Published in final edited form as:

Acad Radiol. 2007 March ; 14(3): 363–370. doi:10.1016/j.acra.2006.12.012.

Independent Evaluation of Computer Classification of Malignant and Benign Calcifications in Full-Field Digital Mammograms

Rich S. Rana, M.D., Yulei Jiang, Ph.D., Robert A. Schmidt, M.D., Robert M. Nishikawa, Ph.D., and Bei Liu, Ph.D.

Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, Chicago, IL 60637

Abstract

Rationale and Objectives—To evaluate whether a computer-aided diagnosis (CADx) technique can accurately classify breast calcifications in full-field digital mammograms (FFDMs) as malignant or benign. The computer technique was developed previously on screen-film mammograms (SFMs) in which individual calcifications were identified manually. The present study evaluated the computer technique independently on a new database of FFDM images with automatic detection of the individual calcifications.

Materials and Methods—We analyzed 49 consecutive FFDM cases (19 cancers) that showed suspicious calcifications. Four mammography radiologists read softcopy mammograms retrospectively and electronically indicated the region of calcifications in each image. The computer then automatically detected the individual calcifications within the indicated region and analyzed eight features of calcification morphology and distribution to arrive at an estimated likelihood of malignancy. The radiologists entered BI-RADS assessments before and after seeing the computer results. Performance was analyzed using receiver operating characteristic (ROC) analysis.

Results—Despite variability in radiologist-indicated regions of calcifications, the computer achieved consistently high performance taking input from the four radiologists (ROC curve area, A_z , of 0.80, 0.80, 0.78, and 0.77; differences not statistically significant). Previous results showed that the computer technique achieved an A_z value of 0.80 on SFMs, which improved radiologists' performance significantly.

Conclusion—The computer technique appears to maintain consistently high performance classifying calcifications in FFDMs as malignant or benign without requiring substantial modification from its initial development on SFMs. The computer performance appears to be robust with respect to variations in radiologists' input.

Corresponding Author and Reprint Request: Yulei Jiang, Ph.D., Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, MC2026, Chicago, IL 60637, Phone: (773) 834-3467, Fax: (773) 702-0371, y-jiang@uchicago.edu.

Author Current Addresses:

Rich S. Rana, M.D., 421 West Huron #909, Chicago, IL 60610

Bei Liu, Ph.D., Department of Radiation Oncology, University of Maryland Medical Center, 22 South Greene Street, Baltimore, MD 21201

Robert Nishikawa and Robert Schmidt are shareholders of, and receive royalty from, Hologic, Inc. (Bedford, MA). Yulei Jiang and Robert Nishikawa receive research funding from Hologic, Inc. It is the University of Chicago Conflict of Interest Policy that investigators disclose publicly actual or potential significant financial interests that may appear to be affected by the research activities.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

computer-aided diagnosis; full-field digital mammography; diagnostic mammography; breast calcifications; breast cancer diagnosis

INTRODUCTION

Breast cancer is the most commonly diagnosed non-skin cancer among women in the United States, with approximately 274,900 new cases (including carcinoma *in situ*) expected in 2006. It is the second leading cause of female cancer deaths, killing more than 40,970 women every year (1). Early detection is important for reducing mortality from breast cancer and, currently, screening mammography is the most effective method for detecting breast cancer early. Recently, full-field digital mammography (FFDM) and computer-aided detection (CAdE) have been adopted clinically as a possible improvement to the conventional screen-film mammography (SFM). FFDM offers potentially better breast cancer detection (2-4) and it better facilitates the application of CAdE by eliminating the film-digitization process necessary for SFMs to be analyzed by a computer (5, 6).

Whereas the goal of CAdE is to help radiologists detect suspect lesions in a mammogram, the goal of computer-aided diagnosis (CAdx) is to help radiologists determine whether a breast lesion is malignant, when the results of computer analysis of the mammogram are provided to the radiologist as a second opinion (7). Previous research indicates that CAdx can potentially help radiologists increase the number of biopsies performed on mildly suspicious but histologically malignant lesions and reduce the number of biopsies performed on mildly suspicious but histologically benign lesions, thereby improving, simultaneously, both sensitivity and specificity of diagnostic mammography (8, 9). Research (8-19) suggests that CAdE and CAdx, together, could serve as an alternative to double reading by radiologists for improving the diagnostic accuracy and consistency of radiologists' mammogram interpretation (20).

