Radiol Phys Technol (2013) 6:70–77
DOI 10.1007/s12194-012-0171-4

Usefulness of presentation of similar images in the diagnosis of breast masses on mammograms: comparison of observer performances in Japan and the USA

Chisako Muramatsu · Robert A. Schmidt · Junji Shiraishi · Tokiko Endo · Hiroshi Fujita · Kunio Doi

Received: 21 May 2012 / Revised: 6 July 2012 / Accepted: 22 July 2012 / Published online: 8 August 2012

Abstract Computer-aided diagnosis has potential in improving radiologists' diagnosis, and presentation of similar images as a reference may provide additional useful information for distinction between benign and malignant lesions. In this study, we evaluated the usefulness of presentation of reference images in observer performance studies and compared the results obtained by groups of observers practicing in the United States and Japan. The results showed that the presentation of the reference images was generally effective for both groups, as the areas under the receiver operating characteristic curves improved from 0.915 to 0.924 for the group in the US and from 0.913 to 0.925 for the group in Japan, although the differences were marginally ($p = 0.047$) and not ($p = 0.13$) statistically significant, respectively. There was a slight difference between the two groups in the way that the observers reacted to some benign cases, which might be due to differences in the population of screenees and in the socio-clinical environment. In the future, it may be worthwhile to investigate the development of a customized system for physicians in different socio-clinical environments.

Keywords Similar images · Computer-aided diagnosis · Breast masses · Mammograms · Image retrieval · Observer study

C. Muramatsu (✉) · H. Fujita
Department of Intelligent Image Information, Graduate School of Medicine, Gifu University, 1-1 Yanagido, Gifu, Japan
e-mail: chisa@fjt.info.gifu-u.ac.jp

R. A. Schmidt · K. Doi
Department of Radiology, The University of Chicago, 5841 S Maryland Ave, Chicago, IL 60637, USA

J. Shiraishi
Department of Medical Information Systems, School of Health Sciences, Kumamoto University, 2-31-1 Kurokami, Kumamoto, Japan

T. Endo
Department of Radiology, Nagoya Medical Center, 4-1-1 Nakaku Sannomaru, Nagoya, Aichi, Japan

K. Doi
Gunma Prefectural Collage of Health Sciences, 328-1 Kamiokimachi, Maebashi, Gunma, Japan

1 Introduction

Breast cancer is the most frequently diagnosed cancer and one of the leading causes of cancer deaths in women in Japan, the United States (US), and European countries [1–3]. Mammography is considered the most effective screening method for early detection of breast cancer for women at normal risk [4–6]. For improving the diagnostic accuracy and efficiency, computer-aided detection (CADe) was introduced [7–10], and its potential usefulness was indicated in an observer performance study [11] and in prospective studies [12–16]. Once a suspicious lesion is found, radiologists may determine whether it should be biopsied or followed up. However, diagnosis on mammograms can be difficult and requires proper training and reading experience. For assisting radiologists' reading, investigators have suggested computer-aided diagnosis (CADx), in which a computer provides the likelihood of malignancy of an unknown lesion and have reported the potential usefulness of CADx in distinguishing between benign and malignant lesions on mammograms [17–19]. In these studies, radiologists' performance in terms of the area (AUC) under the receiver operating characteristic (ROC) curve was improved with use of CADx; however, the studies indicated that the AUCs by many observers with CADx were lower than the AUC by the computer alone. One of the reasons might be that the result of computer analysis was summarized only in one numeral, i.e., the likelihood of malignancy, and the evidence was not clear to radiologists.

In recent years, mammography practice has been shifting from analogue to digital images. With implementation of Picture Archiving and Communication System (PACS), it became much easier to store and retrieve images from the previous examinations, and an effective use of stored data is expected. Radiologists, on the other hand, are trained and gain experience by reading many images in their clinical practice, in textbooks, and in training courses. Therefore, presentation of images that are similar to an unknown image can be an intuitive guide to reinforce the numerical likelihood of malignancy [20]. Different methods for automated selection of similar images have been investigated for diagnosis of chest radiographs [21, 22], thoracic computed tomographs [23, 24], and mammograms [25–31]. In some studies, reference images were selected on the basis of the predicted diagnosis [21, 22, 27, 28], whereas in other studies, images were selected by the similarity of the feature values [24, 26, 28]. For selecting similar images from the point of view of diagnosis, we have been investigating a method for quantifying the subjective ratings by radiologists [29, 30], as well as, a similarity index that takes into account the subjective similarity rated by radiologists [23, 31–33]. In our method, the similarity measure, called a psychophysical measure, was determined using an artificial neural network (ANN) which would be trained to learn the relationship between the subjective similarity ratings by radiologists and the computer-extracted image features.

