

# Presentation of similar images for diagnosis of breast masses on mammograms: analysis of the effect on residents

Chisako Muramatsu\*<sup>ab</sup>, Robert A. Schmidt<sup>a</sup>, Junji Shiraishi<sup>a</sup>,  
Qiang Li<sup>c</sup>, Hiroshi Fujita<sup>b</sup>, Kunio Doi<sup>a</sup>

<sup>a</sup>Dept. of Radiology, The University of Chicago, 5841 S Maryland Ave, Chicago, IL, USA 60637;

<sup>b</sup>Dept. of Intelligent Image Information, Gifu University, 1-1 Yanagido, Gifu, Japan 501-1194;

<sup>c</sup>Dept. of Radiology, Duke University, 2424 Erwin Rd., Durham, NC, USA 27705

## ABSTRACT

We have been developing a computerized scheme for selecting visually similar images that would be useful to radiologists in the diagnosis of masses on mammograms. Based on the results of the observer performance study, the presentation of similar images was useful, especially for less experienced observers. The test cases included 50 benign and 50 malignant masses. Ten observers, including five breast radiologists and five residents, were asked to provide the confidence level of the lesions being malignant before and after the presentation of similar images. By use of multi-reader, multi-case receiver operating characteristic analysis, the average areas under the curves for the five residents were 0.880 and 0.896 without and with similar images, respectively ( $p=0.040$ ). There were four malignant cases in which the initial ratings were relatively low, but the similar images alerted the residents to increase their confidence levels of malignancy close to those by the breast radiologists. The presentation of similar images may cause some observers falsely to increase their suspicion for some benign cases; however, if similar images can alert radiologists to recognize the signs of malignancy and also help them to decrease their suspicion correctly for some benign cases, they can be useful in the diagnosis on mammograms.

**Key words:** computer-aided diagnosis, similar images, mammography, breast masses, observer study

## 1. INTRODUCTION

Breast cancer is one of the leading causes of cancer death and frequently diagnosed cancer in women in the United States, European countries, and Japan [1, 2]. Regular screening by mammography is considered effective for detecting the cancer early and, therefore, reducing the mortality rate [3-5]. However, diagnosis of breast cancer on mammograms can be difficult. Some cancers are misinterpreted, and many benign lesions are suspected to be malignant findings [6-8].

With the hope of helping radiologists in improving the diagnostic accuracy by providing useful information, computerized analysis of breast lesions on mammograms has been studied by many investigators [9-17]. The results of observer performance studies indicated the potential usefulness of the presentation of the likelihood of malignancy of lesions in radiologists' classification of benign and malignant masses [18, 19] and clustered microcalcifications [20].

Radiologists' diagnostic skills are usually based on the observation of numerous cases in their clinical experience. It may be assumed that experienced radiologists have large image libraries in their brain. Less experienced radiologists, on the other hand, may sometimes consult textbooks or past cases. If similar cases are automatically selected from the database and presented as a reference, they may be useful to radiologists when they diagnose a new, unknown case [21-23]. In fact, some studies [24-27] indicate that radiologists' diagnostic performance can be improved by the presentation of similar images. Several groups [24-32] have investigated image retrieval methods, most of which are based on the distance in image feature space.

For increasing the utility of similar images, we believe that presented images must be really similar from the point of view of diagnosing radiologists. We have been investigating radiologists' impression of similarity for pairs of lesions on radiological images [33-35] for selection of visually similar images. In our previous studies [36-38], by use of subjective

\*chisa@fjt.info.gifu-u.ac.jp; phone +81-58-230-6519; fax +81-58-230-6514; www.fjt.info.gifu-u.ac.jp

similarity ratings by breast radiologists, the psychophysical similarity measures for pairs of breast masses on mammograms were determined, which had a relatively high correlation with the radiologists' similarity ratings. In this study, the effect of the presentation of similar images on the distinction between malignant and benign masses on mammograms was evaluated in an observer performance study.

## 2. MATERIAL AND METHODS

### 2.1 Image database of breast masses

Images of breast masses used in this study were obtained from the Digital Database for Screening Mammography developed by the University of South Florida [39]. For each biopsy-confirmed mass, a five-centimeter square region of interest (ROI) was obtained. Some ROIs were excluded for this study, including those on images with poor quality, large masses which do not fit in a 5-cm square, partial masses, findings of architectural distortion or asymmetry, and masses with microcalcifications which may influence radiologists' judgment. The database included 1568 ROIs with 728 malignant masses and 840 benign masses.

