J Digit Imaging (2015) 28:626–632
DOI 10.1007/s10278-015-9814-4

# An Investigation into the Consistency in Mammographic Density Identification by Radiologists: Effect of Radiologist Expertise and Mammographic Appearance

Yanpeng Li[1] · Patrick C. Brennan[1] · Warwick Lee[2] · Carolyn Nickson[3] · Mariusz W. Pietrzyk[4] · Elaine A. Ryan[1]

Published online: 11 August 2015


**Abstract** The aim of this work is to investigate how radiologist expertise and image appearance may have an impact on inter-reader variability of mammographic density (MD) identification. Seventeen radiologists, divided into three expertise groups, were asked to manually segment the areas they consider to be MD in 40 clinical images. The variation in identification of MD for each image was quantified by finding the range of segmentation areas. The impact of radiologist expertise and image appearance on this variation was explored. The range of areas chosen by participating radiologists varied from 7 to 73 % across the 40 images, with a mean range of 35±13 %. Participants with high expertise were more likely to choose similar areas to one another, compared to participants with medium and low expertise levels (mean range were 19±10 %, 29±13 % and 25±14 %, respectively, $p<0.0001$). There was a significantly higher average grey level for the area segmented by all radiologists as MD compared to the area of variation, with mean grey level value for 8-bit images being 146±19 vs. 99±14, respectively. MD segmentation borders were consistent in areas where there was a sharp intensity change within a short distance. In conclusion, radiologists with high expertise tend to have a higher agreement when identifying MD. Tissues which have a lower contrast and a less visually sharp gradient change at the interface between high density tissue and adipose background lead to inter-reader variation in choosing mammographic density.

**Keywords** Mammographic density (MD) · Observer variation · Density segmentation · Mammography

**Abbreviations**

| | |
|---|---|
| MD | Mammographic density |
| ROI | Regions of interest |
| A% | Percentage area |
| $R$ | Range of segmentation areas |
| $A_{\mathrm{M}}$ | Median segmentation area |
| $A_{\mathrm{Comm}}$ | Common segmentation area |
| $A_{\mathrm{Comb}}$ | Combined segmentation area |
| $A_{\mathrm{Var}}$ | Variation area |
| AGL | Average grey level |
| $D_x$ | Distance of perpendicular lines |
| $D_{\mathrm{L}}$ | The longest distance of perpendicular line |
| $D_{\mathrm{S}}$ | The shortest distance of perpendicular line |

## Introduction

Within mammographic images, breast fibroglandular tissue appears brighter in contrast to the adipose background, and this is referred as mammographic density (MD). High MD is a significant risk factor for breast cancer [1–3]. Increased MD also potentially lowers the sensitivity of mammography by obscuring small masses [4, 5]. Accurate and reliable assessment of MD has crucial clinical significance, as women with high MD are often referred for additional magnetic resonance imaging (MRI) or ultrasound examinations in order to increase the likelihood of early breast cancer detection [6, 7]. Up to now, 19 states of the

✉ Yanpeng Li
yali3465@uni.sydney.edu.au

1 Faculty of Health Sciences, The University of Sydney, Cumberland Campus C42, M219, East St, Lidcombe, NSW 2141, Australia

2 Cancer Institute NSW, 8 Central Avenue, Eveleigh, NSW 2015, Australia

3 Melbourne School of Population and Global Health, The University of Melbourne, Grattan Street, Parkville, Victoria 3010, Australia

4 Institute of Physics, 76 Portland Place, London W1B 1NT, UK

USA have signed legislation requiring that women are told of their MD status when they attend mammographic breast examinations [8]. Giving women an accurate and reliable MD assessment result is of vital importance. However, this is not an easy task as there is no gold standard for MD assessment, and the American College of Radiologists (ACR) has acknowledged that MD assessment is not reliably reproducible [9].

Radiologists' visual assessment of MD is being widely used by clinical practices and is approved for use in a number of states in the USA [10, 11]. However, studies investigating the consistency of radiologist MD reporting only show moderate inter-reader agreement [12–20]. Computer-aided methods to measure MD are currently developed [21–24] but to date are not widely used clinically, as they still require robust validation [14, 25]. Therefore, improving the reliability and reproducibility of radiologists' visual assessments of MD has great significance. To achieve this goal, it is also important to investigate the underlying reasons that lead to any inconsistency between radiologists in identifying MD.