Some of the above studies have indicated the potential usefulness of providing reference images together with other information such as the predicted diagnosis [22–24, 27, 34]. In these studies, it was not clear whether the presentation of reference images itself or the both images and other information together was helpful. Therefore, in order to evaluate the usefulness of providing similar images and to investigate the effect on radiologists in detail, we conducted the observer performance study to evaluate the radiologists' abilities in distinguishing between benign and malignant masses without and with similar images [35]. The result of this study was that, although the presentation of similar images provided beneficial effects, the average AUC was almost unchanged. One of the important findings in this study was that a reference image database must be carefully created so that it does not include "confusing" cases. When textbook-type cases are shown, radiologists will feel comfortable and can confidently and properly react to the given information. However, when atypical cases are presented, radiologists may become anxious, especially if they are cases of cancer. Another finding was that when a new case in question may be a benign-looking malignant case or a malignant-looking benign case, similar images would not be helpful. For these cases, radiologists' initial assessment is likely to be incorrect, and the presentation of similar images would only reinforce their incorrect decisions. This effect is, in fact, one that any type of CADx could have in common. When radiologists' initial judgment for a malignant-looking benign lesion was malignant, and the computer-estimated likelihood was also malignant, radiologists would become confident of their incorrect decision.

In this paper, we report the result from our second observer study after manual refinement of the database to exclude confusing cases. A group of radiologists practicing in the US and another group of radiologists and breast surgeons in Japan participated in the observer study, and the results from the two groups were compared.

2 Materials and methods

2.1 Case selection

Regions of interest (ROIs) including breast masses were used in this study. They were obtained from the Digital Database for Screening Mammography (DDSM) [36], which was made available by the University of South Florida. We initially collected 1568 ROIs, including 728 malignant and 840 benign masses [31]. ROIs containing microcalcifications which may influence the diagnosis of masses were excluded from the database. In the previous observer study [35], potentially confusing cases (benign-looking malignant and malignant-looking benign masses) were excluded from the reference database, which consisted of 365 malignant and 442 benign masses, by use of the computer-estimated likelihood of malignancy. However, because of the imperfect accuracy of the estimated likelihood, the database was suboptimal. In this study, for creating a better reference database, all of the cases were rated by a co-author (C.M.) for their difficulty in distinguishing between benign and malignant from 1 to 4, with 1 being difficult and 4 being easy. The images with unclear masses or with low quality that may not be helpful when presented as a reference were rated 0.

For the observer study, 100 cases, including 50 malignant and 50 benign masses, were selected from those rated 2, 3, and 4 to serve as study cases (unknown cases). They were selected by stratified randomization according to the size distributions of the database with 2.5 mm size bins, excluding those less than 5 mm and those larger than

25 mm. After removing all of the ROIs obtained from the same patients that were selected as the unknown cases, the cases rated 3 and 4 comprised the reference database, which included 429 malignant and 480 benign mass ROIs.

For each unknown image, 8 images each from the malignant and benign groups were selected as “similar” reference images, although 4 images each were presented in the monitor, and the next 4 images were provided only if an observer requested it. In the selection of similar reference images, the size criterion of no more than 50 % difference in the effective diameters was applied first. From the remaining cases, reference images were selected on the basis of the psychophysical similarity measures, which were determined by the ANN trained with 300 sample pairs in the previous study [32]. In the training of the ANN, image features characterizing the shape, contrast, and margin were used as input, and the subjective similarity data by radiologists based on the overall impression for diagnosis were used as the teacher. For avoiding having the same image presented more than 5 times as the first 4 images in 100 cases, the top 10 images with the highest similarity measures were preselected, and 4 of them were used. Note that they were called the reference images because, in some cases, there may be no “similar” images with very high similarity measures, especially the benign reference images for the malignant unknown cases and the malignant reference images for the benign unknown cases.

2.2 Observer performance studies

Observer studies for evaluating the usefulness of presenting reference images in the distinction between benign and malignant masses were conducted at the University of Chicago, Chicago, USA, and at Nagoya Medical Center, Nagoya, Japan. During the studies, the images were shown on a monochrome liquid crystal display monitor (ME511L/P4, 21.3 in., 2048 × 2560 pixels, 410 cd/m^2 luminance; Totoku Electric Co., Ltd.). The readings were conducted in the sequential reading mode, in which an observer was asked to provide his/her confidence level of a lesion being malignant on a continuous rating scale from 0.00 to 1.00, corresponding to “definitely benign” and “definitely malignant,” respectively. Immediately after the first rating, four “similar” benign images and four “similar” malignant images were presented on the right and left sides of the unknown case, and the observer was asked to reconsider his/her confidence level. If an observer requested it, next four benign and four malignant images would be shown.

