**Radiological technologists'
performance for the detection of
malignant microcalcifications in
digital mammograms without and
with a computer-aided detection
system**

Rie Tanaka
Miho Takamori
Yoshikazu Uchiyama
Junji Shiraishi

Radiological technologists' performance for the detection of malignant microcalcifications in digital mammograms without and with a computer-aided detection system

Rie Tanaka,^{a,*} Miho Takamori,^b Yoshikazu Uchiyama,^c and Junji Shiraishi^c

^aKanazawa University, College of Medical, Pharmaceutical, and Health Sciences, School of Health Sciences, 5-11-80 Kodatsuno, Kanazawa 920-0942, Japan

^bIshikawa Prefectural Hospital, Department of Radiology, 2-1 Kuratsuki-Higashi, Kanazawa 920-8530, Japan

^cKumamoto University, Faculty of Life Sciences, 4-24-1 Kuhonji, Kumamoto 862-0976, Japan

Abstract. The aim of this study was to investigate the diagnostic performance of radiological technologists (RTs) in the detection of malignant microcalcifications and to evaluate how much computer-aided detection (CADE) improved their performances compared with those by expert breast radiologists (BRs). Six board-certified breast RTs and four board-certified BRs participated in a free-response receiver operating characteristic observer study. The dataset consisted of 75 cases (25 malignant, 25 benign, and 25 normal cases) of digital mammograms, selected from the digital database for screening mammography provided by the University of South Florida. Average figure of merit (FOM) of the RTs' performances was statistically analyzed using jack-knife free-response receiver operating characteristic and compared with that of expert BRs. The detection performance of RTs was significantly improved by using CADE; average sensitivity was increased from 46.7% to 56.7%, with a decrease in the average number of false positives per case from 0.19 to 0.13. Detection accuracy of an average FOM was improved from 0.680 to 0.816 ($p = 0.001$) and the difference in FOMs between RTs and radiologists failed to reach statistical significance. RTs' performances for the identification of malignant microcalcifications on digital mammography were sufficiently high and comparable to those of radiologists by using CADE. © 2015 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.2.2.024505]

Keywords: radiological technologist; mammogram; malignant microcalcifications; computer-aided detection; jack-knife free-response receiver operating characteristic analysis.

Paper 14147RR received Nov. 6, 2014; accepted for publication May 4, 2015; published online May 27, 2015.

1 Introduction

Breast cancer is one of the most common cancers worldwide. It has risen to first place among all cancers and was the fifth most common cause of cancer-related deaths in Japanese women in 2012.¹ For the early detection of breast cancer, the effectiveness of screening mammography (MMG) has been demonstrated by many research groups.^{2,3} Although the incidence of breast cancer has continued to increase, some countries have successfully reduced breast cancer mortality.^{4,5} Therefore, screening MMG is regarded as an important public health priority.

On the basis of the results of these studies, breast cancer screening using MMG was endorsed in 2000 for women aged over 50 years and in 2004 for those aged over 40 years in Japan. The Ministry of Health, Labor, and Welfare (MHLW) has recommended breast screening programs aiming at a screening rate of 50%. The number of women being screened has consistently risen since 2004.¹ However, a shortage of radiologists has become a major concern in screening MMG.

In order to resolve this problem, the use of computer-aided detection (CADE) and/or reporting by radiological technologists (RTs) as a second opinion for radiologists' readings has been

investigated. Since many research groups have demonstrated the clinical usefulness of various CADE systems for radiologists' readings of digital mammograms,^{6,7} CADE is accepted as the standard of care in the US and is used in ~75% of screening exams.^{8,9} On the other hand, "reporting by RTs" has been recommended as another approach to aid radiologists' readings in several countries.¹⁰ Reporting by RTs has already been tried and tested, mainly in the UK,^{11,12} and several studies have confirmed the benefits of radiographers' reporting for screening mammograms, including cost-effectiveness benefits.^{13,14} Meanwhile, in Japan, reporting by RTs was recommended in 2012 by the MHLW.¹⁵ It was asserted that utilization of RTs' reporting should be promoted as an aid to radiologists' readings as team approach medicine. However, practical utilization of the reporting by RTs has not been established because of variation of RTs' reading skills and insufficient evaluation of the diagnostic accuracy provided by using reporting by RTs.