The overall aim of this study is to investigate what areas within breast images radiologists choose to describe as MD, and the extent of inter-reader variation in defining MD, the impact of radiologist expertise and image appearance on the inconsistency of MD identification.

## Materials and Methods

### Participants and Images

Seventeen radiologists were recruited to participate in this study, and consent was gained before participation. All of the participants currently read mammographic images in routine clinical practice. Six radiologists read more than 5000 mammograms per year, five radiologists annual reading volume is between 2000 and 5000 mammograms, and six participants read less than 2000 mammograms per year. Radiologists were divided into three expertise groups according to above annual reading volume variation, as it has been shown that this reader characteristic (cases per year) most accurately correlates with radiology performance expertise [26].

Digital mammographic images from 40 lesion-free women with a range of MD appearances were selected from the BreastScreen NSW screening program database. The images were selected by the researcher and visually assessed. It was judged that the images chosen included various appearances of mammographic density. The images ranged from a very dark, homogenous, low density appearance, to a very bright high density appearance, and a mixture of that in between. Although this selection is subjective, and not performed by a radiologist, it was deemed to include a range, and this was shown to be the case through the variation shown in the results. Ethical approval through BreastScreen NSW was given in order to access the images for use in this study and retrospective patient consent was not required. For this study, the craniocaudal (CC) view only was utilised, chosen randomly from either the left or right breast, in order to limit the time needed for participants to complete the study. Five images were duplicated and mixed with the original images to test intra-reader consistency.

As radiologists in a clinical environment need to assess MD in processed images, regardless of mammographic systems used, images from various systems were included. The mammograms were acquired from both Digital Radiography (DR) and Computed Radiography (CR) systems, including those manufactured by Sectra (Sectra, Linköping Sweden), Philips (Philips healthcare, Amsterdam, The Netherlands), GE (GE Healthcare Limited, Wisconsin, USA), Agfa (Agfa Healthcare, Mortsel, Belgium) and Fuji (Fuji Film, Tokyo, Japan). Anonymised images were exported from the PACS system and compressed to 8-bit bitmap images prior to participant viewing, in order to allow image analysis using concordant parameters for this study.

### Data Collection

All mammographic images were sequentially displayed on a Wacom Cintique U40 tablet (WACOM Co Ltd, Saitama, Japan), in a reading room with controlled lighting levels. The tablet has a 21.3-in. LCD monitor with a resolution of $1600 \times 1200$ pixels. The aim of this study is primarily to examine how, and to what extent radiologists look for dense areas in an image. The purpose does not focus on the diagnostic quality of the images, and it was deemed to be of a sufficient resolution for the study. The tablet also allowed a unique method of data collection, relevant to the research question. Each participant was asked to segment any regions of MD in the selected mammograms as precisely as possible, using a WACOM interactive stylus that allowed freehand responses to be recorded and saved as regions of interest (ROI) for each image, using custom-made software (Ziltron Dublin, Ireland). The participants were able to segment multiple dense regions or exclude small non-dense regions from larger ROIs. Each participant was able to adjust each ROI but were not able to zoom, pan, or window the images.

### Data Analysis

#### *Quantification of Mammographic Density Segmentation*

The manually segmented ROI for each participant ($n$=17) and each image ($n$=40) were quantified using the percentage area ($A$%), given by Eq. 1, where $P_{\mathrm{ROI}}$ is the number of pixels within the segmented ROI, and $P_{\mathrm{Tot}}$ is the total number of pixels within the breast area. If a participant segmented multiple ROIs within one single image, the numbers of pixels within all the ROIs were combined to find $P_{\mathrm{ROI}}$.

$$A\% = {}^{P_{\mathrm{ROI}}}/_{P_{\mathrm{Tot}}} \times 100 \qquad ([1])$$

Image J image processing software (version 1.47, National Institutes of Health, Bethesda, MD, USA) was used for image analysis and measurements.