The instructions to the observers were: (1) The purpose of this study is to investigate whether providing the similar known images can assist radiologists in the distinction between benign and malignant lesions on mammograms. (2) 100 unknown cases are included in this study. The training session including four cases is provided at the beginning of the study. (3) You are asked to provide your confidence level regarding the malignancy (or benignity) of a lesion with a bar displayed on the monitor by use of a mouse first without similar images, and then after observing the similar images. (4) For each unknown case, four most similar images each from benign and malignant lesions in the database are provided. If desired, you may observe additional four similar images by clicking a “show similar images 5–8” button. (5) There is no time limit.

Ten observers, including two attending breast radiologists, three breast imaging fellows, and five radiology residents, participated in the observer study in the US. The two attendings had 13 and 3 years of experience in reading mammograms, whereas the three fellows were in the first or second year of a breast imaging fellowship, and the residents were third- or fourth-year senior residents who had been trained in the breast-imaging section in their rotation. Eleven observers, including 10 radiologists and breast surgeons who were certified for breast image reading and one in training to be certified, participated in Japan. The ten certified physicians had a mean of 13 years of experience in reading mammograms. The results were evaluated by use of multi-reader multi-case (MRMC) ROC analysis (the University of Chicago, IL, USA) [37].

3 Results

The overall results indicated that AUCs without and with the presentation of the reference images were both high, probably because difficult cases were not included in this study. The AUCs without and with the reference images for the individual observers are listed in Table 1. The mean AUCs were slightly improved by providing the reference images, from 0.915 to 0.924 for the group in the US and from 0.913 to 0.925 for the group in Japan; however, the differences were not statistically significant for the Japan group and marginally significant for the US group. If a change in the confidence level of more than 0.1 in the direction of the correct diagnosis is considered a beneficial effect, on average, there were larger numbers of cases that the presentation of the reference images affected beneficially than those affected detrimentally. Figure 1 shows the numbers of beneficially and detrimentally affected cases for each observer. The average numbers of beneficially and detrimentally changed cases were 13 and 8, respectively, for the US group and 16 and 7, respectively, for the Japan group with p values of 0.04 and 0.01 by paired t test.

It may be noticed in the results that for the US observers, there were more beneficial effects to the malignant cases than to the benign cases, and the majority of the detrimental cases were benign cases. On the other hand, for the

Table 1 Areas under the receiver operating characteristic curves without and with the presentation of the reference images for the individual observers

Observers	US group		Japan group	
	Without	With	Without	With
A	0.951	0.962	0.939	0.942
B	0.972	0.978	0.893	0.924
C	0.940	0.938	0.936	0.947
D	0.947	0.942	0.969	0.941
E	0.942	0.943	0.880	0.880
F	0.906	0.931	0.888	0.919
G	0.874	0.877	0.879	0.888
H	0.887	0.913	0.905	0.958
I	0.871	0.874	0.918	0.940
J	0.863	0.885	0.901	0.885
K			0.936	0.952
Average (p value)	0.915	0.924 ($p = 0.047$)	0.913	0.925 ($p = 0.13$)

