

Reproducibility of Computer-Aided Detection Marks in Digital Mammography

Seung Ja Kim, MD¹
Woo Kyung Moon, MD²
Nariya Cho, MD²
Joo Hee Cha, MD³
Sun Mi Kim, MD²
Jung-Gi Im, MD²

Index terms:

Breast neoplasms, diagnosis
Computers, diagnostic aid
Digital radiography

Korean J Radiol 2007; 8: 198-205

Received December 21, 2005; accepted after revision August 24, 2006.

¹Department of Radiology, Konkuk University Hospital, Seoul 143-701, Korea; ²Department of Radiology, Seoul National University College of Medicine and Clinical Research Institute, Seoul National University Hospital and the Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul 110-744, Korea; ³Department of Radiology, Boramae Municipal Hospital, Seoul 156-707, Korea

This study is supported by KISTEP, Ministry of Science and Technology, Korea.

Address reprint requests to:

Woo Kyung Moon, MD, Department of Radiology, Seoul National University Hospital, 28, Yongon-dong, Chongno-gu, Seoul 100-744, Korea.
Tel. (822) 2072-2584
Fax. (822) 743-6385
e-mail: moonwk@radcom.snu.ac.kr

Objective: To evaluate the performance and reproducibility of a computer-aided detection (CAD) system in mediolateral oblique (MLO) digital mammograms taken serially, without release of breast compression.

Materials and Methods: A CAD system was applied preoperatively to the full-field digital mammograms of two MLO views taken without release of breast compression in 82 patients (age range: 33–83 years; mean age: 49 years) with previously diagnosed breast cancers. The total number of visible lesion components in 82 patients was 101: 66 masses and 35 microcalcifications. We analyzed the sensitivity and reproducibility of the CAD marks.

Results: The sensitivity of the CAD system for first MLO views was 71% (47/66) for masses and 80% (28/35) for microcalcifications. The sensitivity of the CAD system for second MLO views was 68% (45/66) for masses and 17% (6/35) for microcalcifications. In 84 ipsilateral serial MLO image sets (two patients had bilateral cancers), identical images, regardless of the existence of CAD marks, were obtained for 35% (29/84) and identical images with CAD marks were obtained for 29% (23/78). Identical images, regardless of the existence of CAD marks, for contralateral MLO images were 65% (52/80) and identical images with CAD marks were obtained for 28% (11/39). The reproducibility of CAD marks for the true positive masses in serial MLO views was 84% (42/50) and that for the true positive microcalcifications was 0% (0/34).

Conclusion: The CAD system in digital mammograms showed a high sensitivity for detecting masses and microcalcifications. However, reproducibility of microcalcification marks was very low in MLO views taken serially without release of breast compression. Minute positional change and patient movement can alter the images and result in a significant effect on the algorithm utilized by the CAD for detecting microcalcifications.

A computer-aided detection (CAD) system can assist the radiologist in the early detection of breast cancer by highlighting suspicious areas seen on mammography (1, 2). Reproducibility is one of the important factors for determining the quality of a CAD system. Recent CAD studies involving repeated scanning of film mammograms of breast cancer patients showed 39–53% reproducibility, markedly decreased from the 95–99% claimed by the manufacturers of the CAD systems (3, 4). It is likely that variability of marks is primarily caused by small shifts in the film position between sequential digitizations (5).

Recently, the CAD system has been applied to full-field digital mammography and results for breast cancer detection were reported to be similar to those obtained using an analog system (6). When CAD is used with full-field digital mammography, it

eliminates the need for digitization. Because the CAD system uses a computer algorithm, the CAD system will be 100% reproducible when the same CAD scheme is applied repeatedly to the same digital image. However, it will be possible to take more than one digital image of the same breast in a repeated exposure and it is likely that CAD will show variable reproducibility with such repeated digital images. To our knowledge, there have been few studies to evaluate the reproducibility of a CAD system combined to full-field digital mammography.

The purpose of this study was to evaluate the performance and reproducibility of the CAD system in two mediolateral oblique (MLO) digital mammograms taken serially without release of breast compression.