For each mass, a manual outline was provided by one of two radiologists for determination of image features. The contrast and density level of the images were adjusted visually by a breast radiologist for facilitating image viewing and feature determination.

### 2.2 Observer performance study

An observer study was conducted for evaluation of the presentation of similar images in the distinction between malignant and benign masses. The results of our preliminary study suggested that the presentation of similar images would not be useful for atypical cases, such as spiculated benign masses ("malignant-looking" benign masses) and well-circumscribed malignant masses ("benign-looking" malignant masses). These benign and malignant masses are generally considered incorrectly as having a very high and very low likelihood of malignancy, respectively, by radiologists. The presentation of similar images would naturally make them more confident of their decisions. However, the impact of the presentation of similar images on these cases may not be large, because radiologists are less likely to be influenced by the computer aid when they are initially confident of their decisions either correctly or incorrectly. We believe that the target cases for CAD are those in which radiologists may need help, i.e., indeterminate cases. Therefore, in order to evaluate the effect of similar images with a limited number of cases, we did not include these benign-looking malignant and malignant-looking benign cases in this study.

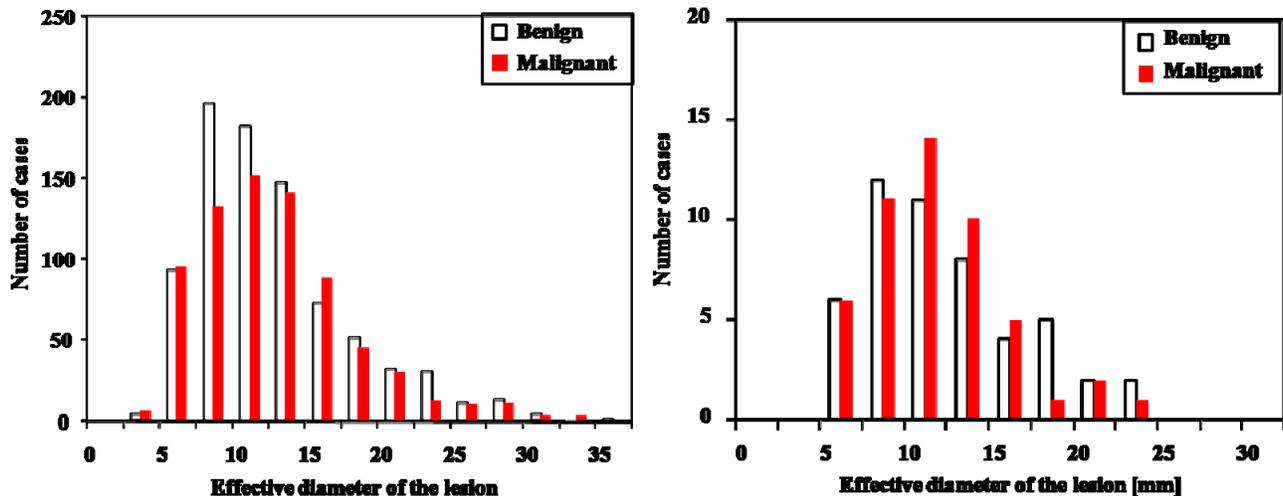


Fig. 1. Size distributions for benign and malignant masses in the database (left) and for those selected as unknown cases in this study (right). The database included a slightly larger number of small masses for the benign cases. The unknown cases were selected semi-randomly based on the original distributions.