Previously, Salfity *et al.* described a computer technique for detecting calcifications in mammograms automatically based on *a priori* information provided by a radiologist (21). The output of that computer technique, representing the locations of individual calcifications, was used as input to another computer technique that classifies calcifications in mammograms as malignant or benign (8, 9, 15). The purpose of the present work was to evaluate the performance of both of these computer techniques, developed previously on digitized SFMs, for classifying calcifications in FFDMs as malignant or benign. In addition, the present work also evaluated the effect of variability in the input that radiologists provide to the computer on the computer performance.

This work improved our prior work in two aspects. First, we evaluated the computer techniques on FFDM images without optimizing the computer techniques for FFDMs, thus making the present evaluation a robustness test of the computer techniques. Second, we improved the computer detection technique by reducing the input required of the radiologist, eliminating the need for the radiologist to estimate the number of individual calcifications present and requiring only that the radiologist to draw a box around the calcifications. Thus, the present evaluation was more difficult for the computer techniques compared to our previous work (8, 15) and it simulated more closely clinical application of the computer techniques. Note, however, this preliminary work was not designed to determine radiologists' performance with and without the computer aid, even though some initial results are presented here. Limitations in this study for measuring radiologists' diagnostic performance are discussed below.

MATERIALS AND METHODS

Study Cases

The database of mammograms used in this study consisted of 81 FFDM images of 49 consecutive patients (see below for inclusion criteria) who underwent diagnostic FFDM studies for evaluation of suspect calcifications at the University of Chicago Hospitals from May 2002 to May 2003, inclusively. These cases included 19 cancers and 30 benign lesions (biopsy was performed on one calcification cluster in each patient even though this was not a criterion for case inclusion). Among the cancers, 13 were ductal carcinoma *in situ* (DCIS) and six were invasive ductal carcinoma (one invasive carcinoma was also a tubular carcinoma). The criteria for case inclusion were that the diagnostic mammography study: (1) was performed using FFDM, (2) depicted a calcification cluster, which subsequently underwent biopsy and for which a definitive pathology report was available, and (3) did not reveal a breast mass associated with the biopsied calcifications. In each case, the original interpreting radiologist reported a final assessment, according to the Breast Imaging Report and Data System (BI-RADS), of either 4 (“suspicious”) or 5 (“highly suggestive of malignancy”). Cases of BI-RADS 1 (“negative”), 2 (“benign”), or 3 (“probably benign”) were excluded from this study because, at the time of the study, the malignant-benign truth status could not be determined definitively. In identifying the study cases, we performed a computer search of all radiology reports during the study time period matching for (a) the word “calcification” and (b) a BI-RADS assessment of 4 or 5. We then reviewed each candidate case in detail according to the inclusion criteria. The study was conducted with approval from the University of Chicago Institutional Review Board.

Each case had two, sometimes one, standard view images available depicting the calcifications of interest (Table I). Whenever possible, we used mediolateral (ML) and craniocaudal (CC) views, although lateromedial (LM), mediolateral oblique (MLO), and laterally exaggerated craniocaudal (XCCL) views were also used when necessary. An experienced researcher of computer-aided diagnosis for mammography selected the images, and an expert mammography specialist reviewed a subset of the cases that were ambiguous as to the most appropriate images. All images were acquired on a General Electric (GE) Senographe 2000D FFDM machine, and subsequently stored and transmitted via a Picture Archiving and Communication System (PACS). The GE processed images were used for computer image analysis and in the observer study because the appearance of the processed images is similar to SFM (22). Magnification views were not included in this study because the computer technique developed on digitized SFM that we evaluated in this study was not designed to analyze magnification views (for SFM, the magnification factor was typically unknown, making it difficult to extract image features). Future studies will include magnification mammograms after we improve the computer technique to analyze these images.