For each image, the smallest percentage area ($A_{\mathrm{S}}$), the largest percentage area ($A_{\mathrm{L}}$) and the median percentage area ($A_{\mathrm{M}}$) of all participants' segmentation were found. The range of segmentation areas ($R$) calculated using Eq. 2 and standard deviation of mean segmentation (StdDev) were used to quantify inter-reader segmentation variation. This was carried out for all participating radiologists and for the three radiologist expertise groups. Statistical analysis was performed to compare $R$ and StdDev between three radiologist groups using a Friedman nonparametric test. Dunn's multiple comparisons post hoc test was used for paired comparisons. In addition, two-factor ANOVA was performed to compare segmented areas difference between three radiologist groups for 40 images. Prism statistical software (GraphPad Prism version 5.00 for Windows, GraphPad Software, San Diego, CA, USA) was used to perform all statistical tests.

$$R = A_{\mathrm{L}} - A_{\mathrm{S}} \qquad ([2])$$

Participants segmented the MD of five images twice and the differences in $A\%$ was used to examine the intra-reader consistency.

The correlation between the range of MD segmentation areas ($R$) and the median segmentation area ($A_{\mathrm{M}}$) was calculated using Pearson's correlation coefficient. This was to investigate if the inter-reader variation in MD segmentation increases with increasing MD.

### *Analysis of Composite Images*

Images with composite segmentations were generated to include all overlaid ROIs drawn by the 17 participants. An example of this is shown by the lines in Fig. 1. For each composite image, a common and a combined area were delineated, as shown by the thick lines in Fig. 1. The common segmentation area ($A_{\mathrm{Comm}}$) consists of pixels which were identified as MD by all 17 participants, which did not cause any disagreement. The combined segmentation area ($A_{\mathrm{Comb}}$) consists of pixels that were included as MD by any participant. The difference between the common and the combined areas is the region which shows the variation between participants, and has been quantified as the 'variation area' ($A_{\mathrm{Var}}$), see Eq. 3. This is the region which has been interpreted as MD by at least one, but not all.

$$A_{\mathrm{Var}} = A_{\mathrm{Comb}} - A_{\mathrm{Comm}} \qquad ([3])$$

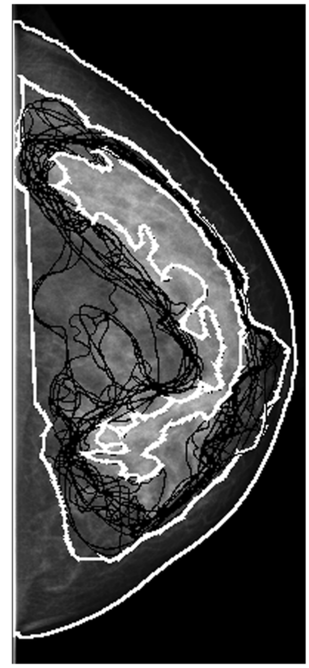

**Fig. 1** An example of an image with composite segmentations, which shows the 17 participants MD segmentation in *thin lines*. The common area ($A_{\mathrm{Comm}}$) is indicated by the *inner thick line* and the combined area ($A_{\mathrm{Comb}}$) by the *middle thick line*. The breast contour is shown in *outer thick line*

A fourth area was also investigated, which was the background area $A_{\mathrm{Back}}$, these are the pixels which were not included by any radiologists in any ROI, and is defined in Eq. 4. $A_{\mathrm{Total}}$ is the area of the breast.

$$A_{\mathrm{Back}} = A_{\mathrm{Total}} - A_{\mathrm{Comb}} \qquad ([4])$$

Average grey level values (AGL) of the common area, variation area, and background area for each image were found, to investigate the brightness differences between them. This was done in order to investigate why some areas that all radiologists agreed to regard them as MD, while other areas cause variation in defining MD. The result can give us an illustration about the brightness difference between these three areas and to test if the difference is significant. A one-way ANOVA was used to compare the grey level differences between the three identified areas, with a Tukey's multiple comparisons post hoc test.

