Using breast radiographers' reports as a second opinion for radiologists' readings of microcalcifications in digital mammography

[Title page]

Title

Using Breast Radiographers' Reports as a Second Opinion for Radiologists' Readings of microcalcifications in Digital Mammography

Short title

Breast Radiographers' Reports as a Second Opinion

Type of Manuscript

Full Papers

Authors *Corresponding author

Rie Tanaka[1]*, PhD., Miho Takamori[2], Yoshikazu Uchiyama[3], PhD.,
Robert M. Nishikawa[4], PhD., Junji Shiraishi[3] PhD.

Affiliations

[1] School of Health Sciences, College of Medical, Pharmaceutical and Health Sciences, Kanazawa University; 5-11-80 Kodatsuno, Kanazawa, 920-0942 Japan

[2] Department of Radiology, Ishikawa Prefectural Hospital; 2-1 Kuratsuki-Higashi, Kanazawa, Japan 920-8530

[3] Faculty of Life Sciences, Kumamoto University; 4-24-1 Kuhonji Kumamoto, Japan 862-0976

[4] Department of Radiology, University of Pittsburgh; 3362 Fifth Ave, Pittsburgh, PA, 15213 USA

[Blind title page]

Title

Using Breast Radiographers' Reports as a Second Opinion for Radiologists' Readings of microcalcifications in Digital Mammography

Short title

Breast Radiographers' Reports as a Second Opinion

Abstract

Objective

The aim of this study was to investigate a practical method for incorporating radiographers' reports with radiologists' readings of digital mammograms.

Methods

This simulation study was conducted using data from a free-response receiver operating characteristic (FROC) observer study obtained with 75 cases　(25 malignant, 25 benign, and 25 normal cases) of digital mammograms. Each of the rating scores obtained by six breast radiographers was utilized as a second opinion for four radiologists' readings with the radiographers' reports. A logical "OR" operation with various criteria settings was simulated for deciding an appropriate method to select a radiographer's report in all combinations of radiologists and radiographers. The average figure of merit (FOM) of the radiologists' performances was statistically analyzed using a jackknife procedure (JAFROC) to verify the clinical utility of using radiographers' reports.

Results

Potential improvement of the average FOM of the radiologists' performances for identifying malignant microcalcifications could be expected when using radiographers' reports as a second opinion. When the threshold value of 2.6 in BI-RADS assessment was applied to adopt/reject a radiographer's report, FOMs of radiologists' performances were further improved.

Conclusion

When using breast radiographers' reports as a second opinion, radiologists' performances potentially improved when reading digital mammograms. It could be anticipated that radiologists' performances were improved further by setting a threshold

value on the BI-RADS assessment provided by the radiographers.

Advances in knowledge

For the effective use of a radiographer's report as a second opinion, radiographer's rating scores and its criteria setting for adoption/rejection would be necessary.

[Main text]

Introduction

The number of mammography examinations has been increasing. In general, double reading by two radiologists is recommended in many countries, resulting in a shortage of radiologists. In order to solve this problem, the use of computer-aided detection (CADe) systems and/or reporting by radiographers as a second opinion for radiologists' readings have been investigated. Since many research groups have demonstrated the clinical usefulness of various CADe systems [1-6], CADe is accepted as the standard of care in the US and is used in approximately 75% of screening exams [6,7]. On the other hand, reporting by radiographers has been recommended as another approach to aid radiologists' readings in several countries [8-10]. Although the use of CADe requires initial costs for introducing CADe systems, reporting by radiographers can be applied in almost all medical institutions without any additional facility costs. However, the practical utilization of radiographers' reports has not been established, because radiographers' reading skills vary and the insufficient evaluation of diagnostic accuracy provided by utilizing radiographer's reports.

The basic concept of "reporting by radiographers" was first developed by the College of Radiographers in 1997 [8]. Since the first practical guidance on reporting by radiographers was published, there have been a number of publications related to reporting by radiographers in the past decade, mainly in the UK [10-14]. To utilize radiographers' reports more practically, a new reading protocol, double reading by two radiographers, was suggested. This reading protocol is a double reading using 'non-discordant radiographer only (double) readings' (NDROR), in which concordant cases are automatically recalled for assessment and discordant cases (where two readers disagree over an interpretation) are arbitrated by an experienced radiologist or breast clinician [15]. Several studies confirmed the benefits of reporting by radiographers in screening mammograms, including cost effectiveness benefits [16-20]. Although a radiographer's report is expected to be a substitute for one of the radiologists in a double reading situation, questions remain, such as how to utilize a radiographer's report in the most effective manner, as well as how to provide the best

services to the patient without loss of quality.