Observer Study

The primary purpose of the observer study was for the observers to provide input to the computer technique, by indicating the general locations of suspect calcifications, so that the computer performance could be evaluated in classifying the calcifications as malignant or benign. Because the observers must interpret the images for this purpose, data of their diagnostic assessment of the suspect calcifications were also recorded secondarily in this study.

Four radiologists, three attending radiologists at the University of Chicago who were experienced mammography specialists and a radiology fellow in breast imaging with about three months of training, participated as observers in this study. All of the observers had

participated in observer studies before the present study. They were told that approximately 40% of the cases were malignant and that a biopsy was performed in all cases. In addition, the observers were told that the computer technique achieved a sensitivity value of 80–90% and a positive predictive value of 55–60% in a previous experiment using SFMs. They were also informed that the computer would provide an independent calculation result for each mammographic view, and that they should concentrate on the view in which the computer result was more suggestive of cancer when considering the computer output.

A Sun Solaris computer displayed the digital mammograms on a Megascan MD5-4820P 5-megapixel grayscale monitor (2048 columns $\times$ 2560 rows) with a DOME Md5/SBX video board. The digital mammograms, displayed at a size of 1448 columns $\times$ 2560 rows, were positioned so that the calcifications in question were approximately centered on the display (an electronic magnifying glass was centered automatically on the calcifications so that the lesion in question was not ambiguous to the observers). For each mammographic view, the observer indicated a region of interest (ROI) containing the calcifications in question by drawing a rectangular box around the calcifications using a computer mouse. After indicating an ROI in each image, the observer indicated his or her BI-RADS assessment based on a consideration of all available mammograms. Only the assessments of 2, 3, 4, or 5 were available to the radiologist because a lesion was present in all cases and because we asked the observers to make a diagnostic decision based on the available images.

The computer analysis of the mammograms, first detecting the individual calcifications and then classifying the calcifications as malignant or benign, based on the radiologist-indicated ROIs, started immediately after the observer entered a BI-RADS assessment. The computer classification results were presented to the observer in the form of a percentage representing an estimate of the likelihood of malignancy (8). A separate result calculated independently by the computer for each image was presented (two calculations for each case that consisted of two views). The observer then had an opportunity to read each image again while considering the computer-estimated likelihood of malignancy and, if the observer felt necessary, to modify one or both ROIs the observer indicated and have the computer repeat its calculations. Subsequent to this repeat interpretation, the observer entered a second BI-RADS assessment that potentially differed from his or her first assessment, reflecting his or her consideration of the results of the computer analysis. The observers were blinded to the pathology results of the study cases. The computer displayed each patient case successively in random order. No time limit was imposed and no other information was made available to the observers.

Computer Detection of Individual Calcifications

The computer calcification-detection technique that we developed previously (21) detected individual calcifications automatically within an ROI indicated by a radiologist. The technique required the radiologist also to estimate the number of calcifications within the ROI in one of four categories: (A) 3–5 calcifications, (B) 6–10 calcifications, (C) 11–30 calcifications, or (D) 31 or more calcifications (21). This estimate of the number of calcifications within the ROI was necessary for the computer to set threshold values for identifying calcifications from background noise: a larger estimate of the number of calcifications corresponds to a lower threshold value, which yields a greater number of “signals” as candidates for calcifications. In this study, we enhanced the computer calcification-detection technique to determine the number of calcifications within an ROI automatically without an estimate from the radiologist. This was accomplished by carrying out the previous calcification-detection operation four times, each time assuming the number of calcifications to be within a different category (A, B, C, or D), as if a radiologist had estimated the number of calcifications to be in that category. The results of the number of calcifications detected following each presumption of the number of calcifications present

were then analyzed in comparison to the boundaries of the four categories. Note, e.g., if 25 calcifications were present, an incorrect presumption of the number of calcifications in category A would likely yield a computer-detection result of close to 5 calcifications (upper bound of category A), an incorrect presumption of the number of calcifications in category B would likely yield a computer-detection result of close to 10 calcifications (upper bound of category B), and a correct presumption of the number of calcification in category C would likely yield a computer-detection result of about 25 calcifications, whereas an incorrect presumption of the number of calcifications in category D would likely yield a computer-detection result of close to 31 calcifications (lower bound of category D). The computer technique used this principle to determine the likely number of calcifications within the ROI by attempting to identify a category such that the computer detection results approached the upper bounds in the categories below, and approached the lower bounds in the categories above, the category that best matches the true number of calcifications (Table II). If such a category could not be identified, then the computer asked the radiologist to estimate the number of calcifications (i.e., in category A, B, C, or D), which occurred in 93 of 332 ROIs (28%). The calcification-detection operations, which were run at least four times per image, were carried out by a second computer, which relayed the results back to the observer-study workstation, reducing the observer's wait time to a few seconds.