### *Localised Segmentation Analysis*

Within each image, it was found that there can be very strong agreement in outlining MD borders between participants in one area, but not in another. In order to assess this, four or five perpendicular lines across segmentation borders on the composite images were drawn in order to quantify the extent of agreement, as shown in Fig. 2a, b. The perpendicular lines were positioned in different regions around the contour of the segmentation borders, usually in superior-anterior, inferior-anterior, superior-posterior, middle-posterior and inferior-posterior areas. The optimal positions for each image were decided on a case by case basis, trying to include the shortest and longest distances for each image. If the usual position for a line

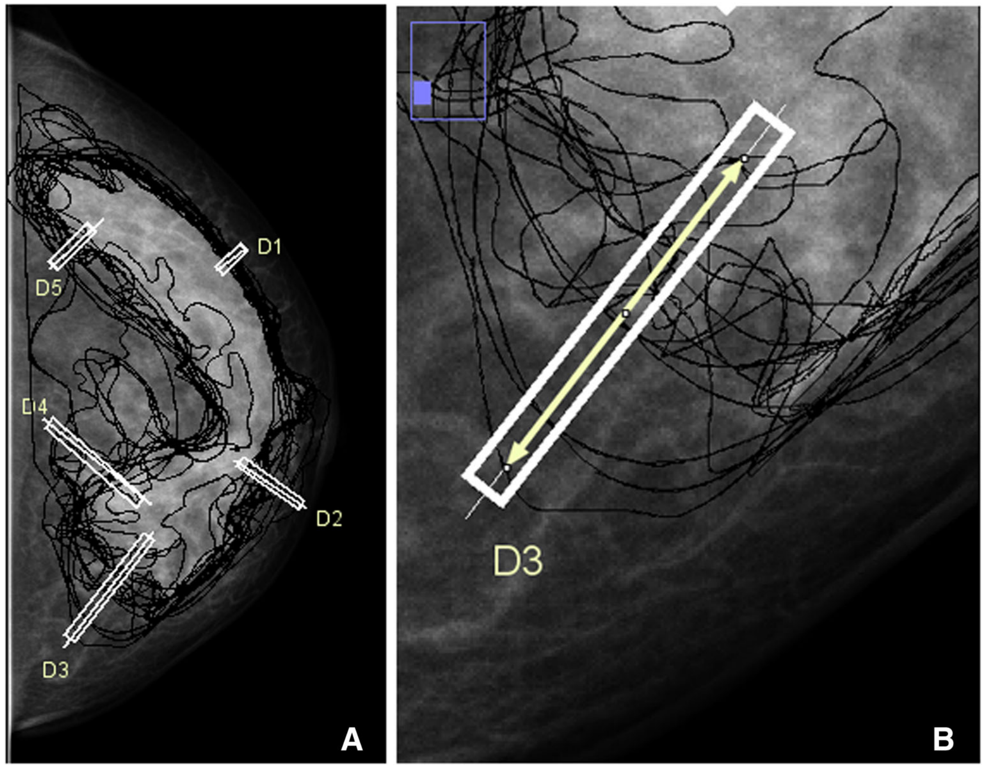


**Fig. 2** **a** An example of five perpendicular lines and surrounding ROIs drawn in different parts of breast to show localised MD segmentation variation. **b** The *arrow* indicates how the MD segmentation variation distance was measured

did not provide a meaningful measurement, the line was repositioned or omitted. It was not meaningful to have standard positions for the lines, due to the nature of the outlining. For example, sometimes a line drawn in the desired position fell in a place where the contours were running parallel, not perpendicular to the line. The length of these perpendicular lines ($D_x$) were used to illustrate inter-reader consistency in choosing MD borders. $D_x$ was measured in numbers of pixels and was normalised to the length of the axillary/lateral border of the breast in order to allow inter-comparison between images.

The intensity changes along these perpendicular lines show that where there is a sharp intensity change between the dense and non-dense areas, the participants tend to segment the MD border more consistently. In contrast, the gradual intensity change between dense and non-dense areas causes more variation between radiologists in how the MD border is defined. In order to quantify this, the intensity change along each of the selected lines was investigated by finding the histogram of a narrow area along each line. An illustration of how these areas were drawn is also shown by Fig. 2a. The histogram of each narrow box shows the distribution of grey level values of the area within the box. Sharp or gradual intensity change resulted in two different patterns of grey level value distribution. Sharp grey level change between dense and non-dense areas resulted in two-peak distribution of grey level values of histograms. One of the peaks was for the grey levels of the dense area, and the other peak was for the grey levels in the non-dense area. These two peaks were usually separated clearly. A gradual intensity change resulted in one-peak distribution of grey level values as the grey level of this local area is from very small values steadily changed to high grey level values. Gaussian peaks were fitted to the histogram using Matlab (MATLAB R2012b, The MathWorks Inc., Natick, MA, USA). This was to objectively decide the number of peaks of histograms, and then a Mann–Whitney test was used to compare $D_x$ between one-peak histograms and two-peak histograms.