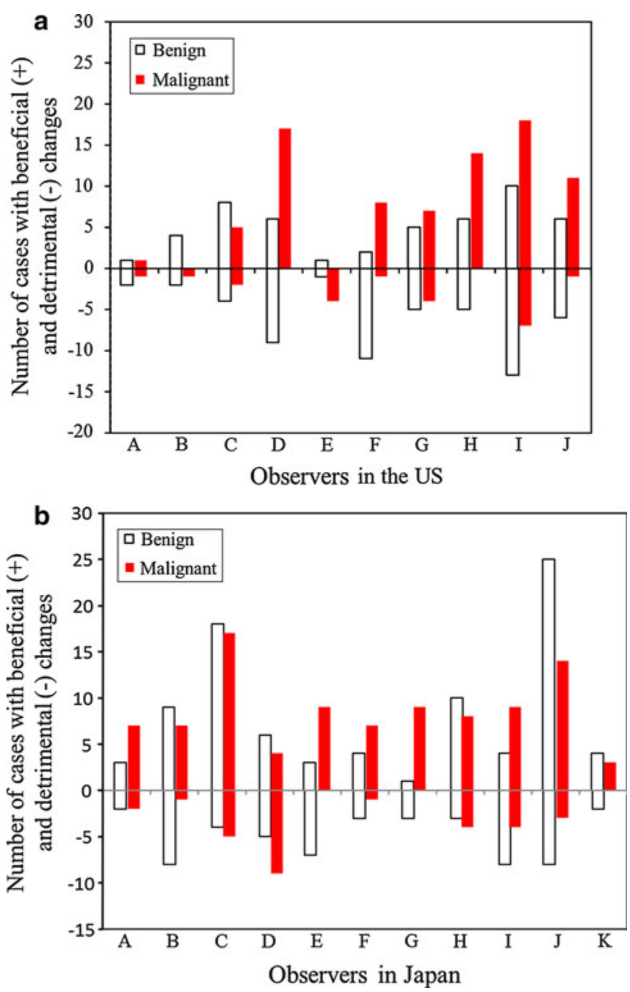

Fig. 1 Numbers of cases that the presentation of the reference images affected beneficially (positive) and detrimentally (negative) for the individual observers. **a** Observers in the US, and **b** observers in Japan

Japanese observers, the presentation of the reference images was helpful for both the malignant and the benign cases. Figure 2 shows the relationships between the average initial confidence levels and their changes, where positive changes correspond to the changes toward a correct diagnosis, for the two groups of observers. It is apparent in the figures that the presentation of the reference images had beneficial effects for many of the malignant cases. For the benign cases, however, it caused the US observers to increase their confidence levels toward malignant (indicated by an arrow in Fig 2a). The results indicate that some observers, regardless of their initial judgments as benign or uncertain, became worried after the reference images were presented. On the other hand, the average initial confidence levels for the benign cases by the Japanese observers were, on average, low for many cases, and the observers became confident of their judgment after the presentation of the reference images (indicated by an arrow in Fig 2b).

Figure 3 shows the relationship between the changes in the confidence levels before and after the presentation of the reference images by the two groups. The points in the right upper quadrant correspond to the cases in which the presentation of the reference images, on average, resulted in the beneficial changes for both groups of observers. An example of such cases is shown in Fig. 4. In this case, the unknown case was malignant. The initial judgments by the observers in both groups were mostly uncertain, and their confidence levels increased after the reference images were presented, with six of them increasing more than 0.1. On the other hand, there are some benign cases in the upper left quadrant in Fig. 3 for which the image presentation resulted in beneficial changes for the Japanese observers, but caused the detrimental changes for the US observers. Figure 5 shows one of such cases. In this case, the observers' initial judgments ranged from somewhat uncertain to likely benign, and the presentation of the reference images caused some US observers to increase their ratings, whereas most Japanese observers remained unchanged.

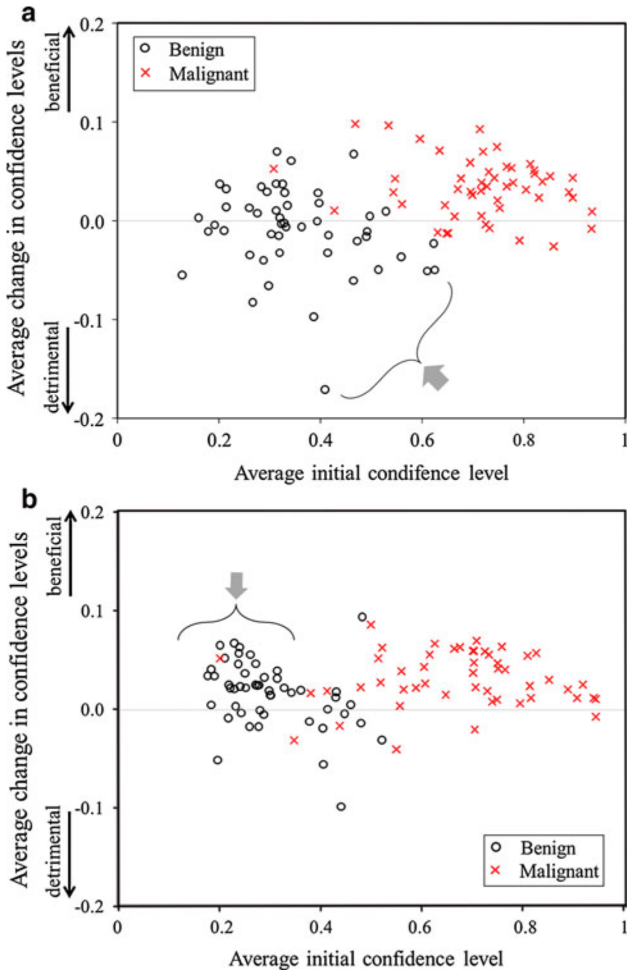

Fig. 2 Relationships between the average initial confidence levels and the changes in confidence levels toward (+) correct and (−) incorrect diagnosis by the **a** observers in the US and **b** observers in Japan. An *arrow* in **a** indicates the benign cases in which presentation of similar images caused detrimental effects, whereas an *arrow* in **b** indicates the benign cases in which the presentation caused beneficial effects