Computer Classification of Calcifications as Malignant or Benign

The technique for computer classification of calcifications as malignant or benign has been described previously (8, 15). This technique, developed on SFM images, was used in this study on FFDM images without modification. Briefly, the classification technique used as input the locations of individual calcifications already detected by the computer and, from these locations, extracted eight features of the calcifications. The features (15) were: (1) calcification cluster circularity, (2) calcification cluster area, (3) number of calcifications within the cluster, (4) average effective volume of individual calcifications (defined as area times effective thickness estimated from contrast), (5) relative standard deviation in calcification effective thickness, (6) relative standard deviation in calcification effective volume, (7) average area of individual calcifications, and (8) second highest calcification-shape-irregularity measure in the cluster. These features were effective for classifying calcifications as malignant or benign presumably because they correlate qualitatively with radiologists' perceptual experience (15). An artificial neural network (ANN) then merged the eight features into an estimate of the likelihood of malignancy (15). It is important to note that for the work described here, the exact same features developed on SFMs and the ANN trained previously on SFMs were used in this study to analyze new FFDM images without modification or re-training.

Data Analysis

The performance of the computer in classifying calcifications as malignant or benign was measured with receiver operating characteristic (ROC) curves computed using the LABROC4 algorithm developed by Charles Metz (23). Four ROC curves were computed for the computer performance, each one based on the input (ROIs identifying the calcifications) from a different observer. When the computer produced two different estimates of the likelihood of malignancy in a case, based on independent analyses of two images, the one estimate more suggestive of malignancy was used in the ROC analysis (8). The performance of each radiologist in classifying calcifications as malignant or benign, with and without the computer aid, was also measured with separate ROC curves, based on the BI-RADS assessments reported by the observer. Comparison between ROC curves in terms of the area under the ROC curve, A_z , was made using the Dorfman-Berbaum-Metz (DBM) method (24).

Because four radiologists indicated ROIs in each image to the computer, we measured the variability in the ROIs indicated by different radiologists. For two ROIs in a given image, we defined their similarity as the ratio of their area of overlap to the average of their areas. For multiple ROIs in a given image, we defined the similarity of the ROIs as the average similarity of all possible pairs of the ROIs. Therefore, for four ROIs in a given image, the similarity of the ROIs was the average of six pairwise-ROI similarities.

RESULTS

The computer performed accurately in classifying calcifications as malignant or benign (Fig. 2). The average A_z value of the computer taking input from each of the four observers was 0.80 (SE¹ 0.05, the DBM method). Without the computer aid, the four radiologists achieved an average A_z value of 0.71 (SE 0.08). The difference between the performance of the computer and the radiologists was not statistically significant (95% confidence interval: -0.12, 0.29). In a previous study on SFMs of 104 patients (8), we found that the computer achieved an A_z value of 0.80 (SE 0.04). Therefore, the computer performance on FFDM images was similar to its performance on SFM images.

Figure 2 also shows that the computer performance was consistent when taking input from different observers. However, the observers' inputs to the computer were less consistent. For the 81 mammograms studied in this work and four ROIs indicated in each image by the four observers, the average similarity of the ROIs was 0.66 (SD² 0.14), which implies that, on average, the ROIs had only 66% of their areas in common. This result suggests that the computer performance was robust with respect to variability in the region of calcifications indicated by different radiologists.

Figure 3 shows composite ROC curves of the computer performance, the performance of the radiologists with the computer aid, and the performance of the unaided radiologists. Of these three ROC curves, the computer performance had the highest A_z value (0.80, SE 0.05), the performance of the radiologists with the computer aid had an intermediate A_z value (0.72, SE 0.08), and the radiologists' unaided performance had the lowest A_z value (0.71, SE 0.08)³. The differences in the A_z values were not statistically significant. The ranking of the ROC curves follow a similar trend observed previously in a study of SFMs (8).