## Results

### Quantification of Mammographic Density Segmentation

The variation of MD segmentation areas between radiologists was quantified using the range of segmentation percentage area for each image, as shown in Fig. 3a. The median and range were also found for three radiologist expertise groups, shown in Fig. 3b–d. The range varied from 7 to 73 % across the 40 images for all participants, with a mean range of 35±13 %. The mean range for three radiologist expertise groups were 19±10 % for the high expertise group, 29±13 % for medium expertise group and 25±14 % for the low expertise group ($p$<0.0001). There were significant differences in the MD segmentation ranges between high expertise group and the medium and low expertise groups. There was no significant difference between the medium and low expertise groups. Standard deviation of mean segmentation (StdDev) was also used to quantify inter-reader consistency in MD segmentation. The mean StdDev for all participants was 10±4 %, the

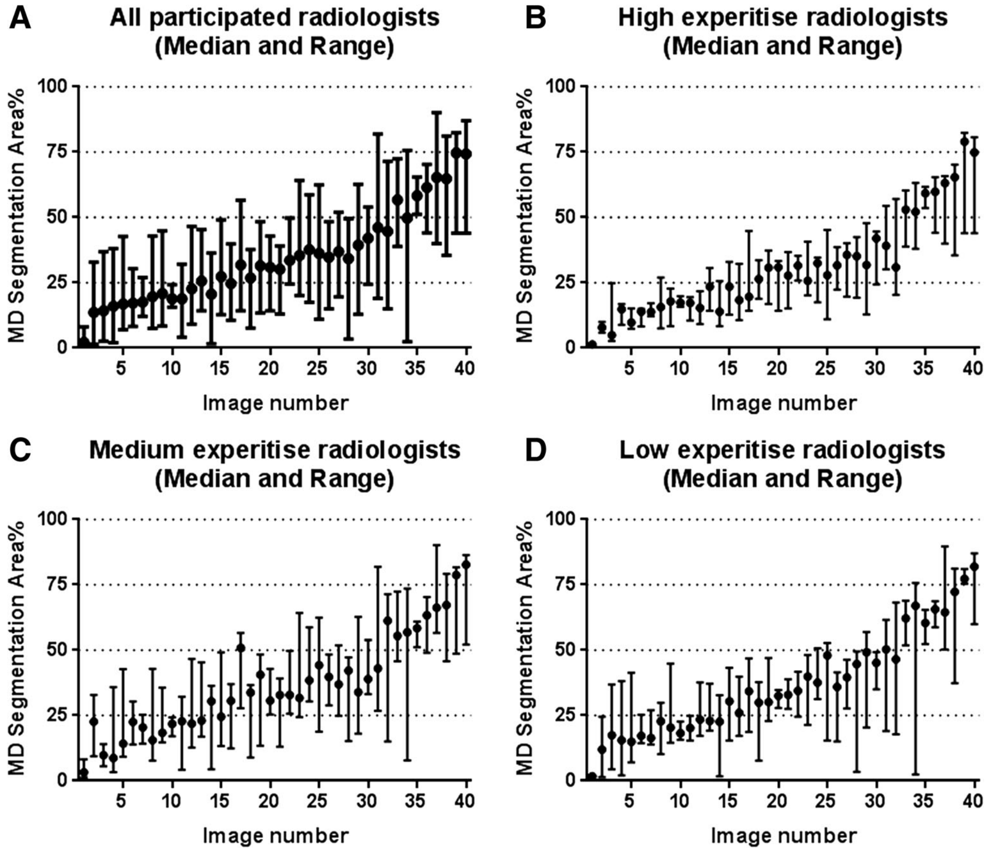


**Fig. 3** Median and range of MD segmentation percentage area for each image, displayed according to increasing median. **a** The results for all participants; **b** the results for high expertise radiologist group; **c** for medium expertise radiologist group; and **d** for low expertise radiologist group

minimum was 2 % and the maximum was 24 %. The mean StdDev for three radiologist expertise groups were 7±4 % for the high expertise group, 13±5 % for medium expertise group and 10±6 % for the low expertise group. There were significant differences between paired radiologist expertise groups ($p<0.0001$). However, the two-factor ANOVA performed using original segmentation areas for the participants from different expertise groups showed no significant difference for three radiologist groups ($F=0.354$, $p=0.722$), as well as no significant difference for pairwise comparisons ($p>0.05$).

