Clinical Radiology (2001) **56,** 150–154
doi:10.1053/crad.2000.0592, available online at http://www.idealibrary.com on IDEAL®

Radiologist Detection of Microcalcifications With and Without Computer-Aided Detection: A Comparative Study

RACHEL F. BREM*, JOELLE M. SCHOONJANS*

*The Breast Imaging and Interventional Center, *The George Washington University Medical Center,, 2150 Pennsylvania Avenue, Washington DC 20037, U.S.A. The Johns Hopkins Medical Institutions, 601 North Caroline Street, Baltimore, MD 21287, U.S.A.*

Received: 6 April 2000 Revised: 8 August 2000 Accepted: 19 August 2000

AIM: To compare the sensitivity and specificity of microcalcification detection by radiologists alone and assisted by a computer-aided detection (CAD) system.
MATERIALS AND METHODS: Films of 106 patients were masked, randomized, digitized and analysed by the CAD-system. Five readers interpreted the original mammograms and were blinded to demographics, medical history and earlier films. Forty-two mammograms with malignant microcalcifications, 40 with benign microcalcifications and 24 normal mammograms were included. Results were recorded on a standardized image interpretation form. The mammograms with suspicious areas flagged by the CAD-system were displayed on mini-monitors and immediately re-reviewed. The interpretation was again recorded on a new copy of the standard form and classified according to six groups.
RESULTS: Forty-one out of 42 (98%) malignant microcalcifications and 32 of 40 (80%) benign microcalcifications were flagged by the CAD-system. There was an average of 1.2 markers per image. The sensitivity for malignant microcalcifications detection by mammographers without and with the CAD-system ranged from 81% to 98% and from 88% to 98%, respectively. The mean difference without and with CAD-system was 2.2% (range 0–7%).
CONCLUSION: No statistically significant changes in sensitivity were found when experienced mammographers were assisted by the CAD-system, with no significant compromise in specificity.
Brem, R. F., Schoonjans, J. M. (2001). *Clinical Radiology* **56**, 150–154.

Key words: microcalcification, CAD, mammogram.

Breast cancer is the most common cancer and the second leading cause of cancer death in women, surpassed only by lung cancer [1]. In 1999, 176 300 new breast cancer cases and 43 700 breast cancer deaths are expected in the U.S.A. [1]. Currently, mammography is the most sensitive method for the detection of breast cancer. Nevertheless, various studies have shown that a significant number of biopsy-proven malignancies are not detected by mammography, emphasizing that mammography provides an imperfect examination [2–4]. Indeed, in the Breast Cancer Detection Demonstration Project (BCDDP), 10% of breast cancers were not detected on mammograms [5]. Additional studies by Beam *et al*. revealed that up to 20% of breast cancer is not visible by mammography [6]. Moreover, there is wide interobserver variability in the interpretation of mammograms; when mammograms are read by two readers the rate of breast cancer detection has been shown to increase by up to 15% [7–9]. The complex structure of breast tissue, in particular dense breasts, oversight, poor quality films and eye fatigue are reasons why cancer is not detected by mammography [10]. Clearly, any methods that would enhance the ability to overcome the difficulties in reading mammograms would be a major advance.

In the last decade, computer-aided detection (CAD) has been developed to improve the sensitivity of mammographic detection of breast cancer [11–14]. The mammograms are digitized and evaluated with algorithms, often using artificial neural networks [15]. The final diagnosis is made by the radiologist using information supplied by the CAD-system. Thus, computer-aided detection can perform as an adjacent to mammographic interpretation by a radiologist which complements rather than replaces the human observer [16,17]. Consequently, mammographic detection should be improved and it is conceivable that cancers could be detected at an earlier stage. Few studies exist to evaluate the ability of CAD-system to assist experienced mammographers to identify lesions that might be missed for any of the reasons described above [16,18,19].

The present study was designed to compare the sensitivity and specificity of interpretation of mammograms by

Author for correspondence and guarantor of study: Rachel F. Brem, M.D., Director, Breast Imaging and Intervention, The George Washington University Medical Center, 2150 Pennsylvania Avenue, Washington, DC 20037, U.S.A. E-mail: radrfb@gwumc.edu

0009-9260/01/020150+05 $35.00/0

experienced mammographers alone and by the same radiologists assisted by the CAD-system for detection of microcalcifications as indicators of breast cancer. We specially chose to assess the effect of CAD on the detection of microcalcifications by radiologists who were experienced mammographers with the assumption that an even greater beneficial effect could be expected when radiologists not specifically trained in mammography utilize CAD for breast cancer detection.