For reporting by radiographers to be effective, i.e., the highest cancer detection rate and the lowest recall rate, it is necessary to investigate the acceptance criteria to adopt or reject a radiographer's report, as well as the acceptable performance level of a radiographer's report. The aim of this study was to investigate a practical method how a radiologist would adopt or reject a report provided by a single radiographer, and to demonstrate the potential usefulness of radiologists reading digital mammograms utilizing radiographers' reports. We used an existing free-response receiver operating characteristic (FROC) observer study [21,22] data set to simulate all possible combinations for evaluating radiologists' performances with and without radiographers' reports.

Materials and Methods

Study design

This simulation study was conducted by using data from an existing previously conducted FROC observer study. The study was originally performed to compare diagnostic performances for identifying malignant clustered microcalcifications between four breast radiologists and six breast radiographers. In order to simplify a comparison between breast radiologists and radiographers, this observer study was aimed to detect clustered microcalcifications only, and there were no case samples with breast masses. The marks obtained in the FROC observer study, which included the location of a lesion and the rating score of the confidence level for malignancy, were used as the radiologists' and radiographers' reports.

In this simulation study, each of the rating scores obtained by the six breast radiographers was utilized as a second opinion for the four radiologists' readings with a radiographer's report. Since there is one radiographer's report in the real clinic, we did not use the average rating scores of the six radiographers, but used the individual rating scores as a radiographer's report for each of the four radiologists, and then, evaluated the performance of the four radiologists by averaging.

A logical "OR" operation with various threshold values to adopt or reject the

radiographers' reports was simulated for all combinations of radiologists and radiographers by using a rating score obtained in the FROC observer study. The average figure of merit (FOM) of radiologists' performances was statistically analyzed using jackknife free-response receiver operating characteristic (JAFROC) analysis to verify the clinical utility of using radiographers' reports [21,22].

FROC Dataset

The mammographic case samples used in the original FROC observer study were from the Digital Database for Screening Mammography (DDSM) provided by the University of South Florida (USF) [23-25]. Each case included two images of each breast, associated patient information (age at time of study, ACR breast density rating, subtlety rating for abnormalities, BI-RADS assessment category[26], ACR keyword description of abnormalities), and "ground truth" information about the locations and types of suspicious regions. In the study there were 75 cases consisting of 50 digital mammograms with clustered microcalcifications (25 malignant and 25 benign) and 25 normal cases, in consideration of observers' workloads. All cases were selected based on "subtlety" to include several challenging cancers that were nevertheless correctly detected by skilled readers. Table 1 displays the number of cases in each subtlety category. Note that the digital mammograms of patients with microcalcifications scattered over 20 mm in diameter were excluded from the case samples to secure location accuracy in the FROC observer study. The normal cases were selected by stratified sampling [27] to match the age distribution to that of cancer cases.

Four board-certified breast radiologists with experience ranging from 6 to 14 years (mean, 11 years) and six board-certified breast radiographers with experience ranging from 3 to 20 years (mean, 10 years) — all of whom are trained and certified for reading mammography by the Japan Central Organization on Quality Assurance of Breast Cancer Screening — participated in the FROC observer study. Observers determined the locations and confidence ratings on 75 digital mammograms (4 views /case). Digital mammograms were displayed on a high resolution liquid crystal display (LCD) for mammography (Nio 5M, BARCO) and observed by using a publically

available computer interface (ROC Viewer ver. 11.4.0.3 developed by Japanese Society of Radiological Technology) [28]. The display allowed the observers to determine the locations and ratings for the confidence level from 0.0 to 1.0 just by clicking on a mammogram.

When the observer found microcalcifications, he/she categorized it to a BI-RADS assessment as usual, and then translated it to a rating score in his/her mind. Figure 1 shows the relationship between the "Assessment" of BI-RADS Atlas and the rating scale used in our observer study. By referring to this figure, the observers were trained in how to translate a BI-RADS assessment into a 0.0 to 1.0 scale, and thus no BI-RADS assessment was given by the observer directly. In the observer study, viewing order of the 75 cases was randomized to minimize reading-order effects [29,30].

Simulation of Radiologists' Readings with a Radiographer's reporting

Each combination of a radiologist's rating and a radiographer's rating was simulated by using a logical "OR" operation with a criteria setting for original ratings obtained in the FROC study and those of the individual radiographers. We simulated all combinations of the six radiographers and four radiologists and averaged the radiologists' performances for each of the six radiographers.

The response of a radiologist reading with a radiographer's report differs depending on with/without the cut-off threshold value. In this logical "OR" operation, concordant cases were automatically assigned the higher rating provided by the radiologist or the radiographer. On the other hand, when the radiologist and radiographer disagreed over the interpretation, the assessment of the "OR" operation without the cut-off threshold value was always "Positive", whereas the logical "OR" operation with the cut-off threshold value could output either "Positive" or "Negative", depending on the rating score provided by the radiologist or the radiographer.

