SHORT COMMUNICATION

Clinical performance metrics of 3D stereoscopic digital mammography compared with 2D digital mammography: observer study

[1,2]AKIKO DAIDOJI, MD, [1]TAKATOSHI AOKI, MD, PhD, [1]SEIICHI MURAKAMI, PhD, [1]MARI MIYATA, MD,
[1]MASAMI FUJII, MD, PhD, [3]TAKEFUMI KATSUKI, MD, PhD, [3]YUZURU INOUE, MD, PhD, [4]YUKO TASHIMA, MD, PhD,
[4]YOSHIKA NAGATA, MD, PhD, [3]KEIJI HIRATA, MD, PhD, [4]FUMIHIRO TANAKA, MD, PhD and [1]YUKUNORI KOROGI, MD, PhD

[1]Department of Radiology, University of Occupational and Environmental Health, Kitakyushu, Kyushu, Japan
[2]Department of Radiology, Yamagata Prefectural Central Hospital, Yamagata, Japan
[3]First department of Surgery, University of Occupational and Environmental Health, Kitakyushu, Kyushu, Japan
[4]Second department of Surgery, University of Occupational and Environmental Health, Kitakyushu, Kyushu, Japan

Address correspondence to: Dr Takatoshi Aoki
E-mail: *a-taka@med.uoeh-u.ac.jp*

Objective: To assess the clinical performance of three-dimensional stereoscopic digital mammography (3DsDM) compared with two-dimensional digital mammography (2DDM) for breast lesion diagnosis with jackknife free-response receiver operating characteristics (JAFROC) observer study.

Methods: 40 pairs of standard-dose 2DDM and their 3DsDM images were used for an observer performance study. A total of 18 lesions were identified as the reference standard of actionable breast lesions (Breast Imaging Reporting and Data System Category 3 or more) by two breast radiologists. Ratings and locations of "lesions" determined by observers were utilized for assessing the statistical significance of differences between eight radiologists' performances with the 2DDM images and with the 3DsDM images in jackknife free-response receiver operating characteristic analysis.

Results: The average figure-of-merit values for all radiologists increased to a statistically significant degree, from 0.859 with the 2DDM images to 0.936 with the 3DsDM images (p < 0.001). The average sensitivity for detecting actionable lesions was improved from 74.3 to 92.4% at a false-positive rate of 0.2 per case by use of the 3DsDM images. The mean reading time per case with 2DDM images was not significantly different from that with 3DsDM images.

Conclusion: The use of 3DsDM would improve the observer performance for breast lesion without considerably extending the reading time.

Advances in knowledge: Use of 3DsDM improves radiologists' performance for breast lesion detection.

INTRODUCTION

Two-dimensional digital mammography (2DDM) has replaced film-screen mammography over the last few decades because of superior performance, image acquisition workflow, improved technologist productivity, and storage. However, diagnostic outcomes were limited by overlapping tissues, especially in dense breasts, due to the 2D nature of the projection images.[1]

With the advent of DM, three-dimensional stereoscopic DM (3DsDM) is now practical and provides direct in-depth views of the internal structure of the breast.[2] 3DsDM is expected to help radiologists overcome overlapping limitation, leading to potential reduction of false readings and thereby, further increasing diagnostic accuracy of breast cancer.

There are a few prior studies dealing with the usefulness of 3DsDM for breast lesion.[3,4] D'Orsi et al conducted a prospective study using prototype 3DsDM system in a high risk screening population and reported that 3DsDM improved specificity and reduced recall rate compared to 2DDM.[3] With recent advances of 3DsDM technology, low dose 3DsDM system has become commercially available. However, the clinical benefit of this system in the non-screening population has not been clarified. In addition, to our knowledge, no observer performance studies have been reported regarding the effect of the 3DsDM.

The purpose of this study was to assess the clinical performance of 3DsDM compared with 2DDM for breast lesion diagnosis with Jack knife free-response receiver operating characteristic (JAFROC) observer study.

PATIENTS AND METHODS

Digital mammography scanner and 3D monitor

Digital mammography was performed by using a clinical unit with a 3DsDM option (Amulet, FUJIFILM Corporation). A standard mediolateral oblique (MLO) mammography image (0-degree image) and its corresponding 4-degree image are treated as a stereo-pair in the 3DsDM. The spatial resolution of both 0- and 4-degree images acquired was 50 μmpixel^{-1}. 0-degree image was acquired with approximately the same dose as that with the standard DM examination and 4-degree image was acquired with approximately, the one-third dose as that with the standard DM examination. Therefore, the stereoscopic examination consisted of a total of six images (bilateral MLO views × 2 images per view and bilateral CC views × 1 image per view), effectively 1.17 times, the X-ray dose compared with the standard DM examination consisted of a total of four images.