MATERIALS AND METHODS

Study Design

The study required evaluation of mammograms by five highly experienced mammographers alone and then by the same mammographers in conjunction with a CAD-system. The original mammograms of 106 patients were consecutively selected (23 two-view and 83 four-view mammograms for a total of 378 original images). The case mix consisted of 24 normal mammograms, 40 with benign microcalcifications and 42 with malignant microcalcifications. Forty-seven malignant masses and 35 benign masses cases were included in the reading session to ensure that all types of suspicious mammographic features were included; however, the masses were not a part of this study analysis. The reference standard required a minimum of 1-year follow-up for normal cases and pathological proof for benign and malignant lesions. Supporting documentation was obtained for all mammograms included in the case sample, including mammogram reports for all studies, 1-year follow-up reports for normal studies and pathology reports for proven benign and malignant lesions. A standard reference form was completed by an expert radiologist (RB) for each case. Technical image quality was reported according to a three-point scale corresponding to the qualitative labels: adequate, good, excellent. Technically inadequate mammograms were excluded. Pathologically proven lesions were classified by lesion type, appearance and location. Subtlety of lesions was classified by a three-point scale corresponding to the qualitative labels: obvious, typical, subtle (Table 1). The lesion size in millimetres was measured on the film and also recorded (Table 1).

CAD-system Components

Original films were labelled and masked for patient confidentiality. The laser film scanner of the CAD-system (Image Checker®, R2 Technology, Inc. Los Altos, CA, U.S.A.) digitized mammograms at high resolution (50 micron pixels), with 12 bit depth (4096 shades of gray) and wide latitude (film density range of 0 to 4.5). The processor unit of the CAD-system analysed the mammograms in real time by utilizing neural network software. It took approximately 10 min to digitize and process 18 by 24-cm mammograms and 15 min for 24 by 30-cm mammograms. Microcalcifications were highlighted by the CAD-system with solid triangles. The display unit highlighted these markers on mini-monitors that flagged areas of interest for the radiologist. Before the study, the digitized mammograms were evaluated by the CAD-system to determine the proportion of microcalcifications that would be flagged by the CAD-system markers. The number of markers per image was counted and the average number of markers per image was calculated. For this study the system was designed to detect clusters consisting of five or more microcalcifications by a propriety algorithm.

Radiologist Selection and Reading Sessions

Each of the five radiologists selected was experienced in mammography and had read a minimum of 4000 mammograms annually for the past 5 years. Two were from academic institutions and three from private practice. Three had had fellowship training in mammography. The radiologists were unaware of the case mix of normals, cancers and benign calcifications in this study. Each underwent a 1-h training session before the study. The training session consisted of reviewing six cases which included normal, benign and malignant microcalcifications, followed by an independent test set of six cases which included normal, benign and malignant microcalcifications followed by a review of the test set results. Four reading sessions were conducted for each radiologist to interpret the mammograms. Approximately 26 cases were interpreted during each session for a total of 94 films per session. Films were given in randomized sequence order generated by computer and were read first by the radiologist alone and then with the assistance of the CAD-system. The radiologists were blinded to the patient demographics and clinical history. No earlier films were available. The reading time per case was not limited. The mean session duration was 57.2 min (range, 52.4 to 63.4 min).

The radiologist first filled out the interpretation form without benefit of the CAD-system. Then, after reviewing highlighted areas, the radiologists made a final diagnosis and completed a second image interpretation form. The form required giving lesion type and assessment using the American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS) assessment terminology (Table 2). In addition, the location of the lesion was documented diagrammatically. Any

Table 1 – Lesion characteristics

Max. size	Microcalcifications distribution			Microcalcifications characteristics			
(mean,mm)	Clustered (%)	Segmental (%)	Occasional (%)	Obvious (%)	Typical (%)	Subtle (%)	
Malignant	10.75 ± 7 SD	36 (85.7)	5 (11.9)	1 (2.4)	7 (16.7)	16 (38.1)	19 (45.2)
Benign	9.4 ± 7 SD	36 (90.0)	3 (7.5)	1 (2.5)	4 (10.0)	17 (42.5)	19 (47.5)

Table 2 – Lesion score according to the American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS)

Negative
(1) Negative (routine follow-up)
(2) Benign finding (routine follow-up)
(3) Probably benign finding (short interval follow-up suggested)
Positive
(4) Suspicious finding (biopsy should be considered)
(5) Highly suggestive of malignancy (appropriate action should be taken)
(0) Need additional imaging evaluation (immediate recall)

lesion, even if documented on only one view, either craniocaudal or medial lateral oblique, was counted.