In this study, the primary aim was to investigate the diagnostic performance of RTs for the detection of malignant microcalcifications by using a free-response receiver operating characteristic (FROC) observer study. In addition, we also evaluated how much the CADE improved RTs' performances

*Address all correspondence to: Rie Tanaka, E-mail: rie44@mhs.mp.kanazawa-u.ac.jp

compared with those obtained from expert breast radiologists (BRs). To our knowledge, there are no reports on the clinical usefulness of a CAde system for RTs' readings.

2 Materials and Methods

2.1 FROC Dataset

All of the digital mammograms used in this study were sampled from the digital database for screening mammography (DDSM) provided by the University of South Florida.¹⁶ The database was provided to facilitate development of computer algorithms to aid in the diagnosis and the development of teaching or training aids, and the current version involves ~2500 studies. Each case includes craniocaudal (CC) and mediolateral oblique (MLO) views of each breast, along with some associated patient information [age at the time of study, American College of Radiology (ACR) breast density rating, subtlety rating for abnormalities, assessment defined by breast imaging reporting and data system (BIRADS),¹⁷ and ACR keyword description of abnormalities] and "ground truth" information about the locations and types of suspicious regions. The lesion type, assessment, and subtlety were specified by an experienced radiologist. Similarly, the outlines for the suspicious regions were derived from markings made on the film by an experienced radiologist.

The database consists of digitized screen film mammograms, while in current clinical practice, full-field digital mammography (FFDM) has become the standard of care. Therefore, case samples were carefully selected by taking into account the image quality of the mammograms. For instance, the mammograms of patients with extremely high-density breast were excluded from FROC dataset, because few microcalcifications represent any visible signs in digitized screen film mammograms of those patients.

The selection process is as follows. First, the digital mammograms of patients with mass were excluded to focus on an evaluation of observers' performances for the detection of microcalcifications. Second, the following case samples were excluded: extremely high-density breast, extremely fatty breast, postoperative deformed breast, breast with diffusely distributed microcalcifications, and deficit breast, to avoid sample cases extremely easy/difficult to detect microcalcifications. The digital mammograms of patients with malignant/benign microcalcifications scattered over 20 mm in diameter were also excluded from the case samples to ensure location accuracy in the FROC observer study. In addition, mammograms obtained with bad positioning, with insufficient image quality, or with wrong marks were excluded. In the results, 66 cases with microcalcifications (36 malignant and 30 benign) and 136 normal cases remained as candidates for FROC data set. The selection was performed by two board-certified RTs who read mammograms in daily clinical practice. These two individuals did not participate in our observer study as observers.

The number of cases included in the dataset was set to be 75, consisting of 50 digital mammograms with microcalcifications (25 malignant and 25 benign) and 25 normal cases in consideration of the observers' workload. Another five cases (two malignant, one benign, and two normal) were also selected for training. The composition ratio (malignant:benign:normal = 1:1:1) was determined by taking into account the reading time for the observer study.

In order to keep stratified sampling fashion for the observer study,¹⁸ all case samples were selected based on the subtlety

rating for abnormalities provided by the DDSM. The subtlety rating was defined on a scale of 1 to 5, where 1 was "subtle" and 5 was "obvious." The case samples included several challenging cancers that were nevertheless correctly detected by skilled readers. A number of cases in each subtlety rating were determined based on stratified sampling¹⁹ in order to have case samples with a range of difficulties likely to produce a normal distribution in the results. Table 1 shows the number of cases in each subtlety. The normal cases were selected to match the malignant cases in terms of the age distribution.

2.2 CAde System

A computer aid was used in this observer study for the detection of clustered malignant microcalcifications. The computerized scheme for such detection in this study was developed based on multiresolution analysis using a filter bank and the Hessian matrix. The CAde system was trained to differentiate malignant microcalcifications from benign ones and only the former findings were finally output. Details of this CAde system have been reported previously.^{20,21} The detection accuracy of this CAde system to detect malignant microcalcifications for the 75 cases was 100% sensitivity, with an average number of false positives (FPs) of 1.43 per case, which is almost equivalent to that of commercially available CAde systems.^{22,23} By contrast, 10 out of 25 benign microcalcifications were incorrectly detected as malignant ones and counted as FPs. The CAde system was designed to provide marks on digital mammograms based on resulting information, such as the center of the location and the effective radius of the detected microcalcifications.