MATERIALS AND METHODS

Patient Selection

Between March 2004 and June 2004, 100 patients with breast cancer underwent full-field digital mammography (Senographe 2000D FFDM, GE Medical Systems, Buc, France) including two serial MLO views without release of breast compression and one craniocaudal view (i.e., three images per each breast) of the bilateral breasts. Of these 100 patients, we selected 82 consecutive patients in whom the malignant lesions were visible in both the craniocaudal view and the MLO view and the CAD system (ImageChecker M1000-DM, version 3.1; R2 Technology) was applied to these mammograms. The patients had a mean age of 49 years (range: 33–83 years). Because two of the 82 patients had bilateral breast cancer, the total number of diseased breasts was 84. In 68 (81%) of the 84 breasts, the tumors were palpable and in 16 (19%) breasts they were nonpalpable. The mean time interval between the serial MLO views was 18 seconds. The serial MLO mammograms of the bilateral breasts were taken by automatic exposure control methods. The average glandular dose to the breasts calculated by the exposure of the six images was about 8.0–9.0 mGy using our mammography system, which was considered to be acceptable considering the guidelines of the American College of Radiology (ACR) for screening mammography, that is, not to exceed 3.0 mGy per view (7). This study was conducted with institutional review board approval and informed consent was obtained from all patients.

Mammographic and Histological Findings

Because the CAD system identifies the mass component and the calcification component separately, we counted the mass component and the calcification component separately for those malignancies that presented with both

a mass and a calcification cluster. On mammograms of the 84 diseased breasts of 82 patients, the total number of visible lesion components was 101. We found 66 masses (mean size, 24 mm) and 35 microcalcifications (mean size, 21 mm) including one mass (n = 50), one microcalcification (n = 20), one mass plus one microcalcification (n = 7), two masses (n = 2), two microcalcifications (n = 2), two masses and one microcalcification (n = 2), and one mass and two microcalcifications (n = 1). The categories of the lesions in the 84 diseased breasts using the American College of Radiology's Breast Imaging Reporting and Data System (ACR BI-RADS) (8) were category 4 (n = 8) and category 5 (n = 76). The size of the mass was ≤ 10 mm (n = 3), 11–20 mm (n = 24), 21–30 mm (n = 27), 31–40 mm (n = 6), and ≥ 41 mm (n = 6) and the size of the calcific cluster was ≤ 10 mm (n = 16), 11–20 mm (n = 9), 21–30 mm (n = 5), 31–40 mm (n = 1), and ≥ 41 mm (n = 4).

Utilizing the density pattern of the ACR BI-RADS, we divided the patients into two subgroups: 1) dense breast group (BI-RADS 3 and 4 density in 50 patients), and 2) fatty breast group (BI-RADS 1 and 2 density in 32 patients).

The final postoperative pathological diagnoses were all malignant (ductal carcinoma in situ [DCIS] [n = 15] and invasive carcinoma [n = 69]).

CAD Mark Evaluation

We applied the CAD system to each set of digital mammograms including the craniocaudal and serial MLO views of the 82 patients and saved the images with CAD marks at a review workstation before sending them to the Picture Archiving Communication System (PACS). We then analyzed the CAD marks of each view by reviewing the original digital mammograms, the images with CAD marks, ultrasound and the pathology reports. The consensus of two breast imaging specialists was used in the interpretation.

The CAD system marks regions suspicious for a mass or a microcalcification cluster by superimposing a small asterisk or a triangle, respectively, on the image. If the asterisk was located within a true positive mass, this mass was considered to have been identified correctly by the CAD system. Similarly, if the triangle overlapped any of the microcalcification areas, the CAD marks were considered to represent true positive detection. Because all patients had ultrasound preoperatively, any mass marks that fell on normal parenchyma identified on ultrasound were considered as false positives. When the CAD system marked typical benign calcifications or crossing lines, we considered them as false positives.

CAD Performance Analysis

The sensitivity, reproducibility and the false positive marks per image of the CAD system in these 82 patients were analyzed. We analyzed sensitivity in the first MLO views, the second MLO views, craniocaudal views, and in a combination of the first MLO views and craniocaudal views. The latter was characterized as case-based sensitivity and the remaining as image-based sensitivity (i.e., sensitivity of each mammographic view).