For example, when both a radiologist and a radiographer marked the same lesion in the FROC study (i.e., two marks were located within a threshold distance), this identification was adopted and the higher rating remained as a radiologist's rating with a radiographer's reporting (A < A' → A'') (Fig. 2a). Note that the threshold distance

was empirically determined as 20mm in this simulation study. If a lesion was marked only by a radiographer and the rating was higher than the pre-determined cut-off threshold value (i.e., 0.20), the mark was adopted and the rating remained as a radiologist's rating with a radiographer's reporting (B → B'', C → reject) (Fig. 2b). In the same way, if a lesion was marked only by a radiologist and the rating was higher than the pre-determined cut-off threshold value (i.e., 0.20), the mark was adopted and the score remained as a radiologist's rating with a radiographer's reporting (D' → D'') (Fig. 2c). In other words, a mark with a lower rating (equal or less than the pre-determined cut-off threshold value) in the original reading was automatically rejected from a radiologist's rating with a radiographer's reporting, even if the mark was identified by the radiologist (E' → reject) (Fig. 2c).

To evaluate the radiologists' performances with a radiographer's report, we changed the cut-off threshold value for the mark selection from 0.1 to 0.5, in increments of 0.1.

Data Analysis

In this study, a malignant case was considered to be actually positive, whereas, benign and normal cases were considered to be actually negative. Please note that a rating of 0.0 (BIRAD 2) was equal to "no action", and thus, the rating was not considered as a false positive (FP) in FROC analysis, but a rating ranging from 0.01 to 0.33 involved a certain percentage of malignancy for the microcalcifications as same as the interpretations of criteria between BIRADS 2 and 3.

A lesion marked by an observer with a distance less than 20mm from the center location of the gold standard was considered a true positive (TP). Otherwise, the mark was considered an FP. "Ground truth information" included in the DDSM was used as the reference standard.

Statistical analysis for a statistically significant difference between the radiologists' diagnostic performances without and with radiographers' reports was performed by using JAFROC ver. 4.2, provided by Chakraborty DP et al [21, 22]. 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 3 (Random Readers and Fixed Cases) with one-tailed test (p=0.05).

Results

Single Reading

Tables 2 and 3 show the diagnostic performances of the four radiologists and the six radiographers. They were obtained from the original FROC observer study for the detection of clustered malignant microcalcifications. There was no significant difference in the average FOM values between the radiologists and the radiographers. However, even though there was no significant difference, the average number of FPs of the radiologists was larger than that of the radiographers.

Simulation of Radiologists' Readings with a Radiographer's report

Figure 3 shows the relationship between the cut-off threshold values and average FOMs of the six radiologists with the individual radiographers' reports (RR). The average performance of the four breast radiologists was highest when the cut-off threshold value of 0.2 (i.e., the threshold value of 2.6 in BI-RADS assessment) was applied.

When the cut-off threshold value of 0.2 was applied for all cases, FOMs of radiologists' performances were generally improved by the radiographer's reports, in most combinations of radiologists and radiographers (Fig. 4).

Table 4 shows the average diagnostic performances of the four radiologists for the detection of malignant microcalcifications by utilizing each of the six radiographer's reports. Almost all the average FOM values were identical or improved by any of the radiographer's reports in comparison to those of conventional radiologists' readings without the radiographer's reports. Significant improvements were found when using three out of six radiographers' reports. In particular, radiologists' performances were significantly improved by using highly skilled radiographers' reports with high FOM or low FP (i.e., RT03 and RT04).

Discussion

Single Reading

In the original FROC observer study, there was no significant difference in the diagnostic performances between the radiologists and radiographers. This is because the observers were asked to find benign and malignant microcalcifications, which supported the fact that radiographers were as competent as radiologists in the detection of certain abnormalities [33]. In fact, Japanese board-certified breast radiographers are well-trained, because it is mandatory for them take essential courses, pass written and practical exams for measuring knowledge of image perception, and renew every five years.

In the meantime, the diagnostic performances obtained either by radiologists or by radiographers were relatively low in the original single reading. This is probably due to the difficulty of the dataset for the original observer study. The database consisted mostly of cases less than "Subtlety 3", i.e., subtle findings and containing several challenging cases, "Subtlety 1" or "Subtlety 2", which were nearly 25% of the total number of malignant cases. For example, four of nine cases, which were not identified by more than five observers, were categorized in "Subtlety 1" or "Subtlety 2." The combination of simple tasks for identifying microcalcifications and challenging cases was thought to be one of the major factors in lowering their diagnostic performances of the original single reading and in allowing for few differences between radiologists and radiographers. In addition, the relatively large number of FPs for benign cases might be another major factor to lower FOMs in radiologists compared to those in radiographers. The FPs in radiologists were mainly benign microcalcifications, while, those in radiographers were lesions other than benign microcalcifications. In this study, however, this enabled an evaluation of the clinical utility of using radiographers' reports for detecting missed cancers and/or eliminating FP findings for benign microcalcifications.