DISCUSSION

The results of this study suggest that the performance of our computer technique classifying calcifications in mammograms as malignant or benign is comparable for digitized SFM images (8) and for GE FFDM images (A_z of 0.80 vs. 0.80, respectively, for SFMs and FFDMs). One possible reason for the similar computer performance on FFDMs and on SFMs is that, whereas in the GE FFDM unprocessed (or raw) image pixel values relate linearly to exposure, in the processed image that radiologists use for interpretation the relationship between pixel value and exposure has undergone a logarithmic transformation to make the FFDM "appear like" SFM (22). In addition, the pixel size of the GE FFDM is the same as the nominal pixel size (100 microns) of the digitized SFM we used previously (8). Therefore, as long as calcifications are not situated near the skin line where an unknown proprietary image processing takes place in the conversion from the raw FFDM image to the processed FFDM image, the calcifications depicted in the processed FFDM images may

¹SE = standard error

²SD = standard deviation

³These A_z values, calculated using the DBM method, are different slightly from those (0.79, 0.74, and 0.70, respectively) of the ROC curves shown in Fig. 3 obtained by averaging the binormal parameters of the ROC curves of the individual readers (25).

appear similar to those depicted in the digitized SFM images, even though the physical characteristics of the FFDM and SFM differ. This similarity between the FFDM and the digitized SFM is important because it suggests that our computer techniques that detect and classify calcifications as malignant or benign developed on digitized SFM images should be applicable to GE FFDM images without substantial modifications. However, others have found that re-optimizations are necessary when converting computer image-analysis techniques from SFM images to FFDM images (26, 27). More research is needed to investigate the migration of CADe and CADx from SFM to FFDM.

The results of this study are encouraging in that this study, an independent evaluation, suggests robust performance of the computer in classifying calcifications as malignant or benign. The computer technique, developed previously on digitized SFM images, was evaluated in this study on a set of new patient cases and on FFDM images, using the exact same image features developed previously and the same ANN trained previously on SFM images. The computer performance found in this study was similar to that found previously (8). These results are, therefore, encouraging.

The results of this study also suggest that the computer performance in classifying calcifications as malignant or benign is robust with respect to variability in radiologists' input to the computer, in the form of an ROI identifying calcifications for the computer to analyze. Although the ROIs in a particular image indicated by different radiologists were not highly consistent in our study, the computer performance based on these ROIs was remarkably consistent. This consistency in the computer performance is critical for the computer to be used clinically by radiologists, as it must be able to function consistently when used by different radiologists. One reason for this consistency is that the computer detects calcifications automatically within the ROI indicated by a radiologist. As long as different radiologists indicate ROIs that enclose the same calcifications, there is a good chance that the computer can detect many, if not all, of the same calcifications. Therefore, the computer's calculation of the likelihood of malignancy can be similar even when the radiologists indicate to the computer ROIs that appear to be different (28). However, the computer obviously cannot produce the same calculation results if the ROIs indicated by different radiologists differ so substantially as to correspond to different subgroups of calcifications. A second reason for the computer's consistency is that removing the requirement for radiologists to estimate the number of calcifications may have helped the computer calculations to be more consistent (in addition to making the computer aid easier to use for the radiologist), because it eliminated an additional source of variability in the observer input in estimating the number of calcifications.

While the results of this study are encouraging, these results are preliminary and the study has several limitations. First, the number of cases, all that was available clinically at the time of the study, is too small to yield sufficient statistical power for this study. Second, the radiologist observers could have recognized a small number of the patient cases, because the radiologists practice clinically at the University of Chicago, where the study cases originated. Third, magnification views, generally considered diagnostically important to radiologists especially for calcifications, were not included in this study, for reason of a technical limitation that the computer technique was not designed to analyze magnification views. Finally and importantly, the observers received little or no training on the use of the computer aid prior to the observer study. We believe that as with most tools, training on the use of the computer aid is important, because training could increase a radiologist's understanding and trust of the computer aid, and help the radiologist derive maximum benefit from the computer aid. For these reasons, this study was not designed for measuring radiologists' diagnostic performance with or without the computer aid. Rather, this study was our first experience with CADx on FFDMs, and it was designed to evaluate the

computer performance alone when taking input from different radiologists. We plan, in the future, to expand this study with a greater number of cases, now available clinically, and a greater number of observers, to test whether the computer aid can improve radiologists' performance interpreting calcifications in FFDMs, after enhancing the study design to include magnification mammograms and after providing radiologists with sufficient training. The future study may also combine computer analysis of both breast masses and calcifications.