For case-based sensitivity, a successful mark for the cancer in either the MLO or craniocaudal view was considered a true-positive identification of the cancer for that case. Breast parenchymal density was used in these analyses in order to show its effect on the sensitivity.

We analyzed the image-based reproducibility and the mark-based reproducibility of the CAD system in serial MLO views of ipsilateral diseased breasts and contralateral normal breasts, respectively, and together. Image-based reproducibility was defined as the identical images and the value was obtained in two different ways: one way by analyzing images irrespective of the existence of CAD marks and another way by analyzing images with CAD marks. The mark-based reproducibility was defined as any marks within the same mass or the same microcalcification in the serial MLO images. For example, as seen in Figures 1A and B, the CAD marks on the first MLO view were not identical with that on the second MLO view. Thus, the two serial MLO views were not reproducible in terms of image-based reproducibility. The true mass mark on the first MLO view was identical with that on the second MLO view, so this CAD mark was reproducible in terms of mark-based reproducibility. We analyzed the reproducibility of true and false positive marks in serial MLO views of bilateral breasts.

The false positive marks per image (of a mass or a microcalcification) were obtained in serial MLO views and craniocaudal views of ipsilateral breasts and contralateral

breasts, respectively and then together.

In order to evaluate of the effect of the variables such as kVp, mAs and breast thickness, we calculated the differences of the exposure parameters between the two MLO views and presented them as percentages in reproducible and in non-reproducible serial MLO sets. In all patients, we also calculated the average glandular dose delivered to the patient during the exposure of the six images (i.e., two serial MLO views and one craniocaudal view of bilateral breasts) by adding the amount of the dose appeared on the digital images.

Statistical Analysis

To compare the differences in the sensitivity of the first MLO views and the second MLO views, the paired *t*-test was applied. To compare the differences of sensitivity for lesions and reproducibility of CAD marks in fatty breast and dense breast and the difference in false positive calcification marks per image in the first MLO views and the second MLO views, the unpaired *t*-test was applied. The differences of the exposure parameters between the two MLO views were compared for reproducible and non-reproducible serial MLO sets using the unpaired *t*-tests.

RESULTS

Sensitivities

Table 1 summarizes the sensitivity of the CAD system in serial MLO views and in craniocaudal views of 82 patients. The case-based sensitivity was 86% (57 of 66) for masses and 100% (35 of 35) for microcalcifications. The case-based sensitivity for masses in the fatty breast group was 93% (25 of 27) and that in the dense breast group was 82% (32 of 39). This difference was not statistically significant (*p* = 0.08). The image-based sensitivity of the CAD system for masses was 71% (47 of 66) in the first MLO views and 68% (45 of 66) in the second MLO views. The

Table 1. Detection Performance of the CAD System in Two Serial Mediolateral Oblique and Craniocaudal Views

	1st MLO (%)		2nd MLO (%)		CC (%)		1st MLO + CC (%)	
	Mass	Ca ⁺⁺	Mass	Ca ⁺⁺	Mass	Ca ⁺⁺	Mass	Ca ⁺⁺
Total (n = 82)	71 (47/66)	80* (28/35)	68 (45/66)	17* (6/35)	70 (46/66)	91 (32/35)	86 (57/66)	100 (35/35)
Fatty (n = 32)	93 [†] (25/27)	75 (9/12)	93 [‡] (25/27)	25 (3/12)	81 (22/27)	100 (12/12)	93 (25/27)	100 (12/12)
Dense (n = 50)	56 [†] (22/39)	83* (19/23)	51 [‡] (20/39)	13* (3/23)	62 (24/39)	87 (20/23)	82 (32/39)	100 (23/23)

Note. — numbers are percentages, with raw data in parentheses.

MLO = mediolateral oblique view, CC = craniocaudal view, 1st = first, 2nd = second Ca⁺⁺ = microcalcification, n = patient number

* This difference was statistically significant (*p* < 0.0001 by two-tailed paired *t*-test).

^{† ‡} These differences were statistically significant ([†]*p* = 0.0011 and [‡]*p* = 0.0003 by two-tailed unpaired *t*-test).