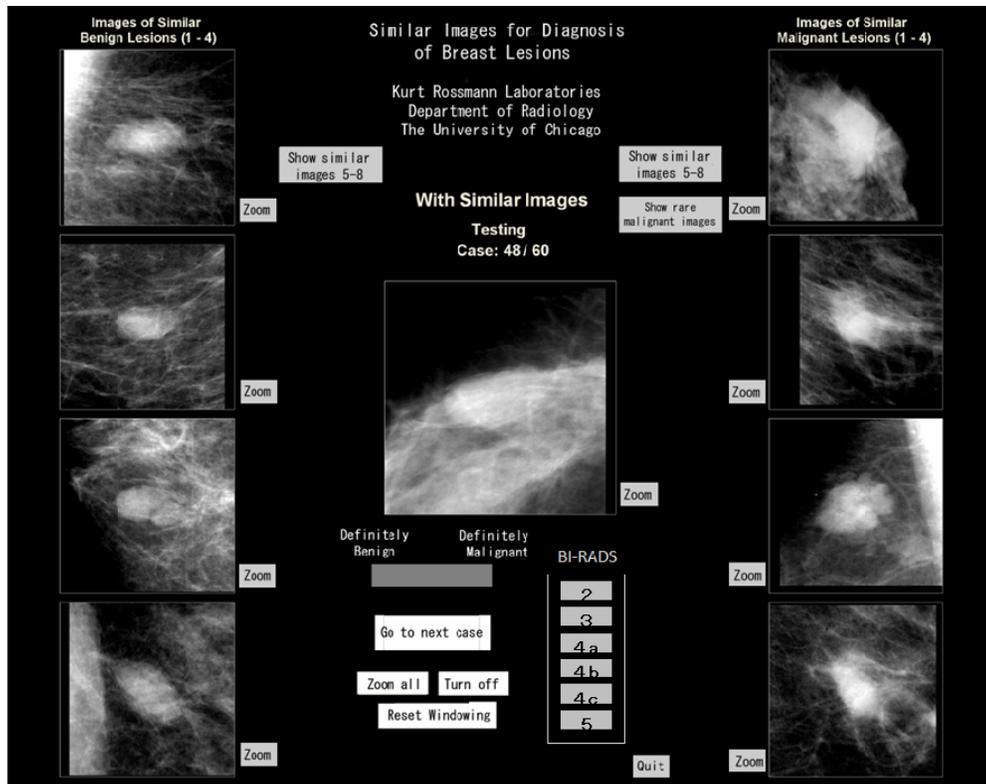


Fig. 2. Observer interface when reading with the similar images. The observers were first shown the center image (unknown image) alone. After the confidence level of malignancy was marked, the similar benign images (left) and malignant images (right) were presented.

In addition, if these benign-looking malignant and malignant-looking benign masses are presented as a reference, they can confuse radiologists, because an unknown lesion may be similar to both malignant and benign masses. Especially, radiologists are afraid to miss a cancer, and therefore, they might increase their suspicion if circumscribed malignant masses are shown. For avoiding confusing cases, benign-looking malignant and malignant-looking benign masses are also excluded from the reference database.

One hundred cases, including 50 malignant and 50 benign masses, were selected as unknown cases semi-randomly based on the size of lesions. The size distributions for malignant and benign masses in the database and for those selected as the unknown masses are shown in Fig. 1. Ten observers, including five breast imaging radiologists and five radiology residents with experience in mammography, participated in the study. Radiologists first provided the confidence level of a lesion being malignant on a continuous rating scale between “definitely benign” and “definitely malignant.” Immediately afterwards, four benign and four malignant similar images were presented, and the radiologists were asked to reconsider their initial confidence levels. These similar images were selected based on the size of lesions and the psychophysical similarity measure [37]. Figure 2 shows the observer study interface for reading by radiologists with the benign similar images (left) and the malignant similar images (right). Another four similar images could be shown if the observers wanted to see additional images. The observers were able to magnify images if they wished. The reading time was not limited.

The results were evaluated by use of multi-reader, multi-case [40] receiver operating characteristic (ROC) analysis with a computer program (Barbaum KS, the University of Iowa [41], and Metz CE, the University of Chicago [42]). The areas under the ROC curves (AUCs) were also determined for the results by a group of five breast radiologists and by a group of five residents. The results of the two groups were compared, and the differences are discussed in this paper.

### 3. RESULTS

The AUCs for distinction between malignant and benign masses without and with the presentation of similar images by each observer and by the groups of observers are summarized in Table 1. For all observers, the average AUC with similar images increased slightly from that without similar images, and the difference was found to be statistically significant ( $p = 0.047$ ). The improvement was larger for the five residents. In fact, the average AUCs without and with similar images for the five breast radiologists are almost comparable. The AUCs for the breast radiologists were already high before the presentation of similar images.