2.3 Observer Study

Six board-certified RTs and four BRs participated in the FROC observer study for the detection of malignant microcalcifications. We employed FROC analysis because conventional receiver operating characteristic (ROC) analysis has limitations for this kind of evaluation (i.e., only one signal can be used per case and the location of the signal cannot be taken into account in the evaluation).²⁴ In the FROC procedure, the observer determined the locations and confidence levels for possible malignant lesions (i.e., clustered microcalcifications) on any of four views of digital mammograms. In addition, the observer could leave a case without any marks if the observer believed that there was no lesion at all.

Digital mammograms were displayed on a high-resolution liquid crystal display (LCD) for MMG (Nio 5M, BARCO). In order to minimize reading-order effects, the viewing order of the 75 case samples was randomized.^{25,26} The case samples were observed by using a publicly available computer interface (ROC Viewer ver. 11.4.0.3, developed by the Japanese Society of Radiological Technology).²⁷ It allows the observer to

Table 1 Number of cases in each subtlety.

	Total	Subtle		Subtlety		Obvious
		1	2	3	4	5
Malignant	25	1	5	12	5	2
Benign	25	1	2	15	6	1

determine the location and confidence level from 0.0 to 1.0 for the lesion candidate just by clicking on each of four views of mammograms. Figure 1 shows the computer interface of the ROC Viewer and a rating window for the FROC observer study. In this FROC observer study, the rating scale for the confidence level was related to the BIRADS assessment, such that 0.00, 0.33, 0.66, and 1.00 correspond to assessments 2 (benign finding), 3 (probably benign), 4 (suspicious abnormality), and 5 (highly suggestive of malignancy), respectively. Therefore, the observers might detect benign microcalcifications with a low confidence level less than 0.33.

2.4 Instruction of the Observer Study

A training session with five mammograms was conducted for each observer to allow them to familiarize themselves with the reading process. The training cases were not included in the 75 mammograms in the test dataset. Before the training and testing sessions, observers were instructed as follows: There are 5 training cases and 75 test cases. When the observer found microcalcifications, he/she categorized it to a BIRADS assessment as usual, and then translated it to a rating score in his/her mind as mentioned above. Therefore, no BIRADS assessment is directly given by the observer. Each case was read without CAde and then read with CAde. The observer is expected to read and rate (interpret) clustered microcalcifications ranked above “assessment 2” in our sequential test. They were told that the computer’s accuracy was nearly 100% sensitivity for malignant microcalcifications involved in the study cases.

Each observer independently reviewed all mammograms with no limits on the reading time and the distance between the observer and the LCD monitor. Gradation change and zooming were allowed. The observers provided written informed consent prior to study participation.

2.5 Determination of true positive or FP

In this study, malignant cases were considered to be positive, whereas benign and normal cases were considered to be negative. A malignant lesion marked by an observer was considered to be true positive (TP) when the distance from the central

location of the ground truth to the lesion was <20 mm. When a malignant lesion could be identified on two views (CC and MLO), it was considered as a TP if it is identified by at least one mark on either view indicating the correct location. If a lesion was identified correctly on both views, a higher rating was selected. All other marks not regarded as TP were considered to be FP. Namely, a benign lesion marked by an observer was counted as FP. “Ground truth information” included in DDSM was used as ground truth.

2.6 Data Analysis

A FROC curve for an individual observer was calculated by varying a threshold confidence level from 1.0 to 0.0, and then all FROC curves were averaged over all observers without and with CAde. The average sensitivity and FPs per case were estimated by using a threshold confidence level of 0.33, which corresponded to BIRADS category 3 (probably benign). Since the observers might detect benign microcalcifications in this FROC study, those findings with a low confidence level <0.33 were not counted as FPs. In other words, some positive findings were not contributed for the sensitivity when the confidence levels for such findings were <0.33.

Statistical analysis was performed using jack-knife free-response ROC (JAFROC) analysis to detect a statistically significant difference between the RTs’ detection performance without and with CAde (JAFROC ver. 4.2, Chakraborty DP, Pittsburg).^{28–30} Figure of merit (FOM) was calculated as a measure of diagnostic accuracy for the detection of malignant microcalcifications and was statistically analyzed by JAFROC analysis 1 (random readers and random cases) with a two-tailed test ($p = 0.05$). For the comparison of detection performances between RTs and radiologists, the 95% confidence interval (CI) of the average FOM value was used. Because the numbers of observers were different for the two groups of RTs and BRs, the effective numbers of degrees of freedom for calculating 95% CI were obtained by using Welch–Satterthwaite equation.