CAD Marks in Digital Mammography

image-based sensitivity for microcalcifications was 80% (28 of 35) in the first MLO views and 17% (6 of 35) in the second MLO views (Figs. 1, 2). The sensitivity differences for microcalcifications in the serial MLO views were statistically significant by the paired *t*-test ($p < 0.0001$).

Reproducibility

The image-based reproducibility was 35% (29 of 84) in the ipsilateral serial MLO mammogram sets (Table 2). When we excluded the six images that had no CAD marks,

the reproducibility in the remaining 78 images was reduced to 29% (23 of 78). The image-based reproducibility of the contralateral breasts was 65% (52 of 80) and when we excluded 41 images that had no CAD marks, the image-based reproducibility fell markedly to 28% (11 of 39). In the serial MLO views of the bilateral breasts, the image-based reproducibility regardless of the existence of CAD marks was 49% (81 of 164) and the reproducibility with CAD marks was 29% (34 of 117).

The mark-based reproducibility for true positive mass

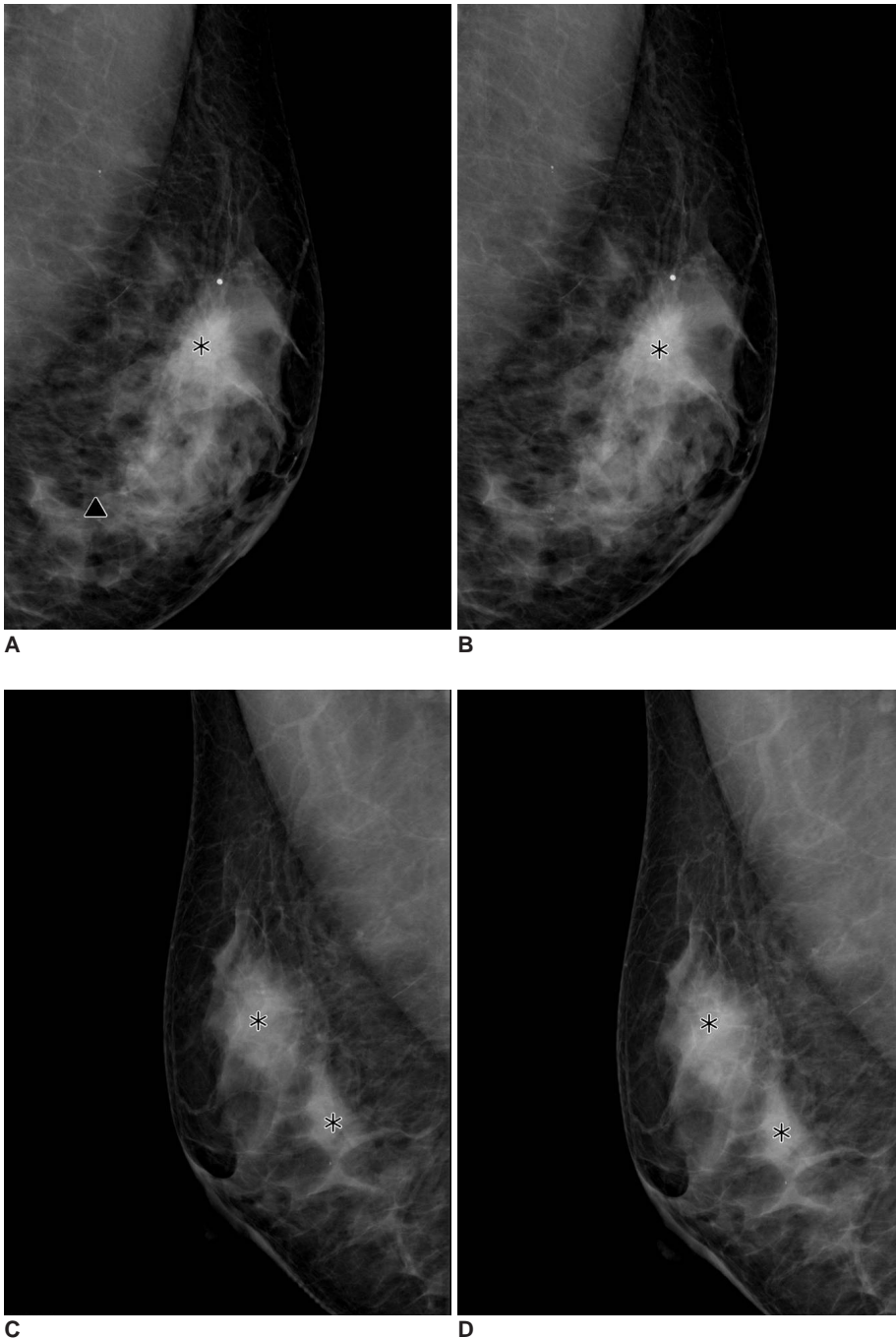


Fig. 1. A 49-year-old woman with an infiltrating ductal carcinoma in the left breast that was confirmed by a biopsy. **A, B.** Serial mediolateral oblique digital mammograms taken without release of breast compression within 17 seconds. The computer-aided detection system marks the mass in the left upper breast correctly on both mediolateral oblique images (asterisks). The false positive calcification mark (triangle) is seen below the mass mark on the first mediolateral oblique view, but not on the second mediolateral oblique view. **C, D.** Serial mediolateral oblique digital mammograms of contralateral breast taken within 17 seconds. The computer-aided detection system marks normal parenchyma areas on the first mediolateral oblique view (**C**) (asterisks) and the computer-aided detection system marks the false positive mass repeatedly on the second mediolateral oblique view (**D**).