The relationships between the average initial confidence level of malignancy of the lesions and the average change in the confidence levels before and after the presentation of similar images by the groups of breast radiologists and residents are shown in Figs. 3(a) and 3(b), respectively. If the changes were in the directions of correct diagnoses, they were considered as beneficial (positive) changes, whereas the negative changes correspond to detrimental changes which were in the direction of incorrect diagnoses. In this study, there were no large changes in confidence levels by the breast radiologists. On the other hand, the presentation of similar images was useful for many malignant cases for the residents. There were four malignant cases, marked as i, ii, iii, and iv in Fig. 3, which the residents initially marked as having relatively low confidence of malignancy; however, some of them increased their suspicion after the presentation of similar images. The average confidence levels for cases i, ii, iii, and iv by the residents were initially 0.24, 0.39, 0.47, and 0.50, respectively, and increased to 0.33, 0.55, 0.59, and 0.62, respectively, after the presentation of similar images. Some of their confidence levels approached the average confidence levels by the breast radiologists, which were 0.39, 0.58, 0.76, and 0.64, respectively.

Figure 4 shows the unknown (center) masses and the similar benign (left) and malignant (right) masses presented in the observer study. They correspond to cases iii and iv in Fig. 3, in which the presentation of similar images was useful to the majority of the residents. It can be predicted that some residents might have recognized the spiculations in these masses when the similar malignant masses were presented.

Table 1. AUCs by each observer and average AUCs with 95% confidence intervals by groups of observers. The p-values for the differences in AUCs without and with similar images are also shown.

		Without similar images	With similar images
Breast radiologist	A	0.951	0.962
	B	0.972	0.978
	C	0.940	0.938
	D	0.947	0.942
	E	0.942	0.943
Average by breast radiologists		0.951 [0.925, 0.976]	0.952 [0.927, 0.978] ( $p = 0.59$ )
Radiology resident	F	0.906	0.931
	G	0.874	0.877
	H	0.887	0.913
	I	0.871	0.874
	J	0.863	0.885
Average by radiology residents		0.880 [0.835, 0.925]	0.896 [0.848, 0.944] ( $p = 0.040$ )
Average for all		0.915 [0.873, 0.957]	0.924 [0.885, 0.963] ( $p = 0.047$ )