In addition to the evaluations of FOM, sensitivity, and numbers of TP and FP, the effect of CAde on the image interpretation of each case was also investigated for both RTs and radiologists. The results of RTs were also compared with

Fig. 1 Screen shot of receiver operating characteristic viewer and relationship between breast imaging reporting and data system (BIRADS) assessment and confidence level in our observer study. The confidence levels of 0.0 and 1.0 are equivalent to BIRADS assessments 2 and 5, respectively. The confidence levels between 0.0 and 1.0 linearly correspond to those between assessments 2 and 5, for example, the confidence level of 0.5 is equivalent to assessment 3.5.

those of radiologists in terms of the beneficial and detrimental effects by CADe. The beneficial effect of CADe includes detection of missed cancer and/or improved confidence level for the detected cancer and decrease in the number of FPs and/or decreased confidence level for FPs, whereas the detrimental effect of CADe includes detection errors and/or lowered confidence level for the detected cancer and increase in the number of FPs and/or increased confidence level for FPs.

3 Results

3.1 Detection Performance of RTs Without and With CADe

The detection performances of RTs were significantly improved by the use of CADe. Table 2 shows the sensitivity and the numbers of FPs of RTs for the detection of malignant microcalcifications with and without CADe. For RT without CADe, average sensitivity was 46.7%, with an average number of FPs of 0.19 per case. For RT with CADe, average sensitivity increased to 56.7% and the average number of FPs decreased to 0.13. Average FOM value was 0.680 for RTs without CADe and 0.816 for those with CADe (Table 3). This difference in FOM value was statistically significant ($p < 0.001$) according to the JAFROC analysis.

For radiologists without and with CADe, average sensitivity also increased from 50.0% to 56.0%, and the average number of FPs decreased from 0.30 to 0.24, as shown in Table 4. Average FOM value was also improved significantly by using CADe, from 0.671 to 0.798 ($p = 0.023$) (Table 5).

Figure 2 shows the average FROC curves of RTs and radiologists, with and without CADe, respectively. The results indicated that CADe improved the detection performance of RTs as well as that of radiologists.

For cases both without and with CADe, the average FOM values of four radiologists were included in the 95% CI of the average FOM values of six RTs. In the same way, the average FOM values of six RTs were included in the 95% CI of the average FOM values of four radiologists. In addition, average differences in FOMs for the detection of clustered malignant microcalcifications between RTs and radiologists without and with CADe were 0.0083 and 0.0178 with 95% CIs of $(-0.0583, 0.0748)$ and $(-0.0503, 0.0860)$, respectively.

Table 3 Figure of merit (FOM) of six RTs for the detection of clustered malignant microcalcifications with/without CADe (CI: confidence interval).

	Without CADe	With CADe
RT01	0.682	0.832
RT02	0.724	0.845
RT03	0.636	0.775
RT04	0.688	0.807
RT05	0.679	0.749
RT06	0.668	0.887
Average (SD)	0.680 ± 0.012	0.816 ± 0.020
95% CI	(0.65, 0.71)	(0.76, 0.87)

Figure 3 indicated the relationship between average confidence levels for all malignant microcalcifications obtained by six RTs and four radiologists without and with CADe. As shown in the figure, the correlation between the confidence levels of RTs and radiologists became higher when the CADe output was provided. The correlation coefficient between the confidence levels of RTs and radiologists became increased from 0.76 to 0.90 when the CADe output was provided.

3.2 Beneficial and Detrimental Cases in RTs' Performance with CADe

Computer aids were beneficial to reduce the number of missed malignant microcalcifications. Some RTs could newly identify malignant microcalcifications in 16 out of 25 malignant cases (12/25 in radiologists) by using CADe. Figure 4 shows a sample case with malignant microcalcifications on the pectoralis major muscle (in a 62-year-old woman). Except for one RT, all other RTs missed the finding without CADe because of its unexpected location. Two more RTs could identify it by using CADe, although CADe detected it only on the MLO view. Figure 5 shows a sample case with malignant microcalcifications located

Table 2 Sensitivity and number of false positives (FPs) of six radiological technologists (RTs) for the detection of clustered malignant microcalcifications with/without computer-aided detection (CADe).