marks in the serial MLO views was 84% (42 of 50) and that for true positive microcalcification marks was 0% (0 of 34) (Table 3) (Figs. 1, 2). The 28 calcific clusters that were correctly marked on the first MLO views were not marked on the second MLO views, while six calcific clusters that were correctly marked on the second MLO views were not marked on the first MLO views. The reproducibility for true positive mass marks was significantly higher in the fatty breast group than in the dense breast group (100% vs. 68%, $p = 0.0015$) (Fig. 3). For ipsilateral breasts, the reproducibility for false positive mass marks was 42% (8 of 19) and that for false positive microcalcification marks was 0% (0 of 19) (Fig. 1). For contralateral breasts, the reproducibility for false positive mass marks was 44% (16 of 36) and that for false positive microcalcification marks was 0% (0 of 18) (Fig. 1). In total, for bilateral breasts, the reproducibility for false positive mass marks was 44% (24 of 55) and that for false positive microcalcification marks was 0% (0 of 37). The reproducibility for false positive mass marks in bilateral breasts was slightly higher in the fatty breast group than in the dense breast group, but that difference was not found

to be statistically significant (48% [12 of 25] vs. 40% [12 of 30], $p = 0.5$).

The false positive marks per image on the first MLO views for bilateral breasts were 0.45, with 0.24 mass marks and 0.21 calcification marks and those for the second MLO views of bilateral breasts were 0.26, with 0.24 mass marks and 0.02 calcification marks. The false positive calcification marks per image were significantly lower in the second MLO views ($p < 0.0001$). The false positives per image on craniocaudal views of bilateral breasts were 0.41, with 0.19 mass marks and 0.22 calcification marks.

The changes of the exposure parameters in 81 reproducible serial MLO mammogram sets and in 83 non-reproducible serial MLO mammogram sets were 0.2% and 0.5% in kVp, 5% and 7% in mAs and 0.5% and 1% in breast thickness. These differences were not statistically significant ($p = 0.2043, 0.2987, 0.1058$, respectively). The two radiologists who analyzed cases in our study found little difference in the two serial MLO mammograms.