The acquired stereopair images are sent to a 3D mammography workstation with a stereo 3D monitor (RadiForce GS521-ST, EIZO Nanao Corporation). The stereo 3D monitor consists of two 5-megapixel greyscale monitors for mammography with a half mirror. Without any reconstructive processes, one of the paired images is displayed on one of the 5-megapixel greyscale monitors and the other image is on the other monitor. The light coming from the top monitor reflected on the half mirror and the light from the bottom monitor comes through the half mirror. The polarization of the light coming through the half mirror rotates 90°. By wearing a pair of polarized glasses, the viewer's visual system fuses the stereo-paired images into a single instant, in-depth, 3D image of a breast (Figure 1). Our institutional review board approved this study and informed consent was waived.

Observer performance study

At retrospective review of the DM file of the 46 patients who underwent DM examinations for suspicion of a breast lesion, two experienced breast radiologists who did not participate in the observer study selected 40 consecutive cases [mean age, 54.7 years ± 12.1 (standard deviation), age range, 30–86 years] for observer performance study. Six patients having a history of breast surgery and/or unacceptable stereoimages due to apparent movement of the compressed breast between the two acquisitions of a stereopair were excluded. According to Breast Imaging Reporting and Data System (BI-RADS) guideline, a total of 18 lesions were identified as the reference standard of actionable breast lesions (BI-RADS Category 3 or more) by the two breast radiologists. 10 of 18 lesions were verified cancer by pathological examinations (8 invasive ductal carcinoma and 2 invasive lobular carcinoma). The remaining eight lesions were not cancers but were identified BI-RADS Category 3 or more by the two experienced breast radiologists with consensus.

Four certificated radiologists and four radiology residents participated in the observer performance study. An independent test consisting of two sessions of two series was performed. In the first session, half the observers interpreted with 2DDM, and the other half interpreted with 3DsDM. In the second session, the same experiment was performed, where data sets were presented in a different order from the one used in the first session. To reduce learning effects, the interval between the two sessions was maintained at least 4 weeks or longer.

Ratings and locations of "mass" were utilized for assessing the statistical significance of differences between radiologists' performances with the 2DDM and with the 3DsDM. Confidence level regarding the presence or absence of actionable lesion was

