A Comparative Study of the Inter-reader Variability of Breast Percent Density Estimation in Digital Mammography: Potential Effect of Reader's Training and Clinical Experience

Emily F. Conant[1], Diane Li[1], Sara Gavenonis[1], Predrag R. Bakic[2], Ann-Katherine Carton[2], Cuiping Zhang[2], Andrew D.A. Maidment[2], and Despina Kontos[1]

[1] Hospital of the University of Pennsylvania, Breast Imaging Division, 1 Silverstein Building HUP, 3400 Spruce St., Philadelphia PA 19104-4206
[2] University of Pennsylvania, Department of Radiology, Physics Section, 1 Silverstein Building HUP, 3400 Spruce St., Philadelphia PA 19104-4206
{Emily.Conant,Diane.Li,Sara.Gavenonis,Predrag.Bakic, Ann-Katherine.Carton,Cuiping.Zhang, Andrew.Maidment,Despina.Kontos}@uphs.upenn.edu

Abstract. The variability of breast percent density (PD%) estimation from digital mammography (DM) images was evaluated using measurements from readers with different training and clinical experience. Post-processed DM images (*PremiumView*™, GE Healthcare) from 40 women were analyzed. Breast PD% estimation was performed using the *Cumulus* software (*Ver. 4.0, Univ. Toronto*). Two groups of readers were considered, one with clinical (*i.e.,* radiologists) and one with non-clinical training (*i.e.,* physicists). Consistency of PD% was analyzed using the Pearson correlation coefficient (r) and ANOVA. Inter-reader agreement was higher among clinical (r=0.91, p<0.001), than non-clinical readers (r=0.83, p<0.001). Intra-reader consistency after repeated reads was on average equally high for both groups (r=0.91, p<0.001). Our results suggest that the reader's experience and training has an effect on the obtained PD% measures. The higher correlation among the clinically trained readers could be attributed to their extensive exposure to post-processed DM images and their knowledge of breast anatomy.

Keywords: Digital mammography, breast percent density, breast cancer risk estimation.

1 Introduction

It is well known that the sensitivity of mammography in detecting breast cancer decreases as parenchymal density increases. There is also growing evidence that suggests that breast density is an independent risk factor for breast cancer [1]. Breast imaging radiologists incorporate a density estimate in all clinical mammographic reports using the 4-tiered Breast Imaging Reporting and Data System (BI-RADS) system and they are very familiar with the wide variety of breast parenchymal patterns that are

J. Martí et al. (Eds.): IWDM 2010, LNCS 6136, pp. 114–120, 2010.

present in clinical mammographic images. Unfortunately, because the assignment of a BI-RADS density category is subjective, there is a high degree of inter-reader variability [2, 3]. Currently, the most widely used methods to quantify breast density rely on measures derived from mammograms using a semi-automated image thresholding method [4]. Most of the studies published to date have been performed using digitized screen-film mammograms [5]. Digital mammography (DM) is increasingly replacing screen-film mammography in breast cancer screening. DM imaging systems typically produced two-types of images; the "FOR PROCESSING" images, which are proportional to the x-ray attenuation (*i.e.,* raw data), and the "FOR PRESENTATION" images, which are post-processed according to vendor-specific image processing algorithms prior to presentation to the radiologist for diagnostic interpretation. Currently, the strategy adopted by most clinical breast imaging divisions is to archive only the post-processed (*i*.e., "FOR PRESENTATION") images due to storage and cost constraints. Therefore, studies are needed to determine the optimal approaches for utilizing the digital data for breast density estimation. Investigating the potential use of post-processed images in breast density estimation could result in a more widely-adopted translation of density-based breast cancer risk assessment in clinical practice. Currently, not many studies exist that have investigated this potential [6].

We performed a study to compare the inter-reader variability of area-based breast percent density (PD%) estimation in post-processed DM images performed by groups of readers with different training and clinical experience. Post-processed DM images were used in our study because they are the DM images that clinicians are most familiar with and are the images that best display breast anatomy. In addition, the post-processed images are often more widely available and accessible in clinical practices. The main goals of our study were to investigate *i*) if post-processed DM images can provide viable means for PD% estimation and *ii*) if the obtained measures are affected by the type of training and the clinical experience of the readers.