4 Discussion

The results of the observer studies showed a somewhat notable difference between the practitioners in the US and those in Japan in the sense of their reaction when the reference images were presented. The differences between the two groups seemed more prominent in the benign than in the malignant cases. One difference we observed during the reading sessions and also obtained in the observers' feedback was that the practitioners in the US primarily and dominantly consider margin characteristics in distinguishing between benign and malignant masses, whereas the practitioners in Japan consider the density of the masses, which relates to their elasticity, in addition to the margin characteristics. This may be due to the fact that Japanese women tend to have dense breasts, and physicians have a

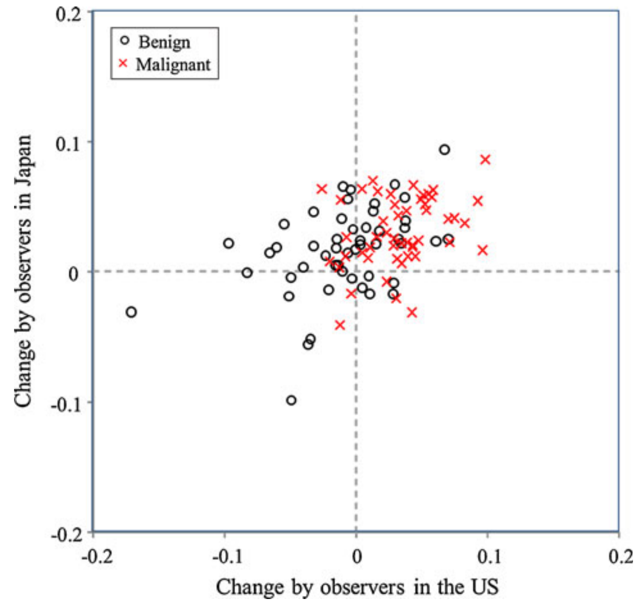

Fig. 3 Relationship between the changes in confidence levels before and after the presentation of the reference images between the two groups of observers

difficult time assessing margins more often than those in the US. It is also related to the fact that the breast cancer incidence rate increases with age in the US, whereas it peaks around the late 40 s in Japan. Therefore, the observers in Japan often complained about the use of ROIs without the availability of whole mammographic views during the observer study. When they read mammograms, the relative mass density in comparison with the normal breast tissue density of the patient is one of the important factors that they consider. However, with the lack of a whole view, it was difficult to see the grandular tissue density of the whole breast and the symmetry against the opposite breast.

Another perspective could be related to the number of law suits on missed cancers in the US. Although nobody wants to miss a cancer, physicians in the US may be particularly sensitive to missing one. This is manifested in the reported higher recall rates in the US than those in other countries. According to the study by the Physicians Insurers Association of America, breast cancer is the subject of the most frequent malpractice lawsuits filed, in which 41 % of all claims resulted in compensation averaging about $438,000 [38]. Dick et al. [39] have reported that, in their surveys, about a half of US radiologists responded that they had had a malpractice claim filed against them. In Japan, based on the statistics by the Supreme Court (http://www.courts.go.jp/saikosai/iinkai/izikankei/index.html), the number of medical lawsuits in each year is about 1000, in which internal medicine, surgery, orthopedics, and gynecology are the top four

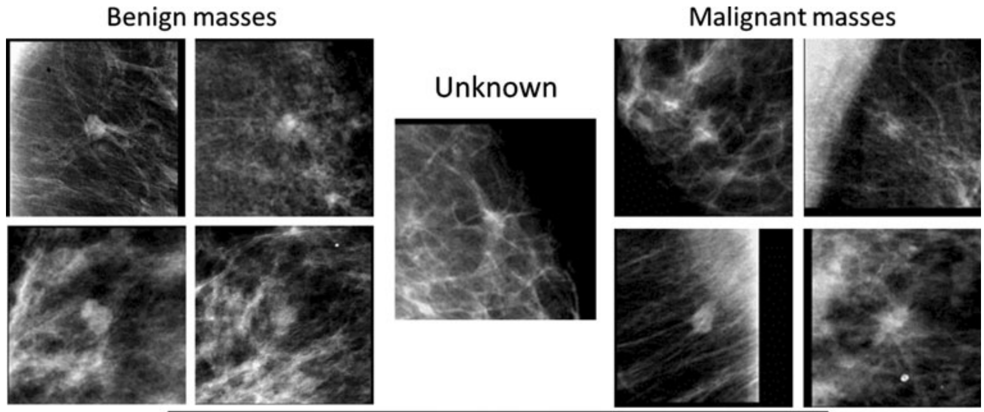

Fig. 4 A malignant case in which the presentation of the reference images resulted in beneficial changes for the both groups