Figure 1. Image acquisition method and monitor of 3DsDM. The acquired stereopair images are sent to a 3D mammography workstation with a stereo 3D monitor (RadiForce GS521-ST, EIZO Nanao Corporation). The stereo 3D monitor consists of two 5-megapixel greyscale monitors for mammography with a half mirror. By wearing a pair of polarized glasses, the viewer's visual system fuses the stereo-paired images into a single instant, in-depth, 3D image of a breast. 3DsDM, three-dimensional stereoscopic digital mammography.
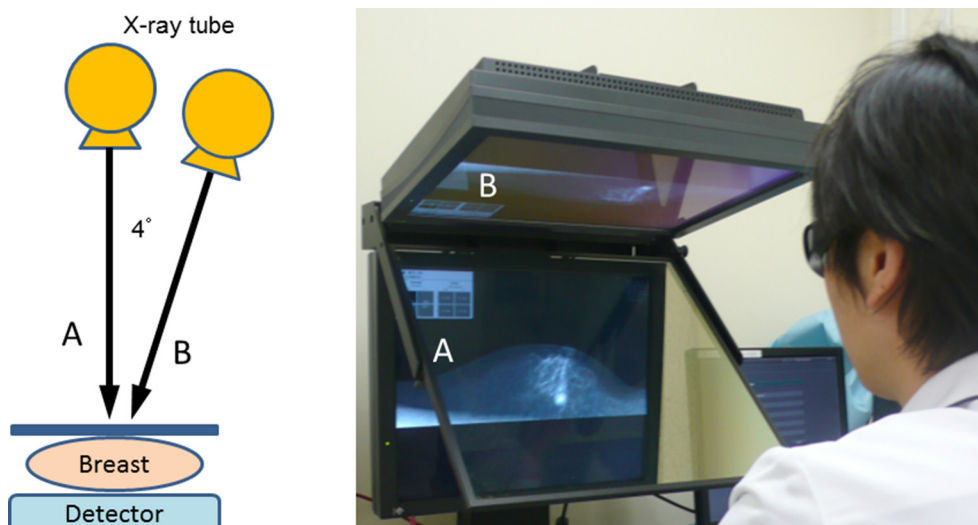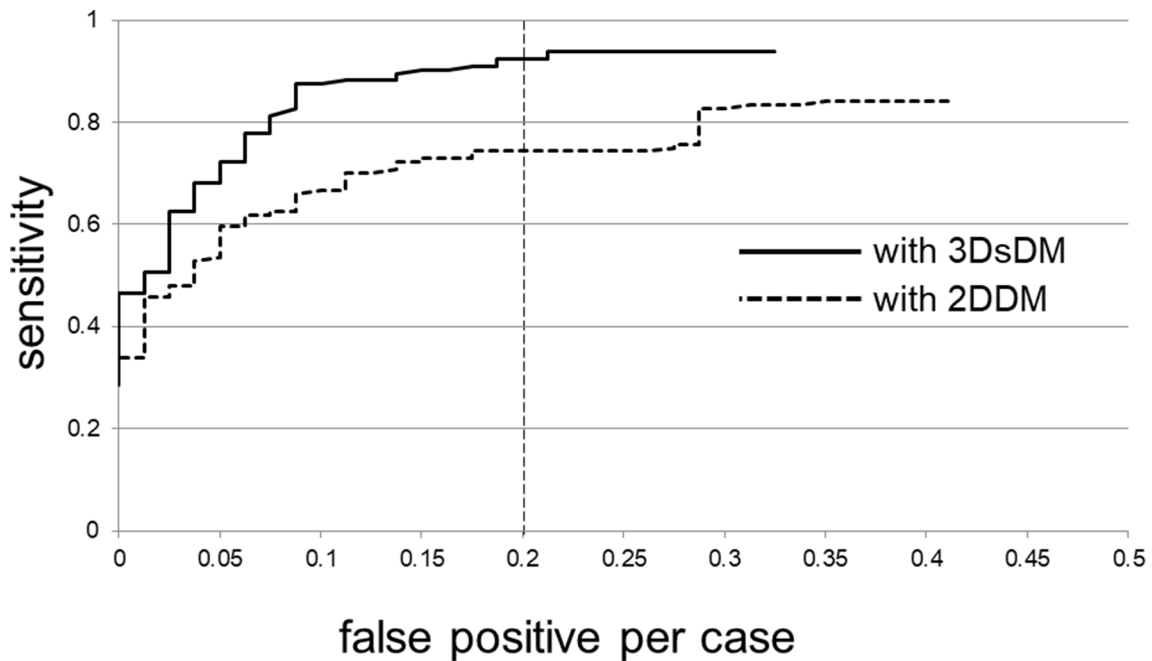
Figure 2. Comparison of average FROC curves for the radiologists' performance with 2DDM images and with 3DsDM images. 2DDM, two-dimensional digital mammography; 3DsDM, three-dimensional stereoscopic digital mammography; FROC, free-response receiver operating characteristic.

marked above both ends of the 10cm line with a black ball-point pen on the line for the rating. The reading time, rating and location for each case was recorded.

Statistical analysis

Assessment of each radiologist's diagnostic accuracy was determined via JAFROC observer study with a continuous rating scale.[5] We calculated a sequence of figure-of-merit (FOM) values for JAFROC analysis, and then built a pseudovalue matrix with these FOM values according to the equations. Finally, the FOM pseudovalue matrix was analysed with the use of analysis of variance techniques[6] for estimating statistically significant differences between the two FROC data sets obtained with 2DDM and with 3DsDM. The statistical significance of differences in the reviewing time was determined by use of a two-tailed paired Student's *t*-test.

RESULTS

The radiologists' performance in the detection of actionable breast lesions (BI-RADS Category 3 or more) with 2DDM and with 3DsDM is illustrated by the average FROC curves in Figure 2. By use of the 3DsDM images, the average sensitivity for detecting actionable lesions was improved from 74.3 to 92.4% at a false-positive rate of 0.2 per case.

The FOM and the reading time with 2DDM and with 3DsDM for each radiologist are shown in Table 1. The average FOM values for all radiologists increased to a statistically significant degree, from 0.859 with the 2DDM images to 0.936 with the 3DsDM images ($p < 0.001$) (Figure 3). When the 3DsDM was used, the average FOM for the residents was more improved (0.913 *vs* 0.825) than for the attending radiologists (0.959 *vs* 0.894).

The average reading time per case with 2DDM images and with 3DsDM images was 58.9 s and 63.3 s, respectively. The mean reading time per case with 2DDM images was not significantly different from that with 3DsDM images ($p = 0.101$). There was no statistically significant difference between the reading times in both residents (61.8 *vs* 69.9 s, $p = 0.39$) and attending radiologists (56.0 *vs* 56.8 s, $p = 0.91$) groups. None of the readers experienced eye fatigue during their reading sessions.