Simulated radiologists' readings with a radiographer's report

In various combinations of a radiologist's reading with a radiographer's report,

performance of a radiologist's reading was likely to be improved, regardless of the diagnostic performances of individual radiologists and radiographers. These results indicate that a radiographer's report could be used as a second opinion in cases of readings of malignant microcalcifications on digital mammography.

We used a modified logical "OR" operation with a cut-off threshold value for this simulation study. When the conventional "OR" operation without the cut-off threshold value was applied, utilization of a radiographer's report always increased the number of FPs, whereas no degradation of the number of TPs was compensated. Therefore, the utilization method of a radiographer's report based on a "recall if one suggests" policy (the conventional logical "OR" operation) always increases recall rate unnecessarily and thus is not practical [34].

To solve this detrimental effect by using the conventional "OR" operation, we introduced a cut-off threshold value to adopt or reject a radiographer's report. For minimizing the number of FPs by keeping highest sensitivity, a radiographer's report should be selected to be utilized for a radiologist's reading by a pre-determined cut-off threshold value on rating scores obtained in an FROC observer study or in a similar way.

In terms of clinical implementation, the cut-off threshold values should be determined practically and correlated to a clinical parameter. The cut-off threshold value used in this study was directly correlated to the "Assessment" value utilized in BI-RADs assessment, and thus, it could be used without any numerical correction and normalization.

The diagnostic performances of the breast radiologists were highest at the cut-off threshold value of 0.2 (i.e. 2.6 in BI-RADS assessment). The results can be explained by the following; high-ranked (i.e., BI-RADS assessment 3) lesions marked by radiologists should remain, because they are very likely true positives, whereas some benign-like lesions marked by radiographers and considered as BI-RAD assessment 2 (i.e., rating scores were less than 0.2) could be rejected. Therefore, radiographers' rating scores and its criteria setting for adoption or rejection, by taking into account the BI-RADS assessment, is necessary for the effective use of a radiographer's reporting

as a second opinion.

On the other hand, because clinically only one radiographer would ever provide a report, there would be no choice for a radiologist whether she/he could be selected as an assisted radiographer. However, our results promised that no radiographer's report affected the performance of a radiologist detrimentally if only the ratings more than the cut-off value are utilized for the assistance of a radiologist's interpretation.

Through this simulation study, we also confirmed that utilizing highly skilled radiographers' reports with a high FOM and low FPs further contributed to improving radiologists' performances. In other words, this result supports the importance of training programs for radiographers reading screening mammograms, as mentioned in the published literature [15, 35]. Further studies are required to expand on the type of abnormalities, such as masses and architectural distortions.

A limitation of this study was that only the detection of microcalcifications was studied. Other relevant lesions, such as masses, architectural distortions, and asymmetries were not included. These non-calcific lesions are a different type of detection task. Microcalcifications are, in general, small and high contrast objects and other than benign calcifications, there are not a lot of normal breast structures in a mammogram that mimic microcalcifications. Soft-tissue masses are often obscured and mimicked by the superposition of normal breast anatomy making the detection of breast masses difficult without a lot of false detections. Whether combining radiologists' ratings with radiographers' ratings for breast masses is beneficial, needs to be studied.

Conclusions

We investigated the potential usefulness of radiologists' readings with breast radiographers' reports by using data from an FROC observer study. The diagnostic performance of the radiologists for identifying malignant microcalcifications was generally (two out of four) improved by utilizing radiographers' reports with no detrimental effect for the performance of all breast radiologists. In particular, reporting by highly skilled radiographers made great contributions to improving radiologists' diagnostic accuracy. To effectively use a radiographer's report as a second opinion, i.e.,

improved accuracy of cancer detection with the lowest recall rate, the radiographer's rating scores and a criteria setting for adoption or rejection is necessary.

References

[1] 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.

[2] Cole EB, Zhang Z, Marques HS, Nishikawa RM, Hendrick RE, Yaffe MJ, et al. Assessing the stand-alone sensitivity of computer-aided detection with cancer cases from the Digital Mammographic Imaging Screening Trial. Am J Roentgenol. 2012;199:W392-401.