Statistical Analysis

Sensitivity and specificity were calculated for interpretation of films by each radiologist alone and then assisted by the CAD-system. The mean difference in sensitivity and specificity across observers (radiologists alone compared to radiologists assisted by the CAD-system) was evaluated by using a paired *t*-test. Statistical significance was inferred at P values less than 0.05. Sensitivity of CAD-system detection of malignant and benign microcalcification was determined. The difference in the CAD-system ability to detect malignant and benign microcalcification was evaluated using Fisher's exact test.

RESULTS

Performance of the CAD System

Overall, there was an average of 1.2 ($\pm$ 1.2 SD, range 0–3) markers per image.

The CAD-system markers flagged 41 out of 42 pathologically proven malignant microcalcifications (98% sensitivity). Five lesions were marked on only one view (craniocaudal or media-lateral oblique), four of which were classified as subtle. The lesion that was not marked on either view was also classified as subtle. There was an average of 1 ($\pm$ 1 SD) false-positive markers per image in the group of malignant cases.

The CAD-system flagged 32 of 40 (80%) pathologically proven benign microcalcifications. Six lesions were marked on only one view, of which four were classified as subtle. Six of the eight benign microcalcifications not detected by the CAD-system were classified as subtle. An average of 0.8 false-positive markers per image was recorded in the group of benign cases. There was a statistically significant difference in the CAD-system's detection of malignant and benign microcalcifications (Fisher's exact test, $P < 0.05$).

The CAD-system correctly identified 11 of 24 (45.8%) normal cases as having no area of concern for the presence of microcalcifications. An average of 0.5 false-positive markers per image was counted in the normal cases.

Of the malignant microcalcifications, 25 (59.5%) were ductal carcinoma in situ, 16 (38.1%) were infiltrating ductal carcinoma, and one (2.4%) was an infiltrating carcinoma with duct and lobular features. Pathology of the benign microcalcifications included 17 cases (42.5%) of fibrocystic changes, seven (17.5%) fibroadenomas, three (7.5%) atypical ductal hyperplasia, and two cases (5%) each of lobular carcinoma *in situ*, adenosis, apocrine metaplasia and papilloma. Other benign causes were responsible for five cases (12.5%) (Table 3).

Lesions were scored as negative if the radiologist determined that the mammogram had no findings and was put into routine follow-up, showed a benign lesion and the patient was put into routine follow-up, or showed a probably benign finding and the patient was asked to return for short-term follow-up. These were based on the American College of Radiology BI-RADS System (Table 2).

Lesions were scored as positive if the radiologist determined that the patient required a call back for additional evaluation. Patients with probably benign findings who were asked to return for immediate recall, those with suspicious findings, and those with findings that suggested the presence of malignancy were included (Table 2).

Analysis of Radiologists' Performance With and Without CAD

The mean sensitivity for detection of malignant microcalcifications increased from 89.6% (range, 81% to 98%) for

Table 3 – Pathology

	%
Malignant	
Ductal carcinoma in-situ	25 (59.5)
Infiltrating ductal carcinoma	16 (38.1)
Infiltrating carcinoma with ductal and lobular features	1 (2.4)
Benign	
Fibrocystic changes	17 (42.5)
Fibroadenomas	7 (17.5)
Atypical ductal hyperplasia	3 (7.5)
Lobular carcinoma in-situ	2 (5)
Adenosis	2 (5)
Apocrine metaplasia	2 (5)
Papilloma	2 (5)
Other	5 (12.5)

Table 4 – Sensitivity for detection of malignant microcalcifications

Radiologist	Microcalcifications detected	
	Radiologist alone (%)	Radiologist with CAD-system (%)
1	88	90
2	98	98
3	81	88
4	88	90
5	93	93
Mean $\pm$ SD	89.6 $\pm$ 6.3	91.8 $\pm$ 3.9

Of the mammograms examined, 42 had malignant microcalcifications. Those classified as Groups 1, 2 and 3 were considered negative; those in Groups 4, 5 and 0 positive. See text for details.

Table 5 – Specificity for detection of benign microcalcification and normal cases

Radiologist	Benign microcalcifications and normal cases	
	Radiologist alone (%)	Radiologist with CAD-system (%)
1	50	48
2	55	53
3	61	61
4	42	41
5	33	33
Mean ± SD	48.2 ± 11.0	47.2 ± 10.8

Of the mammograms examined, 40 had benign microcalcifications and 24 were normal. Those classified as Groups 1, 2, and 3 were considered negative for the presence of cancer; those in Groups 0, 4, and 5 positive. See text for details.

radiologists alone to 91.8% (range, 88% to 98%) for radiologists assisted by the CAD-system (Table 4). The mean specificity for benign microcalcifications and normal cases decreased from 48.2% for radiologists alone to 47.2% for radiologists assisted by the CAD-system with a range in values of both from 33% to 61% (Table 5). Paired *t*-tests were performed to examine the mean difference in sensitivity and specificity across observers (radiologists alone vs with CAD-system). On the basis of this limited sample size, there were no significant differences in sensitivity or specificity between the two groups.