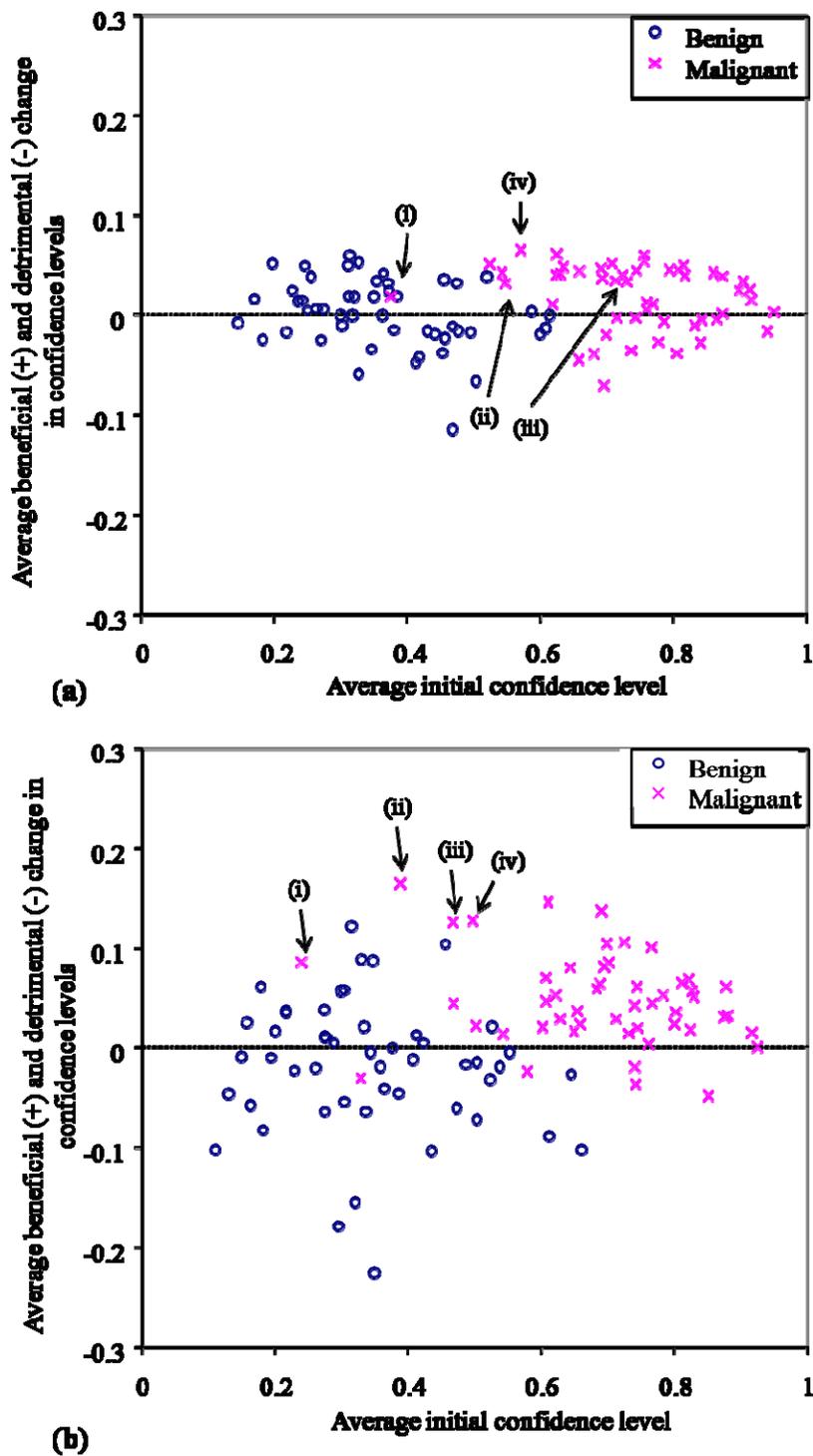


Fig. 3. Relationships between the average initial confidence levels of malignancy of the lesions and the average changes in the confidence levels before and after the presentation of similar images (a) by the five breast radiologists and (b) by the five residents. The positive change corresponds to a beneficial change, which was in the direction of correct diagnosis, whereas the negative change corresponds to a detrimental change, which was in the direction of incorrect diagnosis.