	Sensitivity (%)			FP (/case)			Experience years
	Without CADe	With CADe	ΔSensitivity (%)	Without CADe	With CADe	ΔFP (/case)	
RT01	64.0	64.0	0.0	0.33	0.15	-0.18	10
RT02	64.0	64.0	0.0	0.19	0.19	0.00	11
RT03	24.0	44.0	20.0	0.03	0.04	0.01	20
RT04	32.0	36.0	4.0	0.05	0.03	-0.02	3
RT05	40.0	52.0	12.0	0.15	0.16	0.01	6
RT06	56.0	80.0	24.0	0.37	0.20	-0.17	5
Average (SD)	46.7 ± 17.1	56.7 ± 15.9	10.0 ± 10.4	0.19 ± 0.14	0.13 ± 0.07	-0.06 ± 0.09	10

Table 4 Sensitivity and numbers of FPs of four radiologists for the detection of clustered malignant microcalcifications with/without CADe.

	Sensitivity (%)			FP (/case)			Experience years
	Without CADe	With CADe	ΔSensitivity (%)	Without CADe	With CADe	ΔFP (/case)	
BR01	56.0	59.0	3.0	0.39	0.37	-0.02	14
BR02	60.0	68.0	8.0	0.45	0.33	-0.12	4
BR03	44.0	56.0	12.0	0.24	0.09	-0.16	14
BR04	40.0	40.0	0.0	0.13	0.15	-0.02	11
Average (SD)	50.0 ± 9.5	56.0 ± 11.7	5.8 ± 5.3	0.30 ± 0.14	0.24 ± 0.14	-0.07 ± 0.08	11

Table 5 Figure of merit (FOM) of four radiologists for the detection of clustered malignant microcalcifications with/without CADe (CI: confidence interval).

	Without CADe	With CADe
BR01	0.732	0.766
BR02	0.660	0.858
BR03	0.636	0.784
BR04	0.657	0.784
Average (SD)	0.671 ± 0.021	0.798 ± 0.021
95% CI	(0.60, 0.74)	(0.73, 0.86)

in a dense mammary lesion (in a 75-year-old woman). Two out of six RTs missed the finding, but all RTs could identify it by using CADe. Figure 6 is another sample case with subtle malignant microcalcifications because of high breast density (in an 85-year-old woman). Four out of six RTs missed it in the initial findings without CADe, but three out of these four could identify it by using CADe, and one out of two RTs who found it without CADe missed it by using CADe.

CADe was also useful for RTs to improve the interpretation accuracy of malignant microcalcifications. Note that the

Fig. 2 Average free-response receiver operating characteristic curves of radiological technologists (RTs) and breast radiologists (BRs) with and without computer-aided detection (CADe).

interpretation accuracy was defined as increased confidence levels for malignant cases and/or decreased confidence levels for benign cases. In the malignant cases, there was a significant improvement in the average confidence levels of six RTs between with and without CADe ($p < 0.05$). Especially, average confidence levels of RTs for six malignant cases were increased by using CADe more than 0.33, whereas only one case was increased in the interpretations by four radiologists. On the other hand, some RTs changed confidence levels to less than 0.33 (assessment 2 or 1) by referring to CADe outcomes in 10 out of 25 benign cases (10/25 in radiologists), after giving initial confidence levels equal to or higher than 0.33 (assessment 3) without CADe. In the results, the average number of benign microcalcifications detected by the observers was decreased from 5.7 to 5.0 in RTs and from 9.5 to 6.5 in radiologists, respectively.

CADe sometimes lowered the assessment of malignant microcalcifications and/or even led missed malignant microcalcifications. The detrimental effects were found in 4 out of 25 malignant cases in RTs and in 7 out of 25 malignant cases in radiologists. Figure 7 shows a sample case with malignant

Fig. 3 The relationship between average confidence levels for all malignant microcalcifications obtained by six RTs and four radiologists without and with CADe.