DISCUSSION

In this study, the ability to detect malignant microcalcifications with mammograms by experienced mammographers was compared with detection by the same mammographers together with the computer-aided detection system (CAD-system). We found that no statistically significant improvement in sensitivity was obtained when the CAD-system was used in conjunction with the radiologist, without compromise in specificity. In this study, the improvement in sensitivity was not statistically significant, although one of the five readers had a 7% increase in sensitivity when assisted with the CAD-system, while the others showed a 0 to 2% increase (Table 4). We had specifically chosen to assess the effect of CAD on the detection of microcalification by radiologists who were experienced mammographers with the assumption than an even greater impact could be expected by radiologists not specialized in breast imaging. Of note is that our findings are at variance with a recent study evaluating the impact of CAD on breast cancer detection in five radiologists not specialized in breast imaging which demonstrated a significant improvement in breast cancer detection when using CAD [20].

The CAD-system algorithm used in our study was not designed for the classification of microcalcifications as benign or malignant lesions; however, our results show a statistically significant difference in the CAD-system rate of detection of malignant and benign microcalcifications. This corroborates the hypothesis that there may be specific morphologic difference in, at least some, malignant and benign microcalcifications and that there exists the possibility of differentiating benign from malignant microcalcifications with CAD. Different CAD-systems have been designed to differentiate benign from malignant microcalcifications and/or benign from malignant masses [21–24]. Recently, comparative observer studies have demonstrated that CAD-systems can be used to improve radiologists' classification of microcalcifications and/or masses and thereby possibly increase the positive biopsy yield [25,26].

Use of computer-aided diagnostic modalities has exciting and important implications for breast cancer diagnosis. The possibility that a CAD-system could detect breast cancer earlier has been demonstrated [18]. In this study, 77% of missed cancers were flagged by the CAD-system [18]. In a recent study, 60% of subtle masses, stellate lesions or architectural distortions (representing missed breast cancers in a screening program) were detected by a CAD-system [19]. Earlier detection may highlight breast cancers when they are more amenable to treatment. This is particularly important, not only for patients, but also because delayed diagnosis of breast cancer accounts for the largest number of malpractice awards in the United Stated today [27,28].

As breast cancer screening programmes spread worldwide, the need for expert mammographers may exceed the availability [29]. Furthermore, as larger numbers of mammograms are interpreted, there is a potential need to screen normal mammograms so as not to burden the limited resource of human mammographic expertise. Our results show that 11 out of 24 (42%) normal mammograms were correctly identified by the CAD-system as having no area of abnormality when evaluated for microcalcifications. Only one out of the 82 (1.2%) abnormal mammograms was not flagged by the CAD-system (false- and/or true-positive marks). However, the small number of normal mammograms in our study does not reflect the usual clinical situation where the vast majority of the mammograms are normal. Further large-scale studies are needed to determine whether those mammograms, which have no abnormal areas flagged with CAD-system require no further evaluation by a radiologist.

Computer-aided diagnosis is an exciting and emerging technology for improving the mammographic detection of breast cancer. Although this study does not demonstrate a statistically significant improvement in detection of microcalcifications when using CAD, it was designed to maximally challenge the CAD system by using dedicated mammographers as the readers. Nevertheless, there is evidence that the potential exists not only to detect some of the 10 to 15% of breast cancers currently not identified with mammography, but also potentially to assist in earlier diagnosis. Additionally, further development may allow for improved differentiation of benign lesions from malignant lesions based on analysis of lesion morphology [18,19].

Acknowledgements. We would like to thank Stacey J. Ackerman, MSE, PhD, Erin M. Sullivan, MPH and Earl P. Steinberg MD, MPH from Covance Health Economics and Outcomes Services Inc. for their assistance with the design conduct and analysis of this study; Colleen Cooke, MD, Judy Destouet, MD, Olga Gatewood, MD, H. Rosy Singh, MD and Cynthia Swann, MD, for assistance with mammographic interpretation; Kathy O'Shaughnessy, PhD (R2 Technology, Inc.) for manuscript review,

Pamela Talalay, PhD for assistance with manuscript preparation and Steven Goodman, MD, PhD, for assistance with statistical analysis. This study was supported by a grant from R2 Technology, Inc., Los Altos, CA, U.S.A.