The total average glandular doses delivered to the patients during the exposure of two serial MLO views and one craniocaudal view of bilateral breasts ranged from 4.5

Table 2. Image-based Reproducibility of Medirolateral Oblique Mammograms by the CAD System

	Ipsilateral Breast (%)		Contralateral Breast (%)		Bilateral Breasts (%)	
	Mark (+/-)*	Mark (+)†	Mark (+/-)*	Mark (+)†	Mark (+/-)*	Mark (+)†
Total	35	29	65	28	49	29
(n = 82)	(29/84 ‡)	(23/78)	(52/80)	(11/39)	(81/164)	(34/117)
Fatty	39	39	65	21	52	34
(n = 32)	(13/33)	(13/33)	(20/31)	(3/14)	(33/64)	(16/47)
Dense	31	22	65	32	48	26
(n = 50)	(16/51)	(10/45)	(32/49)	(8/25)	(48/100)	(18/70)

Note.— numbers are percentages, with raw data in parentheses.

n = patient number

*These columns are percentages of the identical images irrespective of the existence of CAD marks.

† These columns are percentages of the identical images with CAD marks.

‡ Two patients had bilateral cancers.

Table 3. Reproducibility of CAD Marks in Serial Medirolateral Oblique Views by the CAD System

	Ipsilateral Breast (%)				Contralateral Breast (%)	
	TP mass	TP ca**	FP mass	FP ca**	FP mass	FP ca**
Total	84*	0	42*	0	44	0
(n = 82)	(42/50)	(0/34)	(8/19)	(0/19)	(16/36)	(0/18)
Fatty	100†	0	55	0	43	0
(n = 32)	(25/25)	(0/12)	(6/11)	(0/10)	(6/14)	(0/6)
Dense	68†	0	25	0	45	0
(n = 50)	(17/25)	(0/22)	(2/8)	(0/9)	(10/22)	(0/12)

Note.— numbers are percentages, with raw data in parentheses.

TP = true positive; FP = false positive; ca** = microcalcification; n = patient number

* † These differences were statistically significant (* $p = 0.0003$ and † $p = 0.0015$ by two-tailed unpaired *t*-test).

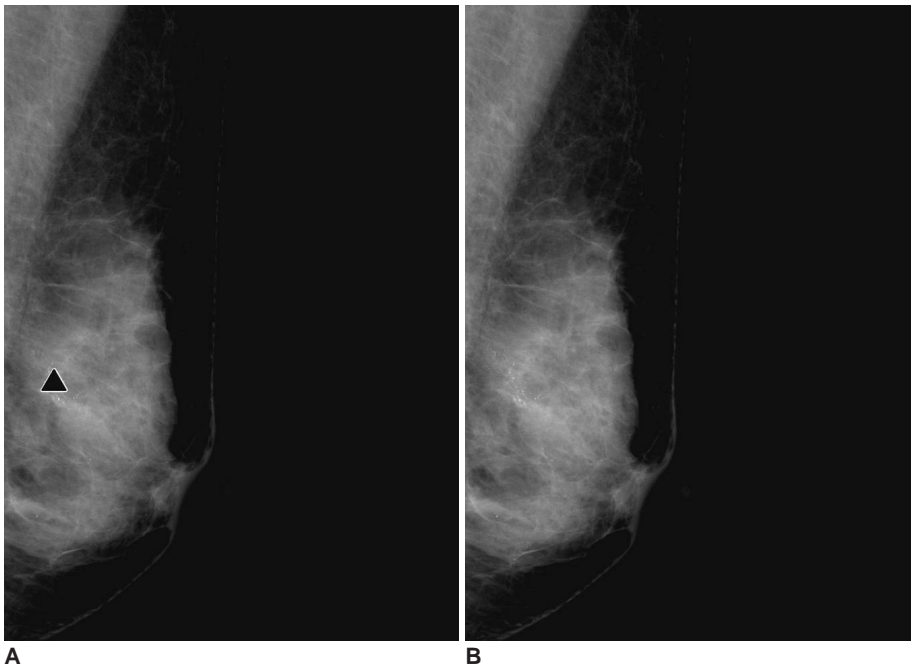


Fig. 2. A 56-year-old woman with malignant microcalcifications in the left breast.
A, B. Serial mediolateral oblique digital mammograms taken without release of breast compression within 17 seconds. The computer-aided detection system marks the microcalcifications correctly on the first mediolateral oblique view (**A**) (triangle). The computer-aided detection system, however, does not mark the microcalcifications on the second mediolateral oblique view (**B**).