	Average change	Number of observers who changed
Observers in the US	0.10 (0.47 → 0.57)	3
Observers in Japan	0.09 (0.50 → 0.59)	3

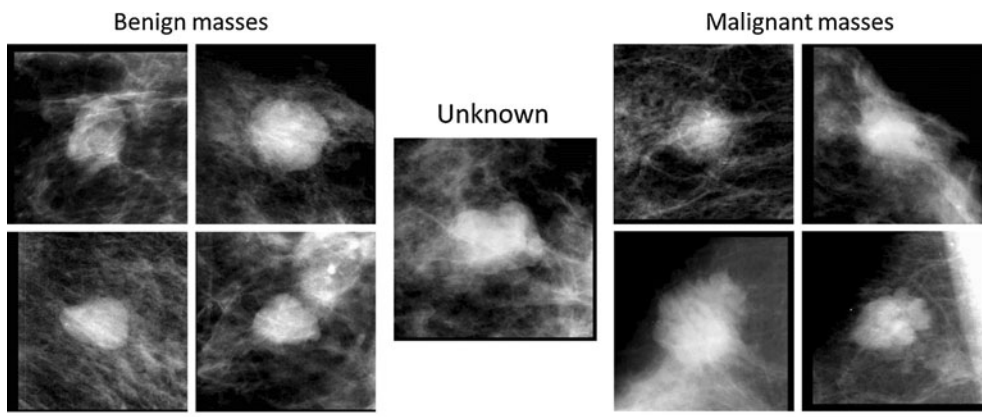

Fig. 5 A benign case in which the presentation of the reference images resulted differently for the observers in the US and those in Japan

	Average change	Number of observers who changed
Observers in the US	-0.10 (0.39 → 0.48)	-5
Observers in Japan	0.02 (0.33 → 0.31)	1

frequently filed, accounting for more than 60 % of cases; no number was provided for radiology. Although it is difficult to compare these statistics, it can be conjectured that US radiologists tend to practice more defensive medicine. The difference is also seen in the diagnostic assessment of probably benign lesions. In the US, cases assessed as BI-RADS 3, “probably benign finding”, should have less than 2 % risk of malignancy, whereas a breast imaging guideline in Japan was created on the basis of the BI-RADS, and cases assessed as category 3 “benign but malignancy can’t be ruled out” may have about a 2–10 % chance of cancer. These facts may explain the tendency of the US observers to give slightly higher ratings than those in Japan.

There were some differences in the years of experience between the two groups. It is difficult to determine whether the different reaction to the benign cases could be due to the years of experience, because US attending radiologists had a tendency to make only small changes in confidence levels. Note that the years of experience is one index; US attending radiologists and fellows only practice in breast imaging section in their routine work, whereas Japanese radiologists, although experts in breast image reading, may also read images of other organs, and surgeons may spend limited time in image reading. In addition, Japanese observers work at several different clinical facilities, and their practice may be somewhat different. The population of the test cases which were obtained in the US and primarily included Caucasians and African Americans may have affected the performance. Although the average years of experience and their background were different between two groups, the mean AUCs without and with similar images were comparable.

One of the limitations in this study was that we excluded difficult cases from the test dataset, which resulted in high AUCs both without and with similar images. Because of the limitation of time, the population of test cases selected for the observer performance studies were generally different from the clinical population. In this study, we excluded difficult cases because it was believed that CAD likely has no impact or detrimental effect on such cases. When a benign lesion looks very similar to typical malignant cases, a computer likely selects similar malignant lesions and outputs a high likelihood of malignancy. Even if a computer provided a low likelihood of malignancy, it is unlikely that radiologists would change their initial decision. Although we believe that such atypical cases are relatively rare, we did not include them in the present study because the number of the study cases was limited. As a result, the impact of the overall beneficial effect observed in this study could be much smaller in an actual clinical population. On the other hand, the high AUCs without similar images might have decreased the chances of gain.