Fig. 4 Beneficial case 1 (malignant microcalcifications, a 62-year-old woman). Solid circle indicates CADe outcome and solid squares indicate the local areas of enlarged views. Numbers of observers correctly identifying the finding: one BR and one RT without CADe, and two BRs and three RTs with CADe.

microcalcifications in the right breast (in a 59-year-old woman). Three out of six RTs identified it in the initial findings without CADe, but two out of these three removed the mark and another decreased her confidence level in half by using CADe. Furthermore, CADe provided 18 more FPs in six RTs (average 3.0) and 11 more FPs in four radiologists (average 2.8) by referring to CADe, although the number of FPs/case as shown in previous Sec. 3.1. Figure 8 is a sample case with benign microcalcifications in the right breast (in a 79-year-old woman). The number of FPs was especially increased in such a case where CADe provided multiple incorrect detections.

4 Discussion

Several research groups have reported the clinical effectiveness of various CADs for radiologists' readings of digital mammograms.⁶⁻⁹ In this study, we found that the detection performance of RTs was significantly improved by the use of CADe, regardless of their number of years of experience. The observers could find missed lesions and exclude FP ones by referring to CADe outcomes. The increased sensitivity and decreased number of FPs indicate a reduction both in detection errors and interpretation errors. In addition, CADe significantly improved the confidence levels in the detection of malignant microcalcifications,

Fig. 5 Beneficial case 2 (malignant microcalcifications, a 75-year-old woman). Solid circles indicate CADe outcomes and solid squares indicate the local areas of enlarged views. Numbers of observers correctly identifying the finding: three BRs and four RTs without CADe, and three BRs and six RTs with CADe.

Fig. 6 Beneficial case 3 (malignant microcalcifications, an 85-year-old woman). Solid circles indicate CAdE outcomes and solid squares indicate the local areas of enlarged views. Numbers of observers correctly identifying the finding: zero BR and two RTs without CAdE, and three BRs and four RTs with CAdE.

resulting in increased detection accuracy of malignant microcalcifications by RTs. We also confirmed that the correlation between the confidence levels of RTs and radiologists became higher when the CAdE output was provided. Therefore, it could be assumed that RTs' detection performances become close to those obtained by radiologists by using CAdE. These results indicate that CAdE is useful for RTs as well as for radiologists.

However, we also confirmed several detrimental effects in the RTs' reading with CAdE. The detrimental effect of CAdE was generally observed as detection errors and/or lowered confidence level of malignant microcalcifications (type II error) and an increase in the number of FPs (type I error). In the case shown in Fig. 7, CAdE indicated malignant microcalcifications only on one of the views (CC). Some RTs might be confused by the

CAdE outcomes, resulting in detection errors and/or lowered confidence level of malignant microcalcifications. On the other hand, in this study, the detection performance obtained by RTs was relatively low in a single reading, whereas an equivalent performance was obtained from four radiologists. First, we need to account for this in terms of difficulty of the dataset for the observer study. The database consisted mostly of cases of less than "subtlety 3," that is, subtle findings, and contained six challenging cases, namely, "subtlety 1" or "subtlety 2." These challenging cases were mostly missed and thought to have been a major factor that lowered the detection performance of the original single reading.

The difference in the detection performance between RTs and radiologists failed to reach statistical significance. This is

Fig. 7 Detrimental case 1 (malignant microcalcifications, a 59-year-old woman). Solid circles indicate CAdE outcomes and solid squares indicate the local areas of enlarged views. Numbers of observers incorrectly identifying the finding: one BR and three RTs without CAdE, and one BR and one RT with CAdE.

Fig. 8 Detrimental case 2 (benign microcalcifications, a 79-year-old woman). Solid circles indicate CADe outcomes and solid squares indicate the local areas of enlarged views. Numbers of observers incorrectly identifying the finding: two BRs and two RTs without CADe, and two BRs and five RTs with CADe.

because the observers were asked to find benign and malignant microcalcifications, which supported the finding that RTs were as competent as radiologists in the detection of certain findings.³¹ In fact, Japanese board-certified breast RTs are well trained because it is mandatory for them to take certain courses and to pass written, reading and practical exams, which must be retaken every five years. In addition, the RTs are actually reading mammograms to provide supporting data for radiologists' readings in daily clinical practice.

When conventional FROC study by a normal rating method was used, the detection performance of observers could not be assessed considering whether identifications are benign or malignant. In the FROC observer study, we used a practical rating method, which was directly correlated to the "assessment" value utilized in BIRADS assessment. This rating method enabled us to evaluate the performance of observers in relation to the confidence in the detection of malignancy of microcalcifications. Therefore, we did not have any numerical correction and normalization on confidence levels obtained in the observer study.