It is encouraging that our computer technique achieved high, consistent, performance on FFDMs as on SFMs, in classifying calcifications as malignant or benign, without modification or re-training from its original development on SFMs. It is also encouraging that the computer performance appears to tolerate variability in the input provided by different radiologists when they use the computer aid. It is further encouraging that the reduced input from the radiologist, which better facilitates clinical use of the computer aid by radiologists, did not compromise the computer performance. However, this study showed only a small increase in radiologists' performance from use of the computer aid that was not statistically significant. Future studies, with greater statistical power and after addressing limitations of the present study, will need to test whether this computer aid can help radiologists improve diagnostic performance.

Acknowledgments

We wish to thank Drs. Hiroyuki Abe, James Chambliss, and Charlene Sennett for participating in this study as observers.

This work was supported in part by the National Institutes of Health (through grants R01 CA92361, T35 DK062719-16, and P30 CA014599-32) and the US Army Medical Research and Materiel Command (through grant DAMD 17-00-1-0197).

References

1. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2006. *CA Cancer J Clin.* 2006; 56:106–130. [PubMed: 16514137]
2. Lewin JM, D'Orsi CJ, Hendrick RE, et al. Clinical comparison of full-field digital mammography and screen-film mammography for detection of breast cancer. *AJR Am J Roentgenol.* 2002; 179:671–677. [PubMed: 12185042]
3. Skaane P, Skjennald A. Screen-film mammography versus full-field digital mammography with soft-copy reading: randomized trial in a population-based screening program--the Oslo II study. *Radiology.* 2004; 232:197–204. [PubMed: 15155893]
4. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med.* 2005; 353:1773–1783. [PubMed: 16169887]
5. Pisano ED, Kuzmiak C, Koomen M. Perspective on digital mammography. *Semin Roentgenol.* 2001; 36:195–200. [PubMed: 11475066]
6. Nishikawa, RM. Computer-aided detection in digital mammography. In: Pisano, ED.; Yaffe, MJ.; Kuzmiak, CM., editors. *Digital Mammography*. Philadelphia: Lippincott Williams and Wilkins; 2003. p. 43-48.
7. Doi K, MacMahon H, Katsuragawa S, Nishikawa RM, Jiang Y. Computer-aided diagnosis in radiology: potential and pitfalls. *Eur J Radiol.* 1999; 31:97–109. [PubMed: 10565509]
8. 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. [PubMed: 9891149]
9. Jiang Y, Nishikawa RM, Schmidt RA, Toledano AY, Doi K. Potential of computer-aided diagnosis to reduce variability in radiologists' interpretations of mammograms depicting microcalcifications. *Radiology.* 2001; 220:787–794. [PubMed: 11526283]