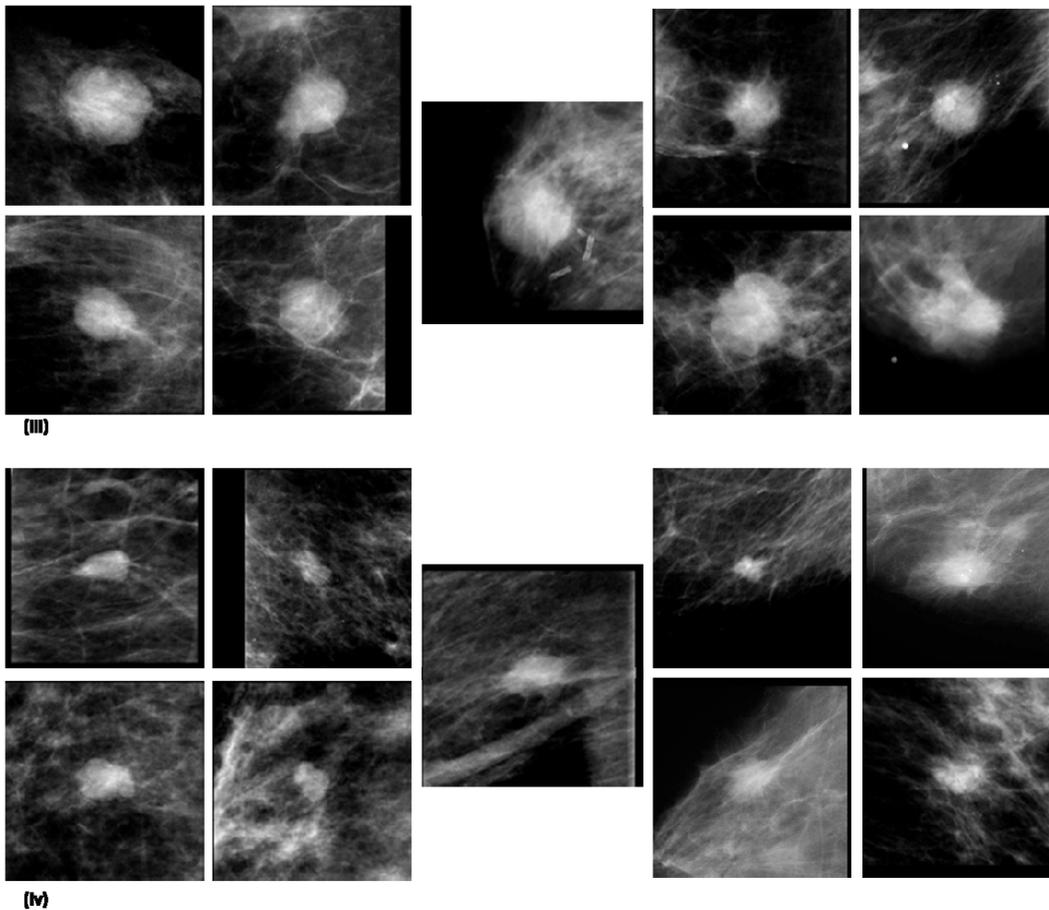


Fig. 4. Unknown masses (center) and the similar benign masses (left) and malignant masses (right) presented in the observer study. These image sets correspond to (iii) and (iv) in Fig. 4. The presentation of similar images caused some residents to increase their confidence levels correctly for these malignant masses.

#### 4. CONCLUSION

The results of this observer study indicated that the presentation of similar images can be useful in the diagnosis of breast masses on mammograms, especially for radiologists with limited experience. The presentation of similar images may alert radiologists and make them recognize some important signs. Similar images may affect some observers falsely to increase their suspicion on benign masses; however, if similar images bring beneficial effects to most cancer cases and to some benign cases, such a CAD system would be useful in the diagnosis of mammograms.

#### ACKNOWLEDGEMENT

The authors are grateful to the following radiologists for their participation in the observer study: H. Abe, M.D., Ph.D., C. Sennett, M.D., A. Shimauchi, M.D., S. Makim, M.D., S. Sung, M.D., J. Carabetta, M.D., S., Konda, M.D., M. Lester, M.D., S. Thiel, M.D., and C. Valentin, M.D.. This study was supported by the Department of Defense Breast Cancer Research Program. The CAD technologies developed in the Kurt Rossmann Laboratory have been licensed to companies including R2 Technology, Riverain Medical, Mitsubishi Space Software Company, General Electric Corporation, Median Technologies, and Toshiba Corporation.