Table 1. Comparison of FOM with 2DDM images and with 3DsDM images

Observers		FOM	
		2DDM	3DsDM
Residents	A	0.847	0.940
	B	0.831	0.926
	C	0.862	0.958
	D	0.759	0.829
	Average	0.825	0.913 ($p < 0.001$)
Attending radiologists	E	0.904	0.963
	F	0.872	0.950
	G	0.881	0.963
	H	0.919	0.967
	Average	0.894	0.959 ($p < 0.001$)
	Average total	0.859	0.936 ($p < 0.001$)

FOM, figure-of-merit; 2DDM, two-dimensional digital mammography; 3DsDm, three-dimensional stereoscopic digital mammography; MLO, mediolateral oblique.

Figure 3. 3DsDM images of a 66-year-old female with invasive ductal carcinoma (arrows). In this case, architectural distortion was more easily derived by 3DsDM than by 2DDM in some observers. 2DDM, two-dimensional digital mammography; 3DsDM, three-dimensional stereoscopic digital mammography; MLO, mediolateral oblique.

Standard MLO image   Corresponding 4-degree image
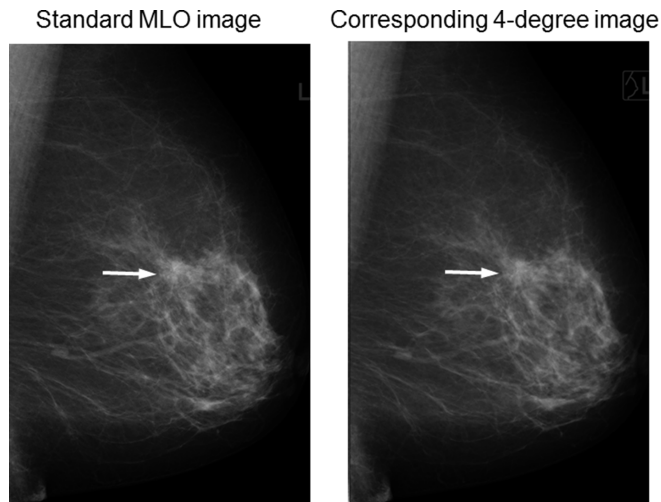
DISCUSSION

Due to the use of 3DsDM, improvement in diagnostic accuracy for breast lesion was experienced by all eight radiologists in the observer performance study. This finding may be considered to be consistent with two of the main predicted effects of stereoscopic DM: reduced susceptibility to tissue superposition and improved analysis of the volumetric structure of suspicious lesions and the surrounding tissue.

Although both observer categories benefited from the 3DsDM, those with less experience tended to benefit more. The average FOM for radiology residents with the 3DsDM images (FOM: 0.913) was higher than that for attending radiologists with the 2DDM images (FOM: 0.894). Attending radiologists can distinguish overlapping normal fibroglandular tissues at different depths from breast lesions comparatively easily, however, which is sometimes difficult for residents and these overlapping tissues tend to be false positive. Because the 3DsDM technique can overcome this limitation, it may be more useful for residents than for attending radiologists.

Digital breast tomosynthesis (DBT) has similar advantage to 3DsDM, and recent results of clinical studies assessing the DBT compared with 2DDM in breast cancer detection demonstrate the superiority of 2DDM plus DBT over 2DDM alone in terms of increased sensitivity and specificity and in reader performance.[7–10]

However, DBT has a disadvantage of additional reading time. According to the comparative study in a population-based screening program by Skaane et al, reading time with 2DDM images and with DBT images was 45 and 91 s, respectively.[11] Wallis et al reported an approximate doubling of average reading time of 67 s for 2DDM and 124 s for DBT.[12] Radiologists' performance in daily practice is affected not only by the diagnostic accuracy but also by interpretation time and this increase in time will have cost implications. In this study, the average reading time per case with 2DDM images and with 3DsDM images was approximately equal in both residents and attending radiologists. Use of 3DsDM may have great potential in a substantial improvement of the observer performance for breast cancer detection.

Although conventional receiver operating characteristics (ROCs) analysis is of limited value because only one lesion can be used per case and the location of the lesion cannot be taken into account in the evaluation, JAFROC allows us to evaluate radiologists' performance in diagnosing medical images by using multiple responses, each with information on the confidence level and location.[5] Therefore, we assessed the clinical performance of 3DsDM compared with 2DDM based on JAFROC observer study.