2 Methods

The MLO DM image from the contralateral (*i.e.,* unaffected) breast of 40 women with recently detected abnormalities and/or previously diagnosed breast cancer, recruited as part of a separate multi-modality imaging clinical trial that has been completed in our department[1], were retrospectively collected and analyzed under HIPAA and IRB approval. All women were study volunteers who had signed informed consent. DM imaging was performed with a GE Healthcare DS full-field DM system (GE Healthcare, Chalfont St. Giles, UK) [7-9]. Images were acquired with 0.1 mm/pixel resolution at 12 bit gray-level. Image post-processing was performed using *PremiumView*™ (GE Healthcare), an embedded adaptive histogram equalization method [8]. Area-based breast percent density (PD%) was estimated using the semi-automated image thresholding technique of *Cumulus* (*Ver. 4.0, Univ. Toronto*) [4]. Using *Cumulus* the image background and the pectoral muscle region are excluded from the breast density calculations and user-defined gray-level intensity thresholds are applied to outline the fibroglandular tissue regions within the breast. PD% is then computed as the percent of the breast occupied by fibroglandular tissue.

[1] Evaluation of Multimodality Breast Imaging, NIH P01 CA85484, PI: M.D. Schnall.

Two groups of readers were considered, one with clinical experience (*i.e.,* breast imaging radiologists) and one with non-clinical training (*i.e.,* medical physicists). Each group included three readers at different levels of experience (*i.e.,* years of training). The clinical readers had varying degrees of clinical experience of breast anatomy and digital mammographic images (*i.e.,* 3 months, 3 years and 20 years of experience). The non-clinical readers had knowledge of breast anatomy and varying degrees of training in medical physics (*i.e.,* 2, 12 and 14 years of experience). Both groups received the training outlined in the manual which accompanies the *Cumulus* software. Each group performed two rounds of readings on the same dataset.

To evaluate the consistency of the obtained PD% measures among readers in the same group (*i.e.,* inter-reader agreement) and between the repeated reads of each reader (*i.e.,* intra-reader agreement), pair-wise Pearson correlation coefficients were computed and analysis of variance (ANOVA) was performed. To compare the PD% measures between the groups of readers the Student's paired t-test was applied on the mean of the PD% estimates of the corresponding groups.

3 Results

Inter-reader correlation was high for both groups (Table 1), with the clinically trained group having a higher average inter-reader Pearson correlation (r=0.91, p<0.001) than the non-clinically trained group (r=0.82, p<0.001). Linear regression analysis also demonstrated lower variability and a stronger association between the PD% estimates obtained by the clinically trained readers than between the non-clinically trained readers (Fig. 1). Overall, the mean of the PD% estimates for the non-clinically trained group was statistically significantly higher (p<0.001) than that of the clinically trained group ($\text{mean}_{\text{Read1}}$=27%, $\text{mean}_{\text{Read2}}$=24%), both for the first ($\text{mean}_{\text{Read1}}$=30%) and the second ($\text{mean}_{\text{Read2}}$=35%) reads. ANOVA showed that the obtained breast PD% measures were more consistent among the clinically trained readers than among the non-clinically trained readers (Fig. 2). The means of the PD% estimates were not statistically significantly different among the clinically trained readers (p_{ANOVA1}=0.39, p_{ANOVA2}=0.05), but statistically significantly different among the non-clinically trained readers (p_{ANOVA1}<0.001, p_{ANOVA2}<0.001). Intra-reader agreement after repeated reads was on average equally high (r=0.90, p<0.001) for both groups (Table 2).

Table 1. Pair-wise Pearson correlations (r) for inter-reader variability in breast PD% estimates between readers in each group. Readers are ordered in increasing order of experience.

	Pearson correlation coefficients (r) for PD% inter-reader variability							
	Clinically Trained Readers				Non-clinically Trained Readers			
	Read 1		Read 2		Read 1		Read 2	
Reader	2	3	2	3	2	3	2	3
1	0.95	0.90	0.83	0.93	0.83	0.85	0.79	0.84
2		0.91		0.92		0.77		0.86