However, there were several limitations in the present study. First, the number of FPs in benign cases might have a major effect of reducing FOMs. The FPs in radiologists were mainly benign microcalcifications, while those in RTs were lesions other than benign microcalcifications. The larger number of FPs counted in benign microcalcifications is thought to be one of the factors to lower FOMs in radiologists compared to those in RTs. This might be one limitation of our FROC study which allowed the observers to detect benign microcalcifications even if they believed they were probably benign. Second, for the comparison of detection performances between RTs and radiologists, the 95% CI of the average FOM value was used. The detection performance between RTs and radiologists failed to reach a statistically significant difference. However, to confirm no statistical significant difference, noninferiority analysis would be performed in the future. Although we calculated 95% CIs for the difference in FOMs between RTs and radiologists, we could not demonstrate an equivalency in terms of diagnostic accuracy for both RTs and radiologists,

because we could not have, at present, a certain value of equivalence margin which was required for demonstrating the equivalency by using a noninferiority test.³² Finally, the observers were asked to read digitized screen film mammograms instead of FFDM mammograms, which are the standard of care in the current clinical practice. The limited type of abnormality, malignant microcalcifications, is another issue to consider. In order to confirm an adequate diagnostic performance of RTs in clinical situations, further studies based on noninferiority analysis are required to expand the type of abnormalities, such as mass and distortion, on FFDM mammograms.

5 Conclusions

The diagnostic performance of RTs for the detection of malignant microcalcifications on mammograms was sufficiently high and improved further by using CADe. These results indicate that CADe could be useful for reporting by RTs, as well as for radiologists' readings of malignant microcalcifications on digital MMG.

Acknowledgments

This study was partially supported by the Foundation for the Promotion of Cancer Research.

References

1. Vital Statistics Japan, Tokyo, "The ministry of health, labor and welfare," <http://ganjoho.jp/professional/statistics/index.html> (17 May 2015).
2. N. Ohuchi et al., "Improved detection rate of early breast cancer in mass screening combined with mammography," *Jpn. J. Cancer Res.* **84**, 807–812 (1993).
3. M. Kawai et al., "Effect of screening mammography on cumulative survival of Japanese women aged 40–69 years with breast cancer," *Breast Cancer* **21**(5), 542–549 (2012).
4. A. Jemal et al., "Global patterns of cancer incidence and mortality rates and trends," *Cancer Epidemiol. Biomarkers Prev.* **19**, 1893–1907 (2010).