There are some limitations in our study. First, reference standard of the presence of actionable breast lesions was established by a consensus panel. However, this limitation is common to radiologists' observer performance studies in clinical situations. Second, novel synthesized two-dimensional mammography (s2DDM), generated via slab reconstruction from the tomosynthesis acquisition, was not evaluated in this study. Although the use of DBT in combination with 2DDM (2DDM/DBT) requires at least a doubling of radiation exposure, the average glandular dose was 39% lower in s2DDM/DBT compared with 2DDM/DBT.[13] Comparative studies between 3DsDM and s2DDM may be warranted. Third, only the MLO view was produced for stereoscopic imaging in consideration of total X-ray dose. Although the stereoscopic examination (total of six images) in this study imposes 17% more dosage compared with the standard DM examination, it is still much lower than 2DDM/DBT and acceptable value. By including CC stereoscopic images, the total X-ray dose increases but the diagnostic performance might be improved. Another limitation is the retrospective nature of this study. Further prospective study of a greater number of cases is likely necessary to confirm the clinical usefulness of 3DsDM.

CONCLUSION

Use of 3DsDM would improve the observer performance for breast lesion without considerably extending the reading time.

REFERENCES

1. Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med* 2005; **353**: 1773–83. doi: https://doi.org/10.1056/NEJMoa052911

2. Goodsitt MM, Chan HP, Hadjiiski L. Stereomammography: evaluation of depth perception using a virtual 3D cursor. *Med Phys* 2000; **27**: 1305–10. doi: https://doi.org/10.1118/1.599008

3. D'Orsi CJ, Getty DJ, Pickett RM, Sechopoulos I, Newell MS, Gundry KR, et al. Stereoscopic digital mammography:

improved specificity and reduced rate of recall in a prospective clinical trial. *Radiology* 2013; **266**: 81–8. doi: https://doi.org/10.1148/radiol.12120382

4. Muralidhar GS, Markey MK, Bovik AC, Haygood TM, Stephens TW, Geiser WR, et al. Stereoscopic interpretation of low-dose breast tomosynthesis projection images. *J Digit Imaging* 2014; **27**: 248–54. doi: https://doi.org/10.1007/s10278-013-9648-x

5. Chakraborty DP, Berbaum KS. Observer studies involving detection and localization: modeling, analysis, and validation. *Med Phys* 2004; **31**: 2313–30. doi: https://doi.org/10.1118/1.1769352

6. 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**: 723–31.

7. Gur D, Abrams GS, Chough DM, Ganott MA, Hakim CM, Perrin RL, et al. Digital breast tomosynthesis: observer performance study. *AJR Am J Roentgenol*

2009; **193**: 586–91. doi: https://doi.org/10.2214/AJR.08.2031

8. Michell MJ, Iqbal A, Wasan RK, Evans DR, Peacock C, Lawinski CP, et al. A comparison of the accuracy of film-screen mammography, full-field digital mammography, and digital breast tomosynthesis. *Clin Radiol* 2012; **67**: 976–81. doi: https://doi.org/10.1016/j.crad.2012.03.009

9. Rafferty EA, Park JM, Philpotts LE, Poplack SP, Sumkin JH, Halpern EF, et al. Assessing radiologist performance using combined digital mammography and breast tomosynthesis compared with digital mammography alone: results of a multicenter, multireader trial. *Radiology* 2013; **266**: 104–13. doi: https://doi.org/10.1148/radiol.12120674

10. Rose SL, Tidwell AL, Bujnoch LJ, Kushwaha AC, Nordmann AS, Sexton R. Implementation of breast tomosynthesis in a routine screening practice: an observational study. *AJR Am J Roentgenol* 2013; **200**:

1401–8. doi: https://doi.org/10.2214/AJR.12.9672

11. Skaane P, Bandos AI, Gullien R, Eben EB, Ekseth U, Haakenaasen U, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology* 2013; **267**: 47–56. doi: https://doi.org/10.1148/radiol.12121373

12. Wallis MG, Moa E, Zanca F, Leifland K, Danielsson M. Two-view and single-view tomosynthesis versus full-field digital mammography: high-resolution X-ray imaging observer study. *Radiology* 2012; **262**: 788–96. doi: https://doi.org/10.1148/radiol.11103514

13. Zuckerman SP, Conant EF, Keller BM, Maidment AD, Barufaldi B, Weinstein SP, et al. Implementation of synthesized two-dimensional mammography in a population-based digital breast tomosynthesis screening program. *Radiology* 2016; **281**: 730–6. doi: https://doi.org/10.1148/radiol.2016160366