Breast Cancer Risk Analysis Based on a Novel Segmentation Framework for Digital Mammograms

Xin Chen, Emmanouil Moschidis, Chris Taylor, and Susan Astley

Centre for Imaging Sciences, Institute of Population Health,
University of Manchester, Oxford Road, Manchester, M13 9PT, UK

Abstract. The radiographic appearance of breast tissue has been established as a strong risk factor for breast cancer. Here we present a complete machine learning framework for automatic estimation of mammographic density (MD) and robust feature extraction for breast cancer risk analysis. Our framework is able to simultaneously classify the breast region, fatty tissue, pectoral muscle, glandular tissue and nipple region. Integral to our method is the extraction of measures of breast density (as the fraction of the breast area occupied by glandular tissue) and mammographic pattern. A novel aspect of the segmentation framework is that a probability map associated with the label mask is provided, which indicates the level of confidence of each pixel being classified as the current label. The Pearson correlation coefficient between the estimated MD value and the ground truth is 0.8012 (p-value<0.0001). We demonstrate the capability of our methods to discriminate between women with and without cancer by analyzing the contralateral mammograms of 50 women with unilateral breast cancer, and 50 controls. Using MD we obtained an area under the ROC curve (AUC) of 0.61; however our texture-based measure of mammographic pattern significantly outperforms the MD discrimination with an AUC of 0.70.

Keywords: Digital mammogram, segmentation, breast cancer risk, mammographic density, texture analysis.

1 Introduction

A major focus of breast cancer imaging research in recent years has been the analysis of mammographic breast density. It has been shown that women with high percentage mammographic density (MD), measured as the proportion of the breast area occupied by dense fibroglandular tissue, have a two to six fold increased breast cancer risk compared to women with low MD [1,2]. Semi-automated computer based tools have been developed where the reader interactively sets thresholds for the breast region and for dense tissue, and the resulting MD is automatically calculated; the most widely used of these is Cumulus [3]. Whilst percentage density measured visually and by Cumulus have been related to cancer risk, these measures are subjective and area-based. Increasing interest in direct and objective measurement of volumes of fat and dense tissue has led to the development of automated volumetric methods such as Cumulus V [5] and Volpara™ [6].

P. Golland et al. (Eds.): MICCAI 2014, Part I, LNCS 8673, pp. 536–543, 2014.

The categorical breast pattern assessment systems originally suggested by Wolfe [7] and Tabár [8] include more complex appearance patterns rather than simply estimating the relative proportion of dense tissue. In recent years, researchers have developed sophisticated texture extraction and analysis methods to characterize mammographic patterns, demonstrating the importance of using texture as a risk factor for breast cancer risk analysis [9].

In our previous work [10], we proposed a novel framework that is able to simultaneously segment the breast region, fatty tissue, pectoral muscle, glandular tissue and nipple region in digital mammograms. Building on this, one contribution of this paper is the development of an automatic MD estimation method, which we show to have discriminatory power for separating the mammograms of women with and without breast cancer. We then describe a texture extraction method for breast cancer risk analysis which outperforms the MD method.

In the following sections, we give a brief description of the segmentation framework followed by introducing the proposed methods for breast cancer risk analysis in section 3. A description of the evaluation methodology and results are given in section 4, and in section 5 we draw conclusions from our work to date, suggesting possible improvements and extensions to the techniques described in the paper.

2 Segmentation Framework

In machine learning method, image features and their associated labels are learnt from a training data set. When an unknown image feeds into the resulting model, the class of pixels in the new image can be estimated according to their corresponding feature descriptors. In [10], we described a system that combines the dual-tree complex wavelet transform (DT-CWT) and random forest (RF) classifier for the purpose of anatomic feature segmentation in mammograms. Here we briefly describe this method which forms the basis of the breast cancer analysis methods described in section 3.