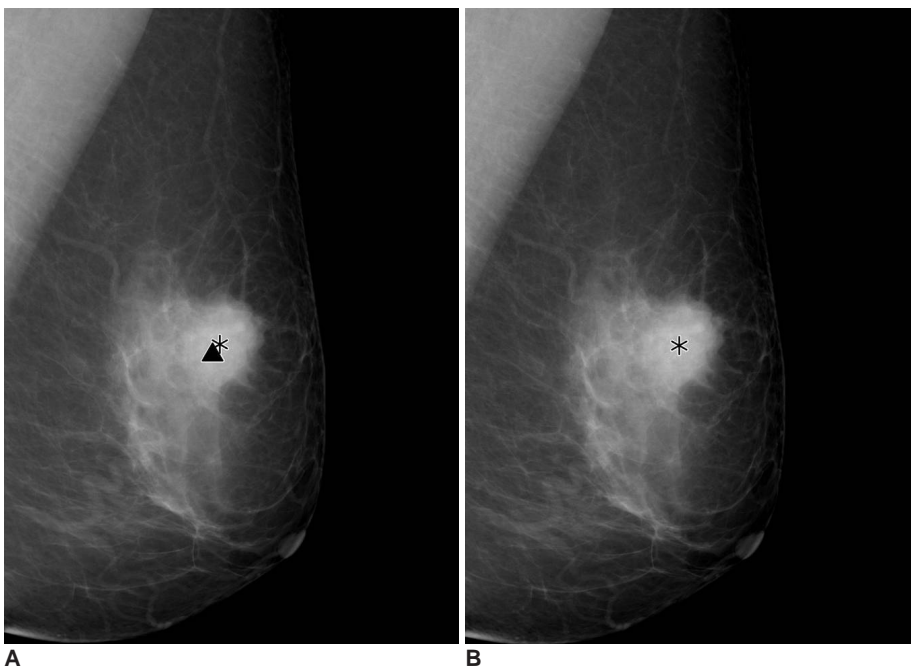


Fig. 3. A 61-year-old woman presented with a screening-detected breast cancer in the left breast.
A, B. Serial mediolateral oblique digital mammograms taken without release of breast compression within 19 seconds. The computer-aided detection system marks the mass component and the microcalcifications component correctly on the first mediolateral oblique view (**A**) (asterisk and triangle). The computer-aided detection system, however, does not mark the microcalcifications on the second mediolateral oblique view (**B**).

mGy to 12.5 mGy, with a mean dose of 8.8 mGy. The doses were less than 8 mGy in 22 patients, 8–12 mGy in 59 patients, and over 12 mGy in only one patient.

DISCUSSION

In our study, patients with breast cancers underwent two serial MLO mammograms of the ipsilateral and contralateral breasts without release of breast compression to evaluate the reproducibility of the CAD system applied to

full-field digital mammography. There should be only minute positional change and patient movement between the two mammograms since the same breast was exposed twice without release of breast compression. The results of our study showed that the sensitivity of the CAD system for malignant masses was similar in two serial MLO views: 71% (47 of 66) and 68% (45 of 66) while the sensitivity for microcalcifications was quite different: 80% (28 of 35) and 17% (6 of 35). The reproducibility of CAD marks for true positive mass in serial MLO views was 84% (42 of 50)

and that for true positive microcalcification was 0% (0 of 34). Our results suggest that even a minute positional change can cause a significant effect on the algorithm of the CAD system for detecting malignant microcalcifications in full-field digital mammography.

In current CAD systems, a binary threshold is typically used to generate detection marks. Each marked region has a computed score that is above a predetermined threshold; lesions with computed scores that are near the threshold are vulnerable to small changes and may be detected in one image and missed in another image (3). In our study, the sensitivity of the CAD system for detecting microcalcifications was significantly lower in the second MLO study. A minimal positional change might have affected pixels thus resulting in the same variability as in repeated scanning of the same films. In addition, motion artifacts and blurring are also possible causes of low reproducibility, especially in microcalcification with detailed structures. In our study, it is likely that the false positive microcalcification marks per image were substantially lower in the second MLO views for the same reason ($p < 0.0001$).