For the intra-reader consistency, the median value for the segmentation areas differences for the 85 pairs was 4 % and the mean was 6 %±6. After excluding two outliers which had area differences of more than 20 %, the mean value of the difference decreased to 5 %±5.

The correlation coefficient r between $R$ and $A_M$ was 0.425.

### Analysis of Composite Images

The variation areas ($A_{Var}$) calculated according to composite segmentations varied from 8 to 82 %, with a mean variation area of 48±15 %.

The mean AGL values was 146±19 for the common area ($A_{Comm}$), 99±14 for the variation area ($A_{Var}$) and 68±20 for the background area ($A_{Back}$) across all 40 images, as show in Fig. 4. There is a significant difference for the AGL values between common area, variation area and the background area, $p<0.0001$. Multiple comparisons showed significant differences between pairs of areas. This result gave us a quantitative illustration about the brightness difference between common area, variation area and background area. Grey level of 146 is visually very bright; therefore, all participants agreed areas around this grey level threshold can be regarded as MD. Grey level of 99 is visually less bright; hence, some participants regarded those less bright areas as MD, while other not. This can help to give us an

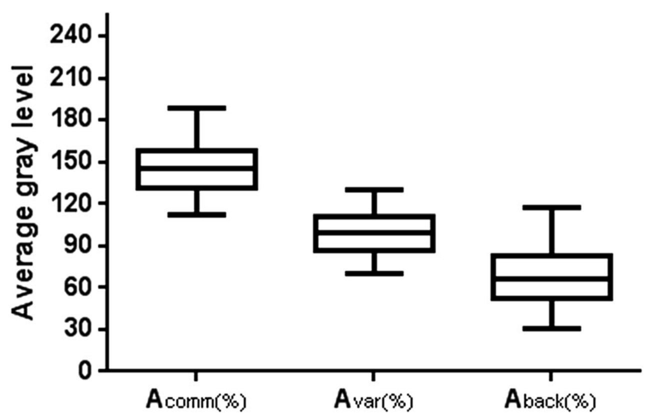


**Fig. 4** A box and whisker plot of AGL values of the common area, variation area and background area. The *box* shows the median (*central line*) and inter-quartile range (*box*), and the *whiskers* show the minimum to maximum range

understanding about what kinds of tissue appearance may cause MD perception difference between radiologists.

### Localised Segmentation Analysis

Within each image, there was an area where the border chosen as MD was consistent between radiologists. Across the 40 images, the shortest variation distances ($D_S$) were between 16 and 50 pixels when the length of lateral border of the breast was normalised to 1000 pixels. However, for most areas within each image, the border chosen was not consistent, with the longest variation distance ($D_L$) varying from 60 to 391, with a mean of 238±85 pixels.

The measurement of the distances along perpendicular lines and number of peaks of the histograms of ROI along the perpendicular lines showed there is significant difference of $D_x$ between two peaks histograms and one-peak histograms ($p<0.001$). For the regions where have sharp changes in intensity between dense and non-dense areas (two-peak histograms), the mean distance was 55 pixels. For the regions where have gradually changes (one-peak histograms), the mean distance was 179 pixels.

## Discussion

An improved understanding of the visual assessment processes used by radiologists can help to discover the possible reasons behind MD reporting inconsistencies that are widely found within the literature [12–20]. Although radiologists within clinical practice only visually assess MD, they do have their own opinions regarding what kinds of tissue appearance should be regarded as MD. Using this MD manual, segmentation method is different to how radiologists assess density in practice, but it is a novel way of expressing how radiologists perceive the importance of density patterns. It is acknowledged that there is no true way of assessing to what extent it really represents clinical practice. Asking radiologists to manually segment the areas they identify as MD is a straightforward way to illustrate the regions they regard as MD. It is also a useful way to show the areas which cause variation in MD identification between radiologists. The results from this study show large inter-reader variation in which areas are considered to be MD. This was not only revealed by the large range ($R$) of MD segmentation areas for each image (mean $R$=35±13 %) but also supported by the large variation area ($A_{Var}$) found in composite images (mean $A_{Var}$=48±15 %) and the long distance between MD segmentation borders (mean $D_L$=238±85 pixels). This variation in MD identification will result in reporting variations. If radiologists report the density category according to their segmented MD, the same image would be put into different density categories by different radiologists. Therefore due to different radiologists choosing various areas as MD, inconsistency in reporting is undoubtedly introduced.