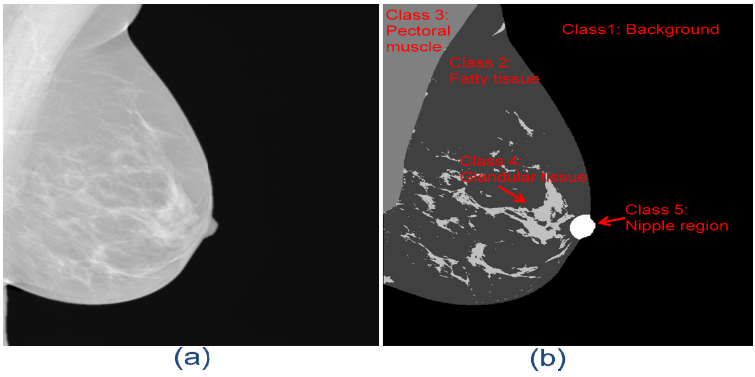

Fig. 1. (a) Logarithm of raw digital mammogram (b) Ground truth segmentation mask

In the training stage, medio-lateral oblique (MLO) view mammograms were segmented into five classes (1: background, 2: fatty tissue, 3: pectoral muscle, 4: glandular tissue, 5: nipple region). The ground truth segmentation was performed by a trained breast radiologist using a semi-automatic software interface; an example segmentation is shown in Fig. 1. For each training image, 500 pixels (as determined experimentally) were randomly selected from each of the five classes. To represent the features of each training pixel, the DT-CWT coefficients together with the normalised X-Y coordinates (origin at the top left corner) and the logarithm of the raw digital mammogram pixel value are used. The feature descriptor for each pixel is a 75-dimensional vector (magnitude and angle parts of DT-CWT × 6 orientations × 6 levels of image pyramid + 2 normalised x, y coordinates + 1 pixel value). Based on the feature vectors and their corresponding classes from all training images, we trained a RF model (200 trees) for classifying unseen mammograms.

When an unseen mammogram is analysed, the same method of calculating a feature vector for each pixel is used. By feeding the feature vector into the trained RF classifier, the probability that a pixel belongs to each of the five classes is obtained. For the input image shown in Fig. 2 (a), Fig. 2 (b)-(f) present the probability maps (brighter pixels represent higher probabilities) obtained for each of the five classes. By assigning the highest probability value of the five classes to each of the pixels, Fig. 2 (g) shows the combined probability map associated with the class labels in Fig. 2 (h). In Fig. 2(h), a small region below the nipple is inaccurately estimated as the nipple. However, the associated probability map (Fig. 2 (g)) shows that the pixels in the misclassified region have very low probabilities which reduce the confidence associated with classification to the assigned label. This smart functionality permits more effective application of the segmentation results in further analysis.

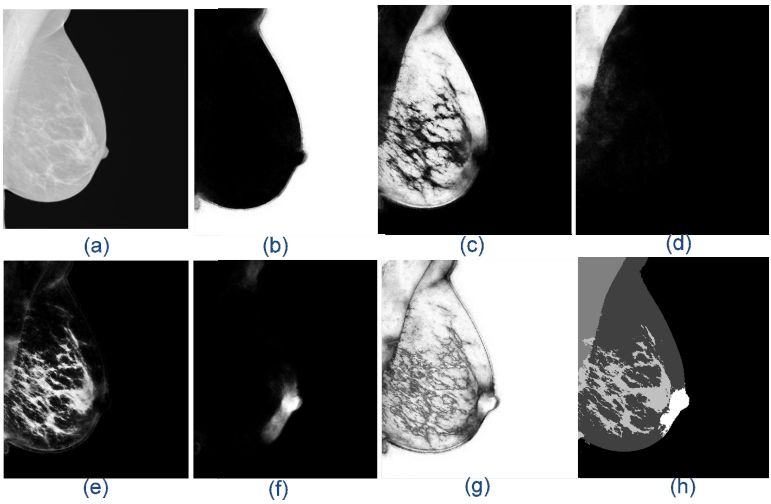

Fig. 2. (a) Input digital mammogram. Outputs from the proposed framework: (b) probability map of background (c) probability map of fatty tissue (d) probability map of pectoral muscle (e) probability map of glandular tissue (f) probability map of nipple region (g) combined probability map (h) label mask. Brighter pixels represent higher probabilities.

3 Breast Cancer Risk Analysis

3.1 Percentage Mammographic Density

A commonly used definition of mammographic density is the area of glandular tissue in a mammogram relative to the area of the breast (denoted as the breast region in this paper). One application of the segmentation framework described in section 2 is fully automatic estimation of MD. Based on the output segmentation mask (Fig. 2(h)), the total number of pixels in the glandular region is calculated and denoted as N_g. The breast region that excluding the pectoral muscle area can be calculated by summing the total number of pixels in the glandular region, fatty region and nipple region, represented by N_b. Hence, $\frac{N_g}{N_b} \times 100\%$ is the estimated MD. We have compared MD obtained automatically in this way with MD derived from the ground truth segmentation, and further investigated the capability to discriminate cancer and non-cancer subjects. The evaluation experiments and results are presented in section 4.

