 6:152–158. doi:[10.1136/jms.6.3.152](https://doi.org/10.1136/jms.6.3.152)
13. Khoo LAL, 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)
14. Garvican L, Field S (2001) A pilot evaluation of the R2 image checker system and users’ response in the detection of interval breast cancers on previous screening films. *Clin Radiol* 10:833–837. doi:[10.1053/crad.2001.0776](https://doi.org/10.1053/crad.2001.0776)
15. Alberdi E, Povvakalo A, Strigini L, Ayton P, Hartswood M, Procter R, Slack R (2005) Use of computer-aided detection (CAD) tools in screening mammography: a multidisciplinary investigation. *Br J Radiol* 78:S31–S40. doi:[10.1259/bjr/37646417](https://doi.org/10.1259/bjr/37646417)
16. Warren Burhenne LJ, Wood SA, D’Orsi CJ, Feig SA, Kopans DB, O’Shaughnessy KF, Sickles EA, Tabar L, Vyborny CJ, Castellino RA (2000) Potential contribution of computer-aided detection to the sensitivity of screening mammography. *Radiology* 215:554–562
17. 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)
18. Nishikawa RM (2007) Current status and future directions of computer-aided diagnosis in mammography. *Comput Med Imaging Graph* 31:224–235. doi:[10.1016/j.compmedimag.2007.02.009](https://doi.org/10.1016/j.compmedimag.2007.02.009)
19. Taylor P, Given-Wilson R, Champness J, Potts HW, Johnston K (2004) Assessing the impact of CAD on the sensitivity and specificity of film readers. *Clin Radiol* 59:1099–1105. doi:[10.1016/j.crad.2004.04.017](https://doi.org/10.1016/j.crad.2004.04.017)
20. Gilbert FJ, Astley SM, Boggis CR, McGee MA, Griffiths PM, Duffy SW, Agbaje OF, Gillan MG, Wilson M, Jain AK, Barr N, Beetles UM, Griffiths MA, Johnson J, Roberts RM, Deans HE, Duncan KA, Iyengar G (2008) Variable size cad prompts and mammography film reader decisions. *Breast Cancer Res* 10:R72. doi:[10.1186/bcr2137](https://doi.org/10.1186/bcr2137)
21. Karssemeijer N, Hupse A, Samulski M, Kallenberg M, Boetes C, Heeten G (2008) An interactive computer aided decision support system for detection of masses in mammograms. In: Krupinski EA (ed) *Lect notes comput sci*. 5116. Springer-Verlag, Heidelberg, pp 273–278. doi:[10.1007/978-3-540-70538-3_38](https://doi.org/10.1007/978-3-540-70538-3_38)
22. Taylor P, Champness J, Given-Wilson R, Johnston K, Potts HW (2005) Impact of computer-aided detection prompts on the sensitivity and specificity of screening mammography. *Health Technol Assess* 9:1–70. doi:[10.3310/hta9060](https://doi.org/10.3310/hta9060)
23. Roelofs AA, van Woudenberg S, Hendriks JH, Everts CJ, Karssemeijer N (2004) Effects of computer-aided diagnosis on radiologists’ detection of breast masses. In: Pisano E (ed) *Proceedings of the seventh international workshop on digital mammography*. Chapel Hill, NC, pp 219–224
24. Kamitani T, Yabuuchi H, Soeda H, Matsuo Y, Okafuji T, Sakai S, Furuya A, Hatakenaka M, Ishii N, Honda H (2007) Detection of masses and microcalcifications of breast cancer on digital mammograms: comparison among hard-copy film, 3-megapixel liquid crystal display (LCD) monitors and 5-megapixel LCD monitors: an observer performance study. *Eur Radiol* 17:1365–1371. doi:[10.1007/s00330-006-0452-6](https://doi.org/10.1007/s00330-006-0452-6)
25. Gur D, Bandos AI, Cohen CS, Hakim CM, Hardesty LA, Ganott MA, Perrin RL, Poller WR, Shah R, Sumkin JH, Wallace LP, Rockette HE (2008) The “laboratory” effect: comparing radiologists’ performance and variability during prospective clinical and laboratory mammography interpretations. *Radiology* 249:47. doi:[10.1148/radiol.2491072025](https://doi.org/10.1148/radiol.2491072025)