Overall, the participating radiologists for this study had varying experience, came from a variety of clinical environments and had no training at the beginning of the study on what should be considered MD. Consequently, the segmentation areas can be considered a reflection of the individual perception of MD. Large inter-reader variation in MD segmentation indicates different radiologists have various opinions about what kinds of tissue appearance should be identified as MD. However for individual participants themselves, they have their own criteria to identify MD as they chose similar areas as MD for the repeated images. The intra-reader variation was only 6±6 % compared to large inter-reader variation of 35±13 %. Large inter-reader variation can be found for almost all participants, although radiologists with higher expertise levels tend to have a higher agreement in choosing MD, with smaller MD segmentation variation range for the images, as illustrated in Fig. 3b. Radiologists with higher expertise read more mammograms and have a better diagnostic performance [27]; hence, it is possible that they have a better idea as to what kinds of image appearance can cause difficulties in diagnosis. As a result, they may have chosen the areas more likely to cause interpretation problems as MD.

Large variation in MD segmentation can be found across most images regardless of whether the median segmentation area ($A_M$) is small or large. Figure 3 shows $A_M$ steadily increases from less than 10 % to over 75 %; however, the range did not increase correspondingly, but was high for all images. The large variation is more crucial for images with a medium quantity of MD, as these images could be classified as a low MD image by one radiologist but as a high MD image by another. Misinterpretation in density assessment may lead to an increase in anxiety for patients and the burden of unnecessary examinations. Therefore, it is worthwhile to further analyse the characteristics of regions within an image that causes discrepancy in MD identification.

Analysis of composite images shows areas where the fibroglandular tissue has an obviously visual contrast to the adipose background tissue were easily to be identified as MD. However, if the region is brighter than the adipose background but not bright enough to be definitely regarded as MD, or if the region has no visually clear sharp gradient change compared to the background, different radiologists have different opinions about whether to identify it as MD. As there is no definite answer about to what extent the brighter fibroglandular tissue appearance should be regarded as MD, this means MD is not a parameter that can be precisely measured. However, it still has clinical application value when the images are classified into different density categories. When radiologists assign a density category for the breast image, not only should the absolute quantity of very dense or less dense areas be considered, but the heterogeneity of the dense area needs be taken into account as well.

A major limitation of the study was the manual segmentation process. It proved difficult for the radiologists to include all the small isolated mammographically dense regions precisely. This was minimised by asking the radiologists to use a pen stylus and tablet to carry out the task, but it still proved to be quite difficult. Moreover, the quantified percentage area of density was affected by how precisely each radiologist outlined the MD area. Hence, the quantified $R$ for inter-reader variation in MD segmentation would be mildly affected. However, the analysis of composite images helps to show where the variation in MD identification was. Another limitation is although 17 radiologists are a relatively large number to investigate inter-reader variation in mammographic density reporting compared to other studies, it was still a small number to infer some information from the segmentation results. However, these results can give us some indication about the reasons that lead to MD identification variation.

In conclusion, there was large variation in identifying the areas to be regarded as MD between radiologists. Radiologists with high expertise tend to have a higher agreement in choosing MD areas. This variation could lead to inter-reader inconsistency in clinical density reporting. Well contained, brightly displayed areas were labelled as MD by most radiologists, but uncertainty was produced by areas where there was no sharp contrast gradient and the area was in low contrast to the adipose background. This indicates that simple thresholding methods, which classify tissue into 'dense' or 'non-dense' areas, are not adequate in interpreting MD as there is no absolute threshold to differentiate these two types of tissues. Areas where there is a gradual drop off in image density and no sharp border with the background will always require a more sophisticated method of interpretation.