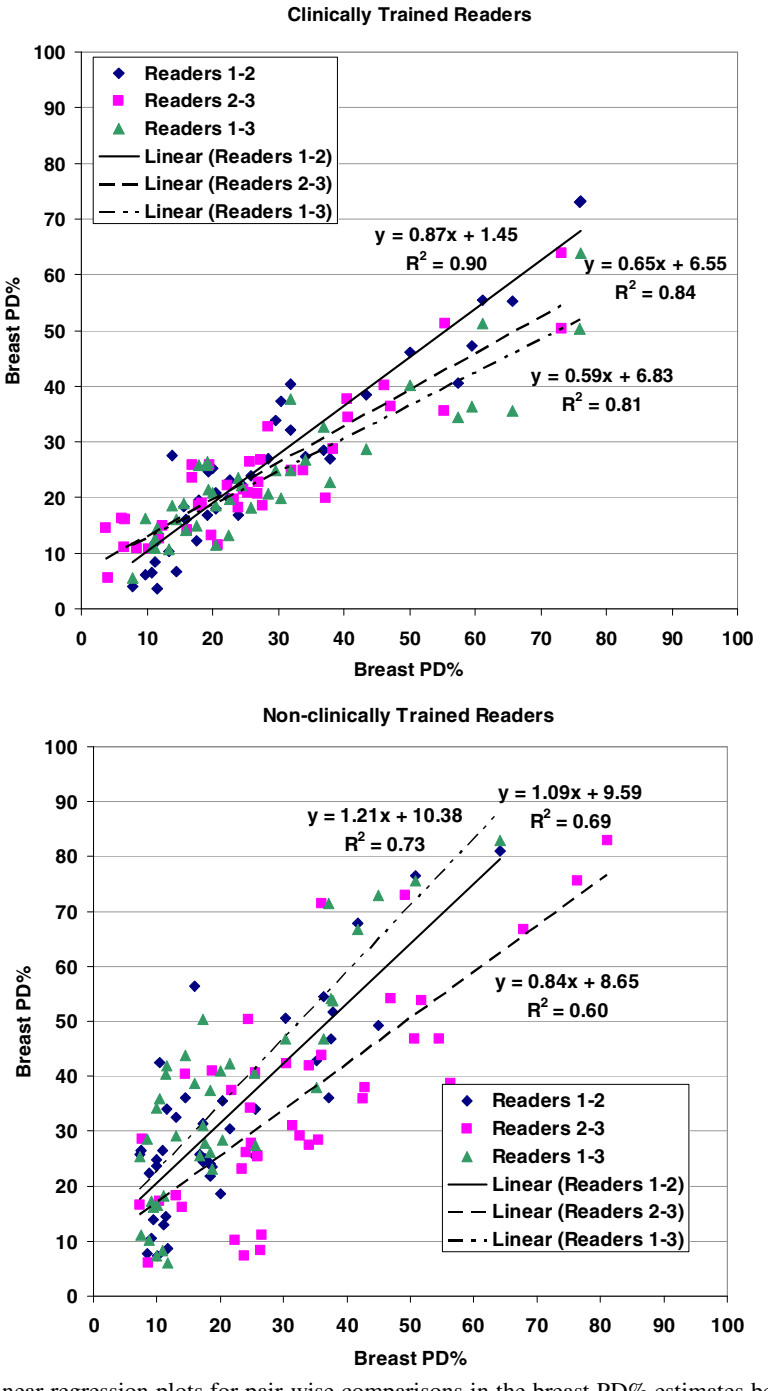

Fig. 1. Linear regression plots for pair-wise comparisons in the breast PD% estimates between the readers in each group (Read 1) with adjusted-R^2 estimates and linear regression equations. Readers in each group are ordered in increasing order of experience.

Table 2. Pearson correlations (r) for intra-reader variability in breast PD% estimates between the two reads of the readers in each group. Readers are ordered in increasing order of experience.

	Pearson Correlations (r) for PD% intra-reader variability					
	Clinically Trained Readers			Non-clinically Trained Readers		
Reader	1	2	3	1	2	3
	0.98	0.79	0.95	0.91	0.91	0.89