## REFERENCES

- [1] American Cancer Society, *Cancer Facts and Figures 2008*, American Cancer Society, Atlanta, 2008.
- [2] Tyczynski, J. E., Bray, F., and Parkin, D. M., "Breast cancer in Europe," European Network of Cancer Registries, Cedex, France, 2002.
- [3] Tabar, L., Fagerberg, G., Duffy, S. W., Day, N. E., Gad, A., and Grontoft, O., "Update of the Swedish two-county program of mammographic screening for breast cancer," *Radiol. Clin. North Am.* 30, 187-210 (1992).
- [4] Shapiro, S., Venet, W., Strax, P., Venet, L., and 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.* 67, 65-74 (1985).
- [5] Humphrey, L. L., Helfand, M., Chan, B.K.S., and Woolf, S. H., "Breast cancer screening: a summary of the evidence for the U.S. preventive services task force," *Annals. Internal Medicine* 137, E347-367 (2002).
- [6] Kopans, D. B., Moore, R. H., McCarthy, K. A., Hall, D. A., Hulka, C. A., Whitman, G. J., Slanetz, P. J., and Halpern, E. F., "Positive predictive value of breast biopsy performed as a result of mammography: there is no abrupt change at age 50 years," *Radiology* 200, 357-360 (1996).
- [7] Gur, D., Wallace, L. P., Klym, A. H., Hardesty, L. A., Abrams, G. S., Shah, R., and Sumkin, J. H., "Trends in recall, biopsy, and positive biopsy rates for screening mammography in an academic practice," *Radiology* 235, 396-401 (2005).
- [8] Rosenberg, R. D., Yankaskas, B. C., Abraham, L. A., Sickles, E. A., Lehman, C. D., Geller, B. M., Carney, P. A., Kerlikowske, K., Buist, D. S. M., Weaver, D. L., Barlow, W. E., and Ballard-Barbash, R., "Performance benchmarks for screening mammography," *Radiology* 241, 55-66 (2006).
- [9] Li, H., Giger, M. L., Yuan, Y., Chen, W., Horsch, K., Lan, L., Jamieson, A. R., Sennett, C. A., and Jansen, S. A., "Evaluation of computer-aided diagnosis on a large clinical full-field digital mammographic dataset," *Acad. Radiol.* 15, 1437-1445 (2008).
- [10] Mu, T., Nandi, A. K., and Rangayyan, R. M., "Classification of breast masses using selected shape, edge-sharpness, and texture features with linear and kernel-based classifiers," *J. Digit. Imaging* 21, 153-169 (2008).
- [11] Delogu, P., Fantacci, M. E., Kasae, P., and Retico, A., "Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier," *Comput. Biol. Med.* 10, 1479-1497 (2007).
- [12] Buchbinder, S. S., Leichter, I. S., Lederman, R. B., Novak, B., Bamberger, P. N., Sklair-Levy, M., Yarmish, G., and Fields, S. I., "Computer-aided classification of BI-RADS category 3 lesions," *Radiology* 230, 820-823 (2004).
- [13] Hadjiiski, L., Sahiner, B., Chan, H. P., Petrick, N., Helvie, M. A., and Gurcan, M., "Analysis of temporal changes of mammographic features: computer-aided classification of malignant and benign breast masses," *Med. Phys.* 28, 2309-2317 (2001).
- [14] Rana, R. S., Jiang, Y., Schmidt, R. A., Nishikawa, R. M., and Liu, B., "Independent evaluation of computer classification of malignant and benign calcifications in full-field digital mammograms," *Acad. Radiol.* 4, 363-370 (2007).
- [15] Nakayama, R., Uchiyama, Y., Watanabe, R., Katsuragawa, S., Namba, K., and Doi, K., "Computer-aided diagnosis scheme for histological classification of clustered microcalcifications on magnification mammograms," *Med. Phys.* 31, 789-799 (2004).
- [16] Veldkamp, W. J. H., Karssemeijer, N., Otten, J. D. M., and Hendricks, J. H. C. L., "Automated classification of clustered microcalcifications into malignant and benign types," *Med. Phys.* 27, 2600-2608 (2000).
- [17] Chan, H. P., Sahiner, B., Lam, D. L., Petrick, N., Helvie, M. A., Goodsitt, M. M., and Adler, D. D., "Computerized analysis of mammographic microcalcifications in morphological and texture feature space," *Med. Phys.* 25, 2007-2019 (1998).
- [18] Huo, Z., Giger, M. L., Vyborny, C. J., and Metz, C. E., "Breast cancer: effectiveness of computer-aided diagnosis – observer study with independent database of mammograms," *Radiology* 224, 560-568 (2002).
- [19] Chan, H. P., Sahine, B., Roubidoux, M. A., Wilson, T. E., Adler, D. D., Paramagul, C., Newman, J. S., and Sanjay-Gopal, S., "Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: an ROC study," *Radiology* 212, 817-827 (1999).
- [20] Jiang, Y., Nishikawa, R. M., Schmidt, R. A., Metz, C. E., Giger, M. L., and Doi, K., "Improving breast cancer diagnosis with computer-aided diagnosis," *Acad. Radiol.* 6, 22-23 (1999).
- [21] Doi, K., "Current status and future potential of computer-aided diagnosis in medical imaging," *Br. J. Radiol.* 78, S3-19 (2005).