REFERENCES

1 Landis SH, Murray T, Bolden S, Wingo PA. Cancer statistics, 1999. *CA Cancer J Clin* 1999;49:8–31.
2 Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. *Radiology* 1992;184:613–617.
3 Harvey JA, Farjado LL, Innis CA. Previous mammograms in patients with impapable breast carcinoma: retrospective vs blinded interpretation. *Am J Roentgenol* 1993;161:1167–1172.
4 Goergen SK, Evans J, Cohen GP, MacMillan JH. Characteristics of breast carcinomas missed by screening radiologists. *Radiology* 1997;204:131–135.
5 Lopez MJ, Smart CR. Twenty-year follow-up of minimal breast cancer from the Breast Cancer Detection Demonstration Project. *Surg Oncol Clin N Am* 1997;6:393–401.
6 Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by US radiologists. *Arch Intern Med* 1996; 156:209–213.
7 Thurjell EL, Lernevall KA, Taube AAS. Benefit of independent double reading in a population-based mammography screening program. *Radiology* 1994;191:241–244.
8 Anderson ED, Muir BB, Walsh JS, Kirkpatrick AE. The efficacy of double reading mammograms in breast screening. *Clin Radiol* 1994;49:248–251.
9 Warren RM, Duffy SW. Comparison of single reading with double reading of mammograms, and change in effectiveness with experience. *Br J Radiol* 1995;68:958–962.
10 Huynh PT, Jarolimek AM, Daye S. The false-negative mammogram. *Radiographics* 1998;18:1137–1154.
11 Giger ML, Doi K, MacMahon H, *et al*. An 'intelligent' workstation for computer-aided diagnosis. *Radiographics* 1993;13:647–656.
12 Vyborny CJ, Giger ML. Computer vision and artifical intelligence in mammography. *Am J Roentgenol* 1994;162:699–708.
13 Schmidt RA, Wolverton DE, Vyborny CJ. Computer-aided diagnosis in mammography. RSNA Categorical Course in Breast Imaging 1995; pp. 199–208.
14 Feig SA, Yaffe MJ. Digital mammography, computer-aided diagnosis and telemammography. *Radiol Clin North Am* 1995;33:1205–1230.
15 Chan HP, Lo SC, Sahiner B, Lam KL, Helvie MA. Computer-aided detection of mammographic microcalcifications: pattern recognition with an artifical neural network. *Med Phys* 1995;22:1555–1567.
16 Chan HP, Doi K, Vyborny CJ, *et al*. Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis. *Invest Radiol* 1990;25:1102–1110.
17 Vyborny CJ. Can computers help radiologists read mammograms? *Radiology* 1994;191:315–317.
18 Warren Burhenne LJ, Wood SA, D'Orsi CJ *et al*. *Radiology* 2000;215:554–562.
19 te Brake GM, Karssemeijer N, Hendriks JHCL. Automated detection of breast carcinomas not detected in a screening program. *Radiology* 1998;207:465–471.
20 Nawano S, Murakami K, Moriyama N, Kobatake H, Takeo H, Shimura K. Computer-aided diagnosis in full digital mammography. *Invest Radiol* 1999;34:310–316.
21 Chan HP, Wei D, Helvie MA, *et al*. Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in tecture feature space. *Phys Med Biol* 1995;40:857–876.
22 Baker JA, Kornguth PJ, Lo JY, Floyd CE. Artifical neural network: improving the quality of breast biopsy recommendations. *Radiology* 1996;198:131–135.
23 Jiang Y, Nishikawa RM, Wolverton DE, *et al*. Malignant and benign clustered microcalcifications: automated feature analysis and classification. *Radiology* 1996;198:671–678.
24 Chan HP, Sahiner B, Lam KL, *et al*. Computerized analysis of mammographic calcifications morphological and texture feature spaces. *Med Phys* 1998;25:2007–2019.
25 Jiang Y, Nishikawa RM, Schmidt RA, Metz CE, Giger ML, Doi K. Improving breast cancer diagnosis with computer-aided diagnosis. *Acad Radiol* 1999;6:22–33.
26 Chan HP, Sahiner B, Helvie MA, *et al*. Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: a ROC study. *Radiology* 1999;212:817–827.
27 Jones TL The worst list. Breast cancer now leading source of medical malpractice claims. *Tex Med* 1996;92:34–36.
28 The 1995 Report of the Physician Insures Association of America. Breast Cancer Study. PIAA 1130 Connecticut Avenue, N.W. Suite 800, Washington, DC, June 1995.
29 Hall FM. Screening mammography: potential problems on the horizon. *N Engl J Med* 1986;314:53–55.