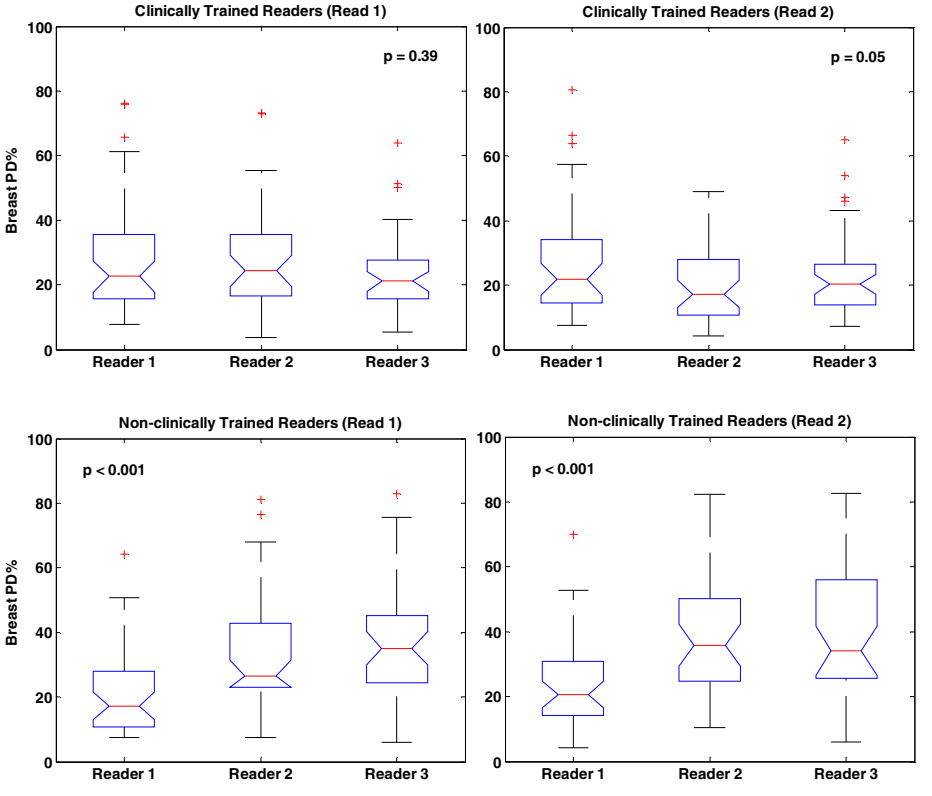

Fig. 2. Box-plots and ANOVA p-values for the breast PD% estimates of the readers in the clinically trained group (up) and the non-clinically trained group (down) after the first and the second breast PD% reads. Readers are ordered in increasing order of experience.

4 Discussion

The observed overall high inter- and intra- reader correlations suggest that area-based breast PD% estimates obtained from post-processed DM images could be a viable means for obtaining breast density measures in DM. However, the level of experience and training of the reader may have an effect on the obtained measures. Our study suggests that clinically trained readers, such as breast imaging radiologists, with minimal training in the use of the semi automated image-thresholding software

(*i.e., Cumulus* Ver. 4.0, Univ. Toronto), have an overall higher inter-reader agreement in performing breast PD% estimation on post-processed DM images. The higher correlation among the clinically trained readers may be attributed to their greater clinical experience with the post-processed DM images and their clinical knowledge of breast anatomy.

The statistically significant difference in the means of the breast PD% measures between the two groups suggests that differences in breast imaging training could potentially impact a patient's risk assessment outcome. We attribute the observed differences in breast PD% estimates between the clinically and the non-clinically trained readers in potentially inherent differences in their corresponding visual perception of the dense breast tissue region in the DM image. Figure 3 illustrates an example where the difference in the thresholded dense tissue region resulted in significantly different breast PD% estimates between the two readers.

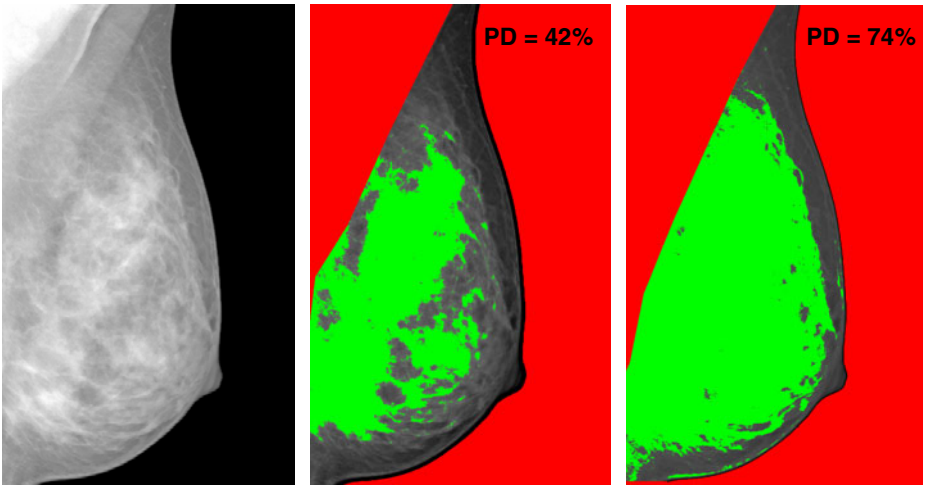

Fig. 3. An example of an MLO DM image and the corresponding *Cumulus* dense tissue thresholding for a clinically trained and a non-clinically trained reader with PD% estimates.