[3] Gilbert FJ, Astley SM, McGee MA, Gillan MG, Boggis CR, Griffiths PM, et al. Single reading with computer-aided detection and double reading of screening mammograms in the United Kingdom National Breast Screening Program. Radiology. 2006;241:47–53.

[4] Gilbert FJ, Astley SM, Gillan MG, Agbaje OF, Wallis MG, James J, et al. Single reading with computer-aided detection for screening mammography. N Engl J Med. 2008;359:1675–1684.

[5] 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.

[6] Boyer B, Balleyguier C, Granat O, Pharaboz C. CAD in questions/answers Review of the literature. Eur J Radiol. 2009;69:24-33.

[7] Rao VM, Levin DC, Parker L, Cavanaugh B, Frangos AJ, Sunshine JH. How widely is computer-aided detection used in screening and diagnostic mammography? J Am Coll Radiol. 2010;7:802-5.

[8] The College of Radiographers. Reporting by Radiographers: a Vision Paper. London: The College of Radiographers, 1997.

[9] Report from the ministry of health, labor and welfare. Tokyo, The ministry of health, labor and welfare (http://www.mhlw.go.jp)

[10] Friedenberg RM. The role of the super technologist. Radiology. 2000;215:630-3.

[11] Paterson AM, Price RC, Thomas A, Nuttall L. Reporting by radiographers: a policy and practice guide. Radiography. 2004;10: 205-212.

[12] The College of Radiographers. Medical Image Interpretation and Clinical Reporting by Non-Radiologists: The Role of the Radiographer. London. London: The College of Radiographers, 2006.

[13] The Royal College of Radiologists. Standards and recommendations for the reporting and interpretation of imaging investigations by non-radiologist medically qualified practitioners and teleradiologists. London: The Royal College of Radiologists, 2011

[14] The College of Radiographers. Preliminary Clinical Evaluation and Clinical Reporting by Radiographers: Policy and Practice Guidance. London: The College of Radiographers, 2013.

[15] NHS Breast Screening Programme. Annual Review 2011. Sheffield: NHS Cancer Screening Programmes, 2011.

[16] Bennett RL, Sellars SJ, Blanks RG, Moss SM. An observational study to evaluate the performance of units using two radiographers to read screening mammograms. Clin Radiol. 2012;67:114-21.

[17] Duijm LE, Groenewoud JH, Fracheboud J, van Ineveld BM, Roumen RM, de

Koning HJ. Introduction of additional double reading of mammograms by radiographers: effects on a biennial screening programme outcome. Eur J Cancer. 2008;44:1223-8.

[18] van den Biggelaar FJ, Nelemans PJ, Flobbe K. Performance of radiographers in mammogram interpretation: a systematic review Breast. 2008;17:85-90.

[19] Wivell G, Denton ER, Eve CB, Inglis JC, Harvey I. Can radiographers read screening mammograms? Clin Radiol. 2003;58:63-7.

[20] Duijm LE, Groenewoud JH, Fracheboud J, de Koning HJ. Additional double reading of screening mammograms by radiologic technologists: impact on screening performance parameters. J Natl Cancer Inst. 2007;99:1162-70.

[21] Chakraborty DP, Berbaum KS. Observer studies involving detection and localization: modeling, analysis, and validation. Med Phys. 2004;31:2313-30.

[22] Chakraborty DP. Validation and Statistical Power Comparison of Methods for Analyzing Free-response Observer Performance Studies, Acad Radiol 2008;15: 1554-1566.

[23] University of South Florida Digital Database for Screening Mammography (DDSM) http://marathon.csee.usf.edu/Mammography/Database.html

[24] Heath M, Bowyer KW, Kopans D. Current status of the Digital Database for Screening Mammography. In: Digital Mammography, edited by Karssemeijer N., Thijssen M., Hendriks J., and L.van Erming. Kluwer Academic, Dordrecht, 1998;457-460.

[25] Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer WP. The Digital Database for Screening Mammography, In: Proceedings of the Fifth International Workshop on Digital Mammography, Yaffe MJ, ed., Medical Physics Publishing, 2001;212-218.

[26] American College of Radiology. ACR BIRADS: Breast Imaging Reporting and Data System: breast imaging atlas, Fourth Edition. Reston, VA: 2003. ACR BI-RADS: Mammography

[27] Chaudhuri A. 3. Methods of Sample Selection and Estimation with their properties. In Essentials of Survey Sampling. PHI Learning. 2010

[28] Shiraishi J, Fukuoka D, Hara T, Abe H. Basic concepts and development of an all-purpose computer interface for ROC/FROC observer study. Radiol Phys Technol. 2013;6:35-41.

[29] Gur D, Rockette HE, Armfield DR, Blachar A, Bogan JK, Brancatelli G, et al. Prevalence effect in a laboratory environment. Radiology. 2003;228:10-4.