- [22] Doi, K., "Computer-aided diagnosis in medical imaging: historical review, current status, and future potential," *Comput. Med. Imaging. Graph.* 31, 198-211 (2007).
- [23] Muramatsu, C., Investigation of similarity measures for selection of similar images in computer-aided diagnosis of breast lesions on mammograms, Ph.D. Dissertation, The University of Chicago, Chicago, IL, ProQuest/UMI (2008).
- [24] Horsch, K., Giger, M. L., Vyborny, C. J., Lan, L., Medelson, E. B., and Hendrick, E. R., "Classification of breast lesions with multimodality computer-aided diagnosis: observer study results on an independent clinical data set," *Radiology* 240, 357-368 (2006).
- [25] Li, Q., Li, F., Shiraishi, J., Katsuragawa, S., Sone, S., and Doi, K., "Investigation of new psychophysical measures for evaluation of similar images on thoracic CT for distinction between benign and malignant nodules," *Med. Phys.* 30, 2584-2593 (2003).
- [26] Aisen, A. M., Broderick, L. S., Winer-Muram, H., Brodley, C. E., Kak, A. C., Pavlopoulou, C., Dy, J., Shyu, C. R., and Marchiori, A., "Automated storage and retrieval of thin-section CT images to assist diagnosis: system description and preliminary assessment," *Radiology* 228, 265-270 (2003).
- [27] Sklansky, J., Tao, E. Y., Bazargan, M., Ornes, C. J., Murchison, R. C., and Teklehaimanot, S., "Computer-aided, case-based diagnosis of mammographic regions of interest containing microcalcifications," *Acad. Radiol.* 7, 295-405 (2000).
- [28] Zheng, B., Mello-Thoms, C., Wang, X. H., Abrams, G. S., Sumkin, J. H., Chough, D. M., Ganott, M. A., Lu, A., and Gur, D., "Interactive computer-aided diagnosis of breast masses: computerized selection of visually similar image sets from a reference library," *Acad. Radiol.* 14, 917-927 (2007).
- [29] El-Naqa, I., Yang, Y., Galatsanos, N. P., Nishikawa, R. M., and Wernick, M. N., "A similarity learning approach to content-based image retrieval: application to digital mammography," *IEEE Trans. Med. Imaging* 23, 1233-1244 (2004).
- [30] Kawata, Y., Niki, N., Ohmatsu, H., and Moriyama, N., "Example-based assisting approach for pulmonary nodule classification in three-dimensional thoracic computed tomography images," *Acad. Radiol.* 10, 1402-1415 (2003).]
- [31] Qi, H., and Snyder, W. E., "Content-based image retrieval in picture archiving and communications systems," *J Digit. Imaging* 12, 81-83 (1999).
- [32] Swett, H. A., Fisher, P. R., Cohn, A. I., Miller, P. L., and Mutalik, P. G., "Expert system-controlled image display," *Radiology* 172, 487-493 (1989).
- [33] Kumazawa S., Muramatsu, C., Li, Q., Li, F., Shiraishi, J., Caligiuri, P., Shmidt, R. A., MacMahon, H., and Doi, K., "An investigation of radiologists' perception of lesion similarity: observations with paired breast masses on mammograms and paired lung nodules on CT images," *Acad. Radiol.* 15, 887-894 (2008).
- [34] Muramatsu, C., Li, Q., Schmidt, R. A., Shiraishi, J., Suzuki, K., Newstead, G. M., and 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* 34, 2890-2895 (2007).
- [35] Muramatsu, C., Li, Q., Schmidt, R. A., Suzuki, K., Shiraishi, J., Newstead, G. M., and Doi, K., "Experimental determination of subjective similarity for pairs of clustered microcalcifications on mammograms: Observer study results," *Med Phys* 33, 3460-3468 (2006).
- [36] Muramatsu, C., Li, Q., Schmidt, R. A., Shiraishi, J., and 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.* (in press).
- [37] Muramatsu, C., Li, Q., Schmidt, R. A., Suzuki, K., Shiraishi, J., Newstead, G. M., and Doi, K., "Determination of subjective and objective similarity for pairs of masses on mammograms for selection of similar images," *Proc SPIE* 6514, 65141I 1-65141I9 (2007).
- [38] Muramatsu, C., Li, Q., Suzuki, K., Schmidt, R. A., Shiraishi, J., Newstead, G. M., and Doi, K., "Investigation of psychophysical measure for evaluation of similar images for mammographic masses: Preliminary results," *Med. Phys.* 32, 2295-2304 (2005).
- [39] Heath, M., Bowyer, K., Kopans, D., Moore, R., Kegelmeyer, W. P., The digital database for screening mammography, in *Proceedings of the Fifth International Workshop on Digital Mammography*, Yaffe, M. J., ed. Medical Physics Publishing, 212-218 (2001).
- [40] Dorfman, D. D., Berbaum, K. S., Metz, C.E., "Receiver operating characteristic rating analysis: Generalization to the population of readers and patients with the jackknife method," *Invest. Radiol.* 27, 723-731 (1992).
- [41] Medical Image Perception Laboratory, The University of Iowa. (<http://perception.radiology.uiowa.edu>)
- [42] Kurt Rossmann Laboratory for Radiologic Image Research, The University of Chicago. (<http://xray.bsd.uchicago.edu/krl>)