5 Conclusion

The results of the observer studies indicate a potential utility of presenting reference images in the distinction between benign and malignant masses on mammograms by physicians. The overall effects in terms of the mean AUC were comparable for the observers in both counties. However, there was a slight difference in the reactions by the observers for some benign cases. This difference could be due to the differences in the patient population and the diagnostic environment in the two countries. In this study, the similarity measures used for the selection of reference images were based on the subjective similarity ratings determined by breast radiologists who practice in the US. Although subjective similarities noted by different groups of observers were expected to be comparable for most of the cases, there could be some differences in the impression due to the diagnostic environment. For improving the utility of computer-aided diagnosis systems, it may be worthwhile to investigate the development of a customized CAD system with an effective image selection scheme for physicians in different socio-clinical environments.

Acknowledgments This study was partly supported by the US Department of Defense Breast Cancer Research Program and Grants-in-Aid for Scientific Research in Japan. The authors are grateful to the physicians who participated in the observer studies.

References

1. Matsuda T, Marugame T, Kamo KI, Katanoda K, Ajiki W, Sobue T, The Japan Cancer Surveillance Research Group. Cancer incidence and incidence rates in Japan in 2006: based on data from 15 population-based cancer registries in the Monitoring of Cancer Incidence in Japan (MCIJ) Project. Jpn J Clin Oncol. 2011;42:139–47.
2. Americal Cancer Society. Cancer facts & figures 2012. Atlanta: American Cancer Society; 2012.
3. Ferlay J, Autier P, Boniol M, Heanue M, Colombet M, Boyle P. Estimates of the cancer incidence and mortality in Europe in 2006. Ann Oncol. 2007;18:581–92.
4. Tabar L, Fagerberg G, Duffy SW, Day NE, Gad A, Grontoft O. Update of the Swedish two-county program of mammographic screening for breast cancer. Radiol Clin North Am. 1992;30: 187–210.
5. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Selection, follow-up, and analysis in the health insurance plan study: a randomized trial with breast cancer screening. J Natl Cancer Inst Monogr. 1985;67:65–74.
6. Humphrey LL, Helfand M, Chan BKS, Woolf SH. Breast cancer screening: a summary of the evidence for the U.S. preventive services task force. Ann Intern Med. 2002;137:E-347–67.
7. Doi K, Giger ML, MacMahon H. Computer-aided diagnosis: development of automated schemes for quantitative analysis of radiographic images. Smin Ultrasound CT MRI. 1992;13:140–52.
8. Vyborny CJ. Can computers help radiologists read mammograms? Radiology. 1994;191:315–7.
9. Giger ML, Huo Z, Kupinski MA, Vyborny CJ. Computer-aided diagnosis in mammography. In: Fitzpatrick JM, Sonka M, editors. The handbook of medical imaging, medical imaging processing and analysis, vol 2. SPIE;2000. p. 915–1004.
10. Doi K. Diagnostic imaging over the last 50 year: research and development in medical imaging science and technology. Phys Med Biol. 2006;51:R5–27.
11. Chan HP, Doi K, Vyborny CJ, Schmidt RA, Metz CE, Lam KL, Ogura T, Wu Y, MacMahon H. Improvement in radiologists' detection of clustered microcalcifications on mammograms. Invest Radiol. 1990;25:1102–10.
12. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. Radiology. 2001;220:781–6.
13. Birdwell RL, Bandodkar P, Ikeda DM. Computer-aided detection with screening mammography in a university hospital setting. Radiology. 2005;236:451–7.
14. Cupples TE, Cunningham JE, Reynolds JC. Impact of computer-aided detection in a regional screening mammography program. Am J Roentgenol. 2005;185:944–50.
15. Morton MJ, Whaley DH, Brandt KR, Amrami KK. Screening mammograms; interpretation with computer-aided detection—prospective evaluation. Radiology. 2006;239:375–83.
16. Dean JC, Ilvento CC. Improved cancer detection using computer-aided detection with diagnostic and screening mammography: prospective study of 104 cancers. Am J Roentgenol. 2006;187: 20–8.
17. Chan HP, Sahiner B, Roubidoux MA, Wilson TE, Adler DD, Paramagul C, Newman JS, Sanjay-Gopal S. Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: an ROC study. Radiology. 1999;212: 817–27.
18. Huo Z, Giger ML, Vyborny CJ, Metz CE. Breast cancer: effectiveness of computer-aided diagnosis—observer study with independent database of mammograms. Radiology. 2002;224: 560–8.
19. Jiang Y, Nishikawa RM, Schmidt RA, Metz CE, Giger ML, Doi K. Improving breast cancer diagnosis with computer-aided diagnosis. Acad Radiol. 1999;6:22–33.
20. Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imaging Graph. 2007;31:198–211.