[30] Gur D, Bandos AI, Cohen CS, Hakim CM, Hardesty LA, Ganott MA, et al. The "laboratory" effect: comparing radiologists' performance and variability during prospective clinical and laboratory mammography interpretations. Radiology. 2008;249:47-53

[31] Zheng B, Chakraborty DP, Rockette HE, Maitz GS, Gur D. A comparison of two data analyses from two observer performance studies using jackknife ROC and JAFROC. Med Phys 2005; 32:1031–1034.

[32] Chakraborty DP, Winter LH. Free-response methodology: alternate analysis and a new observer-performance experiment. Radiology 1990; 174:873–881.

[33] Wivell G, Denton ER, Eve CB, Inglis JC, Harvey I. Can radiographers read screening mammograms? 2003;58:63-7.

[34] Beam CA, Sullivan DC, Layde PM. Effect of human variability on independent double reading in screening mammography. Acad Radiol. 1996;3:891-7.

[35] Pauli R, Hammond S, Cooke J, Ansell J. Radiographers as film readers in screening mammography: an assessment of competence under test and screening conditions. Br J Radiol. 1996;69:10-4.

Figure Legends

Fig. 1: Relationship between BI-RADS Assessment and a rating scale for the confidence level in our observer study.

Fig. 2   Example of selections of responses for simulated radiologist's rating utilizing a radiographer's report

Fig. 3   Relationship between the threshold value for cut-off rating values and average FOMs of 4 radiologists with individual radiographer's report (BR: breast radiographer, RT: radiographer)

Fig. 4   Change of the figure of merit (FOM) of radiologists' performances with and without each of 6 radiographers' report (BR: breast radiographer, RT: radiographer)

Table 1 Number of cases in each subtlety

Table 2   Diagnostic performance of 4 breast radiologists (BR) for the detection of malignant microcalcifications

Table 3   Diagnostic performance of 6 radiographers (RT) for the detection of malignant microcalcifications

Table 4   FOMs of 4 radiologists with and without each of 6 radiographers' report (RR) at the cut-off threshold value of 0.2 (BR: breast radiographer, RT: radiographer)

Table 1 Number of cases in each subtlety

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

Table 2 Diagnostic performance of 4 breast radiologists (BR) for the detection of malignant microcalcifications

	FOM	Sensitivity (%)	FP [/case]	Experience years
BR01	0.732	56.0	0.39	14
BR02	0.660	60.0	0.45	4
BR03	0.636	44.0	0.24	14
BR04	0.657	40.0	0.13	11
Average	0.671±0.021	50.0±9.5	0.30±0.14	10.8

Table 3 Diagnostic performance of 6 radiographers (RT) for the detection of malignant microcalcifications

	FOM	Sensitivity (%)	FP [/case]	Experience years
RT01	0.682	64.0	0.33	10
RT02	0.724	64.0	0.19	11
RT03	0.636	24.0	0.03	20
RT04	0.688	32.0	0.05	3
RT05	0.679	40.0	0.15	6
RT06	0.668	56.0	0.37	5
Average	0.680±0.012	46.7±17.1	0.19±0.14	9.2

Table 4 FOMs of 4 radiologists with and without each of 6 radiographers' report (RR) at the cut-off threshold value of 0.2

	without RR	with RR					
		RT01	RT02	RT03	RT04	RT05	RT06
BR01	0.732	0.722	0.733	0.732	0.756	0.731	0.733
BR02	0.660	0.682	0.703	0.684	0.700	0.670	0.692
BR03	0.636	0.651	0.654	0.654	0.672	0.652	0.635
BR04	0.657	0.706	0.710	0.676	0.700	0.709	0.678
Average	0.671	0.690	0.700	0.686	0.707	0.690	0.685
p value		0.107	0.048	0.032	0.002	0.094	0.095