5. P. Autier et al., "Disparities in breast cancer mortality trends between 30 European countries: retrospective trend analysis of WHO mortality database," *Br. Med. J.* **341**, c3620 (2010).
6. F. J. Gilbert et al., "Single reading with computer-aided detection for screening mammography," *N. Engl. J. Med.* **359**, 1675–1684 (2008).
7. K. Doi et al., "Computer-aided diagnosis in radiology: potential and pitfalls," *Eur. J. Radiol.* **31**, 97–109 (1999).
8. B. Boyer et al., "CAD in questions/answers Review of the literature," *Eur. J. Radiol.* **69**, 24–33 (2009).
9. V. M. Rao et al., "How widely is computer-aided detection used in screening and diagnostic mammography?" *J. Am. Coll. Radiol.* **7**, 802–805 (2010).
10. The College of Radiographers, *Reporting by Radiographers: A Vision Paper*, The College of Radiographers, London (1997).
11. A. M. Paterson et al., "Reporting by radiographers: a policy and practice guide," *Radiography* **10**, 205–212 (2004).
12. The College of Radiographers, *Preliminary Clinical Evaluation and Clinical Reporting by Radiographers: Policy and Practice Guidance*, The College of Radiographers, London (2013).
13. NHS Breast Screening Programme, *Annual Review*, NHS Cancer Screening Programmes, Sheffield (2011).
14. F. J. van den Biggelaar, P. J. Nelemans, and K. Flobbe, "Performance of radiographers in mammogram interpretation: a systematic review," *Breast* **17**, 85–90 (2008).
15. "Report from the ministry of health, labor and welfare," Ministry of Health, Labor and Welfare, Tokyo, <http://www.mhlw.go.jp/english/> (17 May 2015).
16. University of South Florida Digital Database for Screening Mammography (DDSM), <http://marathon.csee.usf.edu/Mammography/Database.html> (17 May 2015).
17. American College of Radiology, *ACR BIRADS: Breast Imaging Reporting and Data System: Breast Imaging Atlas*, 4th ed., ACR BI-RADS: Mammography, Reston, Virginia (2003).
18. A. Chaudhuri, Ed., "Methods of sample selection and estimation with their properties," in *Essentials of Survey Sampling*, PHI Learning, Delhi, India, 2010 <http://phindia.com/index.php>.
19. C. E. Metz, "Some practical issues of experimental design and data analysis in radiological ROC studies," *Invest. Radiol.* **24**(3), 234–245 (1989).
20. R. Nakayama et al., "Computer-aided diagnosis scheme for histological classification of clustered microcalcifications on magnification mammograms," *Med. Phys.* **31**, 789–799 (2004).
21. R. Nakayama et al., "Computer-aided diagnosis scheme using a filter bank for detection of microcalcification clusters in mammograms," *IEEE Trans. Biomed. Eng.* **53**, 273–283 (2006).
22. A. Lauria, "GPCALMA: implementation in Italian hospitals of a computer aided detection system for breast lesions by mammography examination," *Phys. Med.* **25**, 58–72 (2009).
23. A. M. Scarnelo et al., "Sensitivity of a direct computer-aided detection system in full-field digital mammography for detection of microcalcifications not associated with mass or architectural distortion," *Can. Assoc. Radiol. J.* **61**, 162–169 (2010).
24. B. Zheng et al., "A comparison of two data analyses from two observer performance studies using jackknife ROC and JAFROC," *Med. Phys.* **32**, 1031–1034 (2005).
25. D. Gur et al., "Prevalence effect in a laboratory environment," *Radiology* **228**, 10–14 (2003).
26. D. Gur et al., "The 'laboratory' effect: comparing radiologists' performance and variability during prospective clinical and laboratory mammography interpretations," *Radiology* **249**, 47–53 (2008).
27. J. Shiraishi et al., "Basic concepts and development of an all-purpose computer interface for ROC/FROC observer study," *Radiol. Phys. Technol.* **6**, 35–41 (2003).
28. D. P. Chakraborty and K. S. Berbaum, "Observer studies involving detection and localization: modeling, analysis, and validation," *Med. Phys.* **31**, 2313–2330 (2004).
29. D. P. Chakraborty, "Validation and statistical power comparison of methods for analyzing free-response observer performance studies," *Acad. Radiol.* **15**, 1554–1566 (2008).
30. D. P. Chakraborty and L. H. Winter, "Free-response methodology: alternate analysis and a new observer-performance experiment," *Radiology* **174**, 873–881 (1990).
31. G. Wivell et al., "Can radiographers read screening mammograms?," *Clin. Radiol.* **58**(1), 63–67 (2003).
32. E. Walker and A. S. Nowacki, "Understanding equivalence and non-inferiority testing," *J. Gen. Intern. Med.* **26**, 192–196 (2011).

Rie Tanaka is an assistant professor of radiological technology at Kanazawa University in Japan. She has a PhD in health sciences from Kanazawa University. She has been working on quantitative image analysis of dynamic chest radiography and its clinical evaluation. Her research interests also include observer study and image-guided radiation therapy. The research presented here was funded by the Cancer Research Foundation.

Miho Takamori is a radiographer from Ishikawa prefectural hospital. She has been active in the area of imaging, quality assurance, and quality control of digital mammograms, as a chief administrator of digital mammogram readings in Ishikawa prefecture from 2004. Her current research centers around clinical evaluation of digital breast tomosynthesis.

Yoshikazu Uchiyama is an associate professor of the faculty of life science at Kumamoto University in Japan. He has a PhD in computer science from Miyazaki University in Japan. He has research interests in medical image processing, image recognition, and computer-aided diagnosis.

Junji Shiraishi is a professor of the faculty of life sciences at Kumamoto University, Japan. Before he joined Kumamoto University, he worked at the Kurt Rossmann Laboratories for Radiologic Image Research in the University of Chicago for 8 years as an assistant professor of research associate. He has been working in the areas of computer-aided diagnosis for chest radiography, nuclear medicines, ultrasonography, and the evaluation of their performances by utilizing observer performance studies with ROC/FROC analysis.