21. Swett HA, Fisher PR, Cohn AI, Miller PL, Mutalik PG. Expert system-controlled image display. Radiology. 1989;172:487–93.
22. Aisen AM, Broderick LS, Winer-Muram H, Brodley CE, Kak AC, Pavlopoulou C, Dy J, Shyu CR, Marchiori A. Automated storage and retrieval of thin-section CT images to assist diagnosis: system description and preliminary assessment. Radiology. 2003;228:265–70.
23. Li Q, Li F, Shiraishi J, Katsuragawa S, Sone S, Doi K. Investigation of new psychophysical measures for evaluation of similar images on thoracic CT for distinction between benign and malignant nodules. Med Phys. 2003;30:2584–93.
24. Kawata Y, Niki N, Ohmatsu H, Moriyama N. Example-based assisting approach for pulmonary nodule classification in three-dimensional thoracic computed tomography images. Acad Radiol. 2003;10:1402–15.
25. Swett HA, Mutalik PG, Neklesa VP, Horvath L, Lee C, Richter J, Tocino I, Fisher P. Voice-activated retrieval of mammography reference images. J Digit Imaging. 1998;11:65–73.
26. Qi H, Snyder WE. Content-based image retrieval in picture archiving and communications systems. J Digit Imaging. 1999; 12:81–3.
27. Sklansky J, Tao EY, Bazargan M, Ornes CJ, Murchison RC, Teklehaimanot S. Computer-aided, case-based diagnosis of mammographic regions of interest containing microcalcifications. Acad Radiol. 2000;7:395–405.
28. Giger ML, Huo Z, Vyborny CJ, Lan L, Bonta I, Horsch K, Nishikawa RM, Rosenbourgh I. Intelligent CAD workstation for breast imaging using similarity to known lesions and multiple visual prompt aids. Proc SPIE. 2002;4684:768–73.
29. Muramatsu C, Li Q, Schmidt RA, Suzuki K, Shiraishi J, Newstead GM, Doi K. Experimental determination of subjective similarity for pairs of clustered microcalcifications on mammograms: observer study results. Med Phys. 2006;33:3460–8.
30. Muramatsu C, Li Q, Schmidt RA, Shiraishi J, Suzuki K, Newstead GM, Doi K. Determination of subjective similarity for pairs of masses and pairs of clustered microcalcifications on mammograms: comparison of similarity ranking scores and absolute similarity ratings. Med Phys. 2007;34:2890–5.
31. Muramatsu C, Li Q, Suzuki K, Schmidt RA, Shiraishi J, Newstead G, Doi K. Investigation of psychophysical measure for evaluation of similar images for mammographic masses: preliminary results. Med Phys. 2005;32:2295–304.
32. Muramatsu C, Li Q, Schmidt RA, Shiraishi J, Doi K. Determination of similarity measures for pairs of mass lesions on mammograms by use of BI-RADS lesion descriptors and image features. Acad Radiol. 2009;16:443–9.
33. Muramatsu C, Li Q, Schmidt RA, Shiraishi J, Doi K. Investigation of psychophysical similarity measures for selection of similar images in the diagnosis of clustered microcalcifications on mammograms. Med Phys. 2008;35:5695–702.
34. Horsch K, Giger ML, Vyborny CJ, Lan L, Mendelson EB, Hendrick ER. Classification of breast lesions with multimodality computer-aided diagnosis: observer study results on an independent clinical data set. Radiology. 2006;240:357–68.
35. Muramatsu C, Schmidt RA, Shiraishi J, Li Q, Doi K. Presentation of similar images as a reference for distinction between benign and malignant masses on mammograms: analysis of initial observer study. J Digit Imaging. 2010;23:592–602.
36. Heath M, Bowyer K, Kopans D, Moore R, Kedelmeyer P. Current states of the digital database for screening mammography. Digital mammography. Dordrecht: Kluwer; 1998.
37. 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.
38. Physician Insurers Association of America. PIAA 2002 breast cancer study. Rockville: Physician Insurers Association of America; 2002.
39. Dick JF III, Gallagher TH, Brenner RJ, Yi JP, Reisch LM, Abraham L, Miglioretti DL, Carney PA, Cutter GR, Elmore JG. Predictors of radiologists' perceived risk of malpractice lawsuits in breast imaging. Am J Roentgenol. 2009;192:327–33.