3.2 Mammographic Pattern Analysis

The DT-CWT has been shown to provide a suitable representation of linear structures in mammographic images [11]. We have also demonstrated that it is capable of discriminating different anatomical features in mammograms (section 2 and [10]). We therefore further investigate the use of the DT-CWT to characterize mammographic patterns for breast cancer risk analysis.

Similarly to the segmentation framework described in section 2, our method consists of a model training stage and a classification stage. In the training stage, the DT-CWT is applied to a number of selected pixel locations in the input training images. At each pixel location, the DT-CWT coefficients are calculated on 6 levels of the image pyramid and for 6 different orientations at each level. Therefore, a 72-element (2 magnitude and phase components of the DT-CWT× 6 orientations × 6 levels of image pyramid) feature descriptor is obtained for each selected pixel. Using the output of our segmentation method, the pixel selection process can be random sampling from the breast region, or selective sampling from the fatty and/or glandular regions depending on classification probability values. A comparison of different sampling strategies is given in section 4.3. By feeding the feature descriptors and their corresponding image types (cancer or non-cancer) to a RF classifier, a discrimination model can be trained. To determine whether an unseen image is likely to be from a woman with or without cancer, the same sampling strategy is used to extract a number of pixels. The same 72-element DT-CWT feature descriptor is calculated for each pixel and fed into the trained RF model. Based on the votes of trees from the RF model, the probability of each sample pixel belonging to a cancer (or non-cancer) mammogram can be calculated. The average probability of all the sample pixels is output as a breast cancer risk score.

4 Evaluation

4.1 Data and Pre-processing

We use a balanced case-control dataset of 50 cancer cases and 50 controls. All images are anonymised full-field digital screening mammograms obtained from GE Senographe Essential mammography systems with a pixel size of 94.1 μm. The cancer cases were selected randomly from the most recent available screen-detected malignant breast cancers identified, excluding interval cancers and mammograms showing bilateral breast cancer. The medio-lateral oblique (MLO) view of the contralateral breast was analysed as a surrogate for the prior mammogram. Controls were selected randomly from normal screening mammograms where a subsequent normal mammogram was available, using MLO views in the same ratio of left and right breasts as in the cancer cases. To minimise the effects of machine parameter variations and other image intensity variations, all the raw (unprocessed) digital mammograms were pre-processed by a normalisation algorithm that is embedded in the commercially available software Volpara^{TM} [6]. The density maps output by Volpara^{TM} are used as the input to our method. If not otherwise stated, all the experiments in section 4 were performed in a 5-fold cross validation manner. The dataset was randomly organised into five subgroups, each with 10 cancer cases and 10 controls. Four groups of images were used for training and testing was performed on the remaining group, repeating until all groups have been tested.

4.2 Evaluation of Automated Mammographic Density Estimation

We have compared MD from the automatic method described in section 3.1 with MD derived from the ground truth. The ground truth mask was obtained interactively as described in section 2, in a process similar to that used by Cumulus [3]. The Pearson correlation coefficient between the automated MD and their corresponding ground truth MD is 0.8012 (p-value <0.0001). A Bland-Altman plot and scatter plot of the two sets of values are shown in Fig. 3, which demonstrates the strong correlation between them. Additionally, by varying the threshold of the MD scores to assign images to cancer and non-cancer groups and comparing with known image classes, a receiver operating curve (ROC) can be generated which illustrates the capability of the method for determining whether a mammogram belongs to a cancer case or not. The area under the ROC curve (AUC) values for MD from the ground truth and from the automatic MD are 0.6160 and 0.5812 with the sensitivity and specificity at the equal-error-rate point of 60% and 57% respectively. These results are listed in Table 1 with the results obtained from the texture analysis method in section 4.3.