Fig. 1: Relationship between BI-RADS Assessment and a rating scale for the confidence level in our observer study.
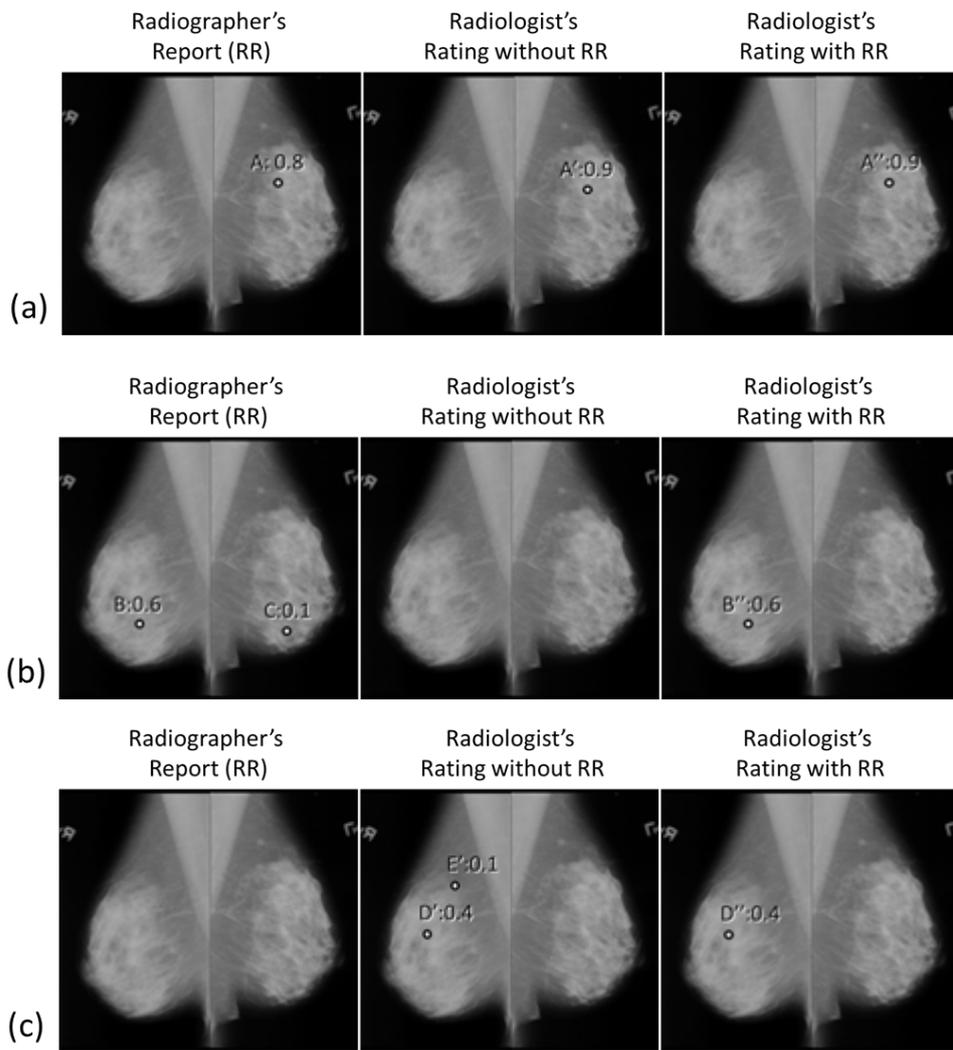
Fig. 2   Example of selections of responses for simulated radiologist's rating utilizing a radiographer's report
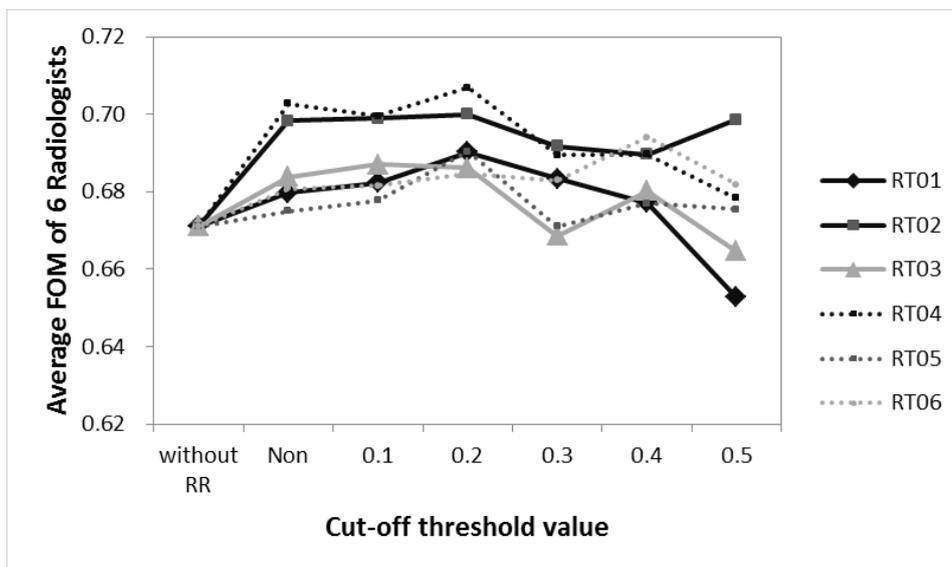
Fig. 3   Relationship between the threshold value for cut-off rating values and average
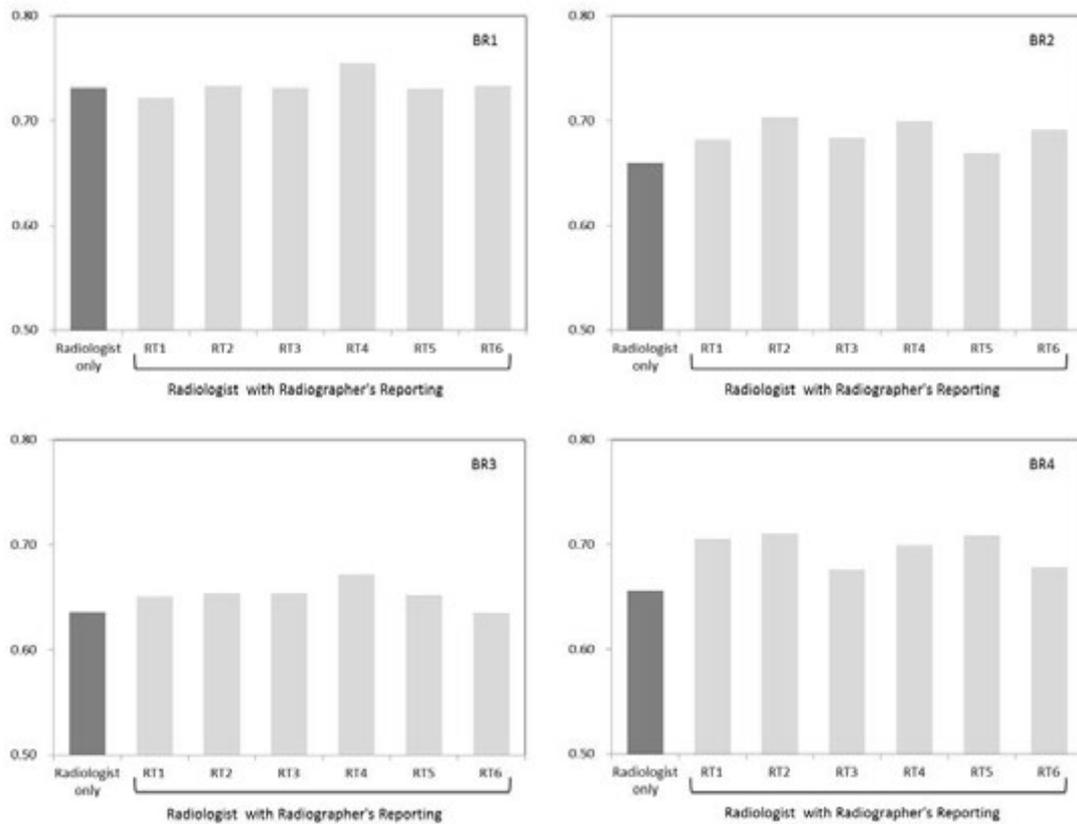FOMs of 4 radiologists with individual radiographer's report