With digital mammography, the CAD system does not require a digitizer and allows display of the CAD markers rapidly after the image acquisition. Because the CAD system uses a computer algorithm, the CAD system is 100% reproducible when the same CAD scheme is applied repeatedly to the same digital image. This is different from a CAD system using film mammography. Our study showed, however, that CAD in full-field digital mammography could show variable reproducibility with repeated images of the same breast.

In our study, the case-based sensitivity of the CAD system for masses and microcalcifications was 86% (57 of 66) and 100% (35 of 35), respectively. This is similar to or better than seen in previous studies using film or digital mammography and a CAD system (3–7, 9–10). It is difficult to compare directly the detection performance of these studies, because different image databases and case selection criteria were used. The image-based sensitivity for masses was higher in the fatty breast group than in the dense breast group in serial MLO views (93%, 93% vs. 56%, 51%) and these differences were statistically significant ($p = 0.0011, 0.0003$). The different sensitivity for masses relating to the parenchymal density is consistent with the results of the study of Brem et al. (10).

There are several limitations to our study. When we calculated the differences of the exposure parameters between the two MLO views in the reproducible and in the non-reproducible serial MLO sets, the changes of parameters such as kVp, mAs and breast thickness in the non-reproducible serial MLO mammogram sets were slightly

greater than those in the reproducible serial MLO mammogram sets. These parameters may affect the reproducibility of the CAD system. Subtraction of the two serial images in reproducible and in non-reproducible serial MLO sets could suggest changes in position and thus a reason for inconsistency of the CAD system, even though the radiologists in this study found no difference between the two serial mammograms. A method to test the reproducibility of computer-aided detection schemes has been recently described for digitized mammograms (11). We selected cases in which lesions were visible in both craniocaudal and MLO views, and thus the overall sensitivity of the CAD system may be overestimated.

In conclusion, the CAD system in full-field digital mammography showed a high sensitivity for detecting masses and microcalcifications related to breast cancer. However, the reproducibility of microcalcification CAD marks was very low in two MLO views taken serially without release of breast compression. Minute positional change and patient movement between serial mammograms might have a significant effect on the algorithm of the CAD system in detecting microcalcifications. Reproducibility of CAD system remains an important issue in full-field digital mammography.

References

1. Warren Burhenne LJ, Wood SA, D'Orsi CJ, Feig SA, Kopans DB, O'Shaughnessy KF, et al. Potential contribution of computer-aided detection to the sensitivity of screening mammography. *Radiology* 2000;215:554-562
2. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology* 2001;220:781-786
3. Zheng B, Hardesty LA, Poller WR, Sumkin JH, Golla S. Mammography with computer-aided detection: reproducibility assessment initial experience. *Radiology* 2003;228:58-62
4. Malich A, Azhari T, Bohm T, Fleck M, Kaiser WA. Reproducibility—an important factor determining the quality of computer-aided detection (CAD) systems. *Eur J Radiol* 2000;36:170-174
5. Taylor CG, Champness J, Reddy M, Taylor P, Potts HW, Given-Wilson R. Reproducibility of prompts in computer-aided detection (CAD) of breast cancer. *Clin Radiol* 2003;58:733-738
6. Baum F, Fischer U, Obenaus S, Grabbe E. Computer-aided detection in direct digital full-field mammography: initial results. *Eur Radiol* 2002;12:3015-3017
7. American College of Radiology. *Mammography quality control manual*. Reston, VA: American College of Radiology, 1999:284-285
8. American College of Radiology. *Breast imaging reporting and data system: BI-RADS atlas*, 4th ed. Reston, VA: American College of Radiology, 2003
9. Baker JA, Lo JY, DeLong DM, Floyd CE. Computer-aided detection in screening mammography: variability in cues. *Radiology* 2004;233:411-417
10. Brem RF, Hoffmeister JW, Rapelyea JA, Zisman G,

CAD Marks in Digital Mammography

- Mohtashemi K, Jindal G, et al. Impact of breast density on computer-aided detection for breast cancer. *AJR Am J Roentgenol* 2005;184:439-444
11. Zheng B, Gur D, Good WF, Hardesty LA. A method to test the reproducibility and to improve performance of computer-aided detection schemes for digitized mammograms. *Med Phys* 2004;31:2964-2972