10. 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. [PubMed: 2079409]
11. Kegelmeyer WP Jr, Pruneda JM, Bourland PD, Hillis A, Riggs MW, Nipper ML. Computer-aided mammographic screening for spiculated lesions. *Radiology.* 1994; 191:331–337. [PubMed: 8153302]
12. 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. [PubMed: 11526282]
13. Gur D, Sumkin JH, Rockette HE, et al. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. *J Natl Cancer Inst.* 2004; 96:185–190. [PubMed: 14759985]
14. Getty DJ, Pickett RM, D'Orsi CJ, Swets JA. Enhanced interpretation of diagnostic images. *Invest Radiol.* 1988; 23:240–252. [PubMed: 3372189]
15. Jiang Y, Nishikawa RM, Wolverton DE, et al. Malignant and benign clustered microcalcifications: automated feature analysis and classification. *Radiology.* 1996; 198:671–678. [PubMed: 8628853]
16. Baker JA, Kornguth PJ, Lo JY, Floyd CE Jr. Artificial neural network: improving the quality of breast biopsy recommendations. *Radiology.* 1996; 198:131–135. [PubMed: 8539365]
17. Chan HP, Sahiner B, Helvie MA, et al. Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: an ROC study. *Radiology.* 1999; 212:817–827. [PubMed: 10478252]
18. Veldkamp WJ, Karssemeijer N, Otten JD, Hendriks JH. Automated classification of clustered microcalcifications into malignant and benign types. *Med Phys.* 2000; 27:2600–2608. [PubMed: 11128313]
19. Huo Z, Giger ML, Vyborny CJ, Metz CE. Breast cancer: effectiveness of computer-aided diagnosis observer study with independent database of mammograms. *Radiology.* 2002; 224:560–568. [PubMed: 12147857]
20. Jiang Y, Nishikawa RM, Schmidt RA, Metz CE. Comparison of independent double readings and computer-aided diagnosis (CAD) for the diagnosis of breast calcifications. *Acad Radiol.* 2006; 13:84–94. Erratum in *Acad Radiol* 2006; 13: 534–535. [PubMed: 16399036]
21. Salfity MF, Nishikawa RM, Jiang Y, Papaioannou J. The use of a priori information in the detection of mammographic microcalcifications to improve their classification. *Med Phys.* 2003; 30:823–831. [PubMed: 12772990]
22. Burgess A. On the noise variance of a digital mammography system. *Medical Physics.* 2004; 31:1987–1995. [PubMed: 15305451]
23. Metz CE, Herman BA, Shen JH. Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Stat Med.* 1998; 17:1033–1053. [PubMed: 9612889]
24. Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method. *Invest Radiol.* 1992; 27:723–731. [PubMed: 1399456]
25. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. *Invest Radiol.* 1989; 24:234–245. [PubMed: 2753640]
26. Li, H.; Giger, ML.; Yuan, Y., et al. Comparison of computerized image analyses for digitized screen-film mammograms and full-field digital mammography images. In: Astley, SM.; Brady, M.; Rose, C.; Zwiggelaar, R., editors. *Digital Mammography*. New York: Springer; 2006. p. 569–575.
27. Ge J, Sahiner B, Hadjiiski LM, et al. Computer aided detection of clusters of microcalcifications on full field digital mammograms. *Med Phys.* 2006; 33:2975–2988. [PubMed: 16964876]
28. Jiang Y, Nishikawa RM, Papaioannou J. Dependence of computer classification of clustered microcalcifications on the correct detection of microcalcifications. *Med Phys.* 2001; 28:1949–1957. [PubMed: 11585226]

Figure 1.

ROC curves of the computer technique in classifying calcifications as malignant or benign. Each ROC curve was obtained from results of the computer-estimated likelihood of malignancy based on input from one observer. The observer input was in the form of an ROI indicating the location of the calcifications in the image.

Figure 2.

Comparison of average ROC curves of the computer technique, the observers with the computer aid, and the observers without the computer aid, in classifying calcifications as malignant or benign. The differences in the A_z values were not statistically significant. Radiologists' diagnostic performance was measured secondarily in this study; study limitations are noted in the Discussion.

Table I

Image view angles of the study cases

1 st Image	2 nd Image	Number of Cases
ML/LM	CC	18
MLO	CC	9
ML/LM	LM/ML	5
CC	CC/XCCL	2
MLO/ML/LM	N/A	13
CC	N/A	2
Total		49

Table II

Algorithm for computer to estimate the number of calcifications within an ROI in which the computer makes four presumptions of the number of calcifications and then determines the most likely number range based on detection results under each presumption

Detection Results Under Each Presumption				
Presumed Number of Calcifications				
3–5	close to 5	close to 5	close to 5	any
6–10	close to 10	close to 10	any	close to 6
11–30	close to 30	any	close to 11	close to 11
> 31	any	close to 31	close to 31	close to 31
Most Likely Number Range	> 31	11–30	6–10	3–5

Note.—When the computer detection results does not fit into one of these four patterns, the computer asks the observer to select a number range for the calcifications.