Fig. 4　Change of the figure of merit (FOM) of radiologists' performances with and without each of 6 radiographers' report

Table 1 Number of cases in each subtlety

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

Table 2　Diagnostic performance of 4 breast radiologists (BR) for the detection of malignant microcalcifications

	FOM	Sensitivity (%)	FP [/case]	Experience years
BR01	0.732	56.0	0.39	14
BR02	0.660	60.0	0.45	4
BR03	0.636	44.0	0.24	14
BR04	0.657	40.0	0.13	11

	FOM	Sensitivity (%)	FP [/case]	Experience years
Average	0.671±0.021	50.0±9.5	0.30±0.14	10.8

Table 3 Diagnostic performance of 6 radiographers (RT) for the detection of malignant microcalcifications

	FOM	Sensitivity (%)	FP [/case]	Experience years
RT01	0.682	64.0	0.33	10
RT02	0.724	64.0	0.19	11
RT03	0.636	24.0	0.03	20
RT04	0.688	32.0	0.05	3
RT05	0.679	40.0	0.15	6
RT06	0.668	56.0	0.37	5
Average	0.680±0.012	46.7±17.1	0.19±0.14	9.2

Table 4 FOMs of 4 radiologists with and without each of 6 radiographers' report (RR) at the cut-off threshold value of 0.2

	without RR	with RR					
		RT01	RT02	RT03	RT04	RT05	RT06
BR01	0.732	0.722	0.733	0.732	0.756	0.731	0.733
BR02	0.660	0.682	0.703	0.684	0.700	0.670	0.692
BR03	0.636	0.651	0.654	0.654	0.672	0.652	0.635
BR04	0.657	0.706	0.710	0.676	0.700	0.709	0.678
Average	0.671	0.690	0.700	0.686	0.707	0.690	0.685
p value		0.107	0.048	0.032	0.002	0.094	0.095