4.3 Evaluation of Mammographic Pattern Analysis

As described in section 3.2, based on the segmentation results, different sampling strategies can be used in the mammographic pattern analysis method. Here we compare different sampling strategies in terms of their discriminatory power for cancer

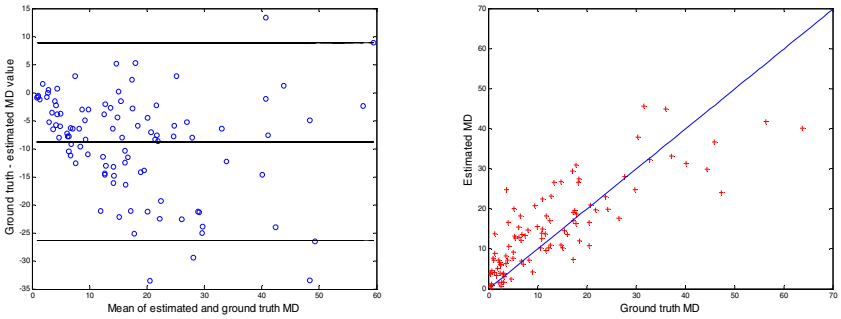

Fig. 3. Estimated MD vs. ground truth MD: (a) Bland-Altman plot (b) Scatter plot

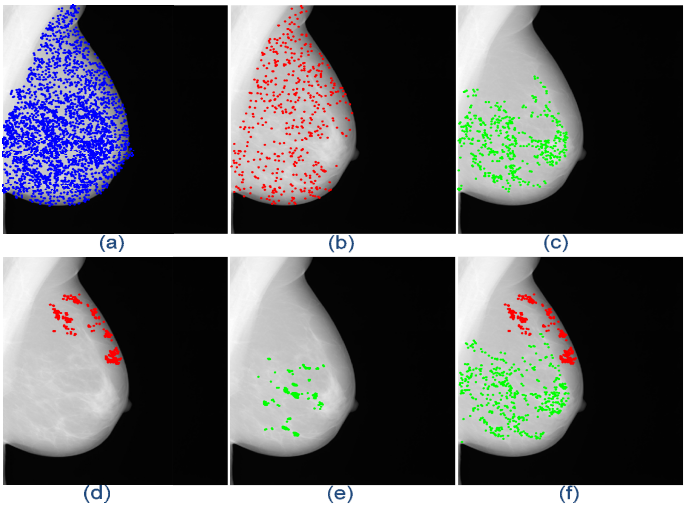

Fig. 4. Sampling strategies (total of 166639 pixels in the whole breast region in this example): (a) 3k random pixels in whole breast region (3KRBreast) (b) 500 random pixels in fatty region (0.5KRFat) (c) 500 random pixels in glandular region (0.5KRGla) (d) 500 pixels with highest probabilities in fatty region (0.5KPFat) (e) 500 pixels with highest probabilities in glandular region (0.5KPGla) (f) 500 pixels with highest probabilities in fatty region and 500 random pixels in glandular region (0.5KPFat_0.5KRGla).

risk analysis. As shown in Fig. 4, we have evaluated our methods using random sampling in the whole breast region, random sampling in the fatty region only, random sampling in the glandular region only, selective sampling in the fatty region (using pixels with a high probability of belonging to the fatty component of the breast), selective sampling in glandular region (using pixels with a high probability of belonging to the glandular component of the breast) and a combination of selective sampling in the fatty region and random sampling in the glandular region (combined the most successful sampling strategies from the two regions). The number of pixels for each

Table 1. ROC performance for breast cancer risk analysis by using different sampling strategies in texture analysis and MD scores

	AUC	*Sensitivity/Specificity at EER*
3KRBreast [(a)]	0.6050	54.0%
0.5KRFat [(b)]	0.6532	58.0%
0.5KRGla [(c)]	0.6032	56.0%
0.5KPFat [(d)]	**0.6912**	**66.0%**
0.5KPGla [(e)]	0.5508	56.0%
0.5KPFat_0.5KRGla [(f)]	**0.7052**	**66.0%**
Ground Truth MD	0.6160	60.0%
Automatic MD	0.5812	57.0%
0.5KPFat_0.5KRGla & Ground Truth MD	0.6800	64.0%

sampling strategy (shown in the caption of Fig. 4) was determined experimentally, with the aim of achieving optimum performance in terms of discriminatory power and computational time.

The ROC performance of each evaluation is listed in Table 1. The first column in Table 1 corresponds to the sampling methods illustrated in Figure 4, together with results for MD from the ground truth and the automatic MD. The last row in Table 1 shows the ROC performance by combining the best texture method and the ground truth MD using logistic regression. It can be seen from the results that the combination of selective sampling in fatty region and random sampling in glandular region produced the best discrimination power, with an AUC of 0.7052 and 66% sensitivity and specificity at the equal error rate (EER) point. Selective sampling in the fatty region alone produced almost as good performance as the combined sampling. We