## References


1. Wolfe JN: Breast patterns as an index of risk for developing breast cancer. American Journal of Roentgenology. 126: 1130–1137, 1976
2. McCormack VA, Dos Santos SI: Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiology, Biomarkers & Prevention 15:1159–1169, 2006
3. Huo CW, et al: Mammographic density—a review on the current understanding of its association with breast cancer. Breast Cancer Research and Treatment 144:479–502, 2014
4. Sickles E, D'Orsi CJ, Bassett LW, et al: ACR BI-RADS® Mammography. In: ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System, Reston, VA, 2013
5. AL Mousa D, Ryan EA, Mello-Thoms C, Brennan PC: What effect does mammographic breast density have on lesion detection in digital mammography? Clinical Radiology 69:333–341, 2014
6. Ho JM, Jafferjee N, Covarrubias GM, Ghesani M, Handler B: Dense breasts: a review of reporting legislation and available supplemental screening options. American Journal of Roentgenology 203:449–456, 2014
7. Scheel JR, Lee JM, Sprague BL, Lee CI, Lehman CD: Screening ultrasound as an adjunct to mammography in women with mammographically dense breasts. American Journal of Obstetrics & Gynecology 212(1):9–17, 2014
8. NINETEEN State Density Laws: is your state one of them? Available at http://www.areyoudense.org/worxcms_published/news_page227.shtml. Accessed November 12 2014.
9. ACR Statement on Reporting Breast Density in Mammography Reports and Patient Summaries. Available at http://www.acr.org/About-Us/Media-Center/Position-Statements/Position-Statements-Folder/Statement-on-Reporting-Breast-Density-in-Mammography-Reports-and-Patient-Summaries. Accessed 24th April 2012.
10. ACR Offers Resources to Physicians on Breast Density Notification. Available at http://www.acr.org/Advocacy/eNews/20131004-Issue/ACR-Offers-Resources-to-Physicians-on-Breast-Density-Notification. Accessed November 12 2014.
11. The State of California: Breast Density Notification Law: Senate Bill No. 1538, 2013
12. Salvatore M, et al: Breast density: comparison of chest CT with mammography. Radiology 270:67–73, 2014
13. Redondo A, et al: Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. British Journal of Radiology 85:1465–1470, 2012
14. Ciatto S, et al: A first evaluation of breast radiological density assessment by QUANTRA software as compared to visual classification. Breast 21:503–506, 2012
15. Tagliafico A, et al: Mammographic density estimation: comparison among BI-RADS categories, a semi-automated software and a fully automated one. The Breast 18:35–40, 2009
16. Ooms EA, et al: Mammography: interobserver variability in breast density assessment. Breast 16:568–576, 2007
17. Martin KE, et al: Mammographic density measured with quantitative computer-aided method: comparison with radiologists' estimates and BI-RADS categories. Radiology 240:656–665, 2006
18. Nicholson BT, LoRusso AP, Smolkin M, Bovbjerg VE, Petroni GR, Harvey JA: Accuracy of assigned BI-RADS breast density category definitions. Academic Radiology 13:1143–1149, 2006
19. Ciatto S, et al: Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories. The Breast 14:269–275, 2005
20. Berg WA, Campassi C, Langenberg P, Sexton MJ: Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. American Journal of Roentgenology. 174:1769–1777, 2000
21. Yaffe MJ: Mammographic density. Measurement of mammographic density. Breast Cancer Research 10:209, 2008
22. Jeffreys M, Warren R, Highnam R, Davey Smith G: Breast cancer risk factors and a novel measure of volumetric breast density: cross-sectional study. British Journal of Cancer 98:210–216, 2008
23. Heine JJ, et al: A novel automated mammographic density measure and breast cancer risk. Journal of the National Cancer Institute 104: 1028–1037, 2012
24. Boyd N, et al: Mammographic density and breast cancer risk: evaluation of a novel method of measuring breast tissue volumes. Cancer Epidemiology, Biomarkers & Prevention 18:1754–1762, 2009
25. Lee HN, Sohn YM, Han KH: Comparison of mammographic density estimation by Volpara software with radiologists' visual assessment: analysis of clinical-radiologic factors affecting discrepancy between them. Acta Radiologica. doi: 10.1177/0284185114554674, 2014.
26. Reed WM, Lee WB, Cawson JN, Brennan PC: Malignancy detection in digital mammograms: important reader characteristics and required case numbers. Academic Radiology 17:1409–1413, 2010
27. Rawashdeh MA, et al: Markers of good performance in mammography depend on number of annual readings. Radiology 269:61–67, 2013