Consistent and reproducible measures of breast PD% will become increasingly necessary as breast density measures become more frequently incorportated in breast cancer risk assessment algorithms. In addition, consistent and reproducible breast density measures will continue to be important in understanding the potential sensitivity of mammographic screening for individual women as we move forward towards adopting personalized screening algorithms for breast cancer detection. Fully-automated methods for estimating breast PD% hold the promise to alleviate the subjectivity introduced by individual readers and result in more accurate quantitative measures [10-12]. Further work is underway to compare the breast PD% estimates obtained from the post-processed DM images with the corresponding measures estimated from the raw (*i.e.,* unprocessed) digital mammograms.

Acknowledgments. This work was partially funded by the Susan G. Komen for the Cure Foundation (Postdoctoral Fellowship KG080756), the DOD (Concept Award BC086591), and the National Institutes of Health/National Cancer Institute (Program Project Grant P01-CA85484, PI M.D. Schnall). We would like to thank Dr. Martin Yaffe for providing the *Cumulus* software and Dr. Johnny Kuo for developing and maintaining the RSNA Medical Imaging Resource Center (MIRC) image archive. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH.

References

1. Boyd, N.F., Guo, H., Martin, L.J., Sun, L., Stone, J., Fishell, E., Jong, R.A., Hislop, G., Chiarelli, A., Minkin, S., Yaffe, M.J.: Mammographic density and the risk and detection of breast cancer. New England Journal of Medicine 356, 227–236 (2007)
2. Nicholson, B.T., LoRusso, A.P., Smolkin, M., Bovbjerg, V.E., Petroni, G.R., Harvey, J.A.: Accuracy of assigned BI-RADS breast density category definitions. Academic Radiology 13, 1143–1149 (2006)
3. Ooms, E.A., Zonderland, H.M., Eijkemans, M.J., Kriege, M., Mahdavian Delavary, B., Burger, C.W., Ansink, A.C.: Mammography: interobserver variability in breast density assessment. Breast 16, 568–576 (2007)
4. Yaffe, M.J.: Mammographic density. Measurement of mammographic density. Breast Cancer Research 10, 209 (2008)
5. McCormack, V.A., dos Santos Silva, I.: Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiology, Biomarkers & Prevention 15, 1159–1169 (2006)
6. Bakic, P.R., Carton, A.-K., Kontos, D., Zhang, C., Troxel, A.B., Maidment, A.D.A.: Breast percent density estimation from mammograms and central tomosynthesis projections. Radiology 252, 40–49 (2009)
7. Kontos, D., Ikejimba, L., Bakic, P.R., Troxel, A.B., Conant, E.F., Maidment, A.D.A.: Digital breast tomosynthesis parenchymal texture analysis for breast cancer risk estimation: Results from a screening trial. Presented at 95th Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, IL (2009)
8. Kontos, D., Bakic, P.R., Carton, A.K., Troxel, A.B., Conant, E.F., Maidment, A.D.A.: Parenchymal texture analysis in digital breast tomosynthesis for breast cancer risk estimation: A preliminary study. Academic Radiology 16, 283–298 (2009)
9. Bakic, P.R., Carton, A.K., Kontos, D., Zhang, C., Troxel, A.B., Maidment, A.D.: Breast percent density: estimation on digital mammograms and central tomosynthesis projections. Radiology 252, 40–49 (2009)
10. Jeffreys, M., Warren, R., Highnam, R., Smith, G.D.: Initial experiences of using an automated volumetric measure of breast density: the standard mammogram form. Br. J. Radiol. 79, 378–382 (2006)
11. Heine, J.J., Carston, M.J., Scott, C.G., Brandt, K.R., Wu, F.F., Pankratz, V.S., Sellers, T.A., Vachon, C.M.: An automated approach for estimation of breast density. Cancer Epidemiol Biomarkers Prev. 17, 3090–3097 (2008)
12. van Engeland, S., Snoeren, P.R., Huisman, H., Boetes, C., Karssemeijer, N.: Volumetric breast density estimation from full-field digital mammograms. IEEE Trans. Med. Imaging 25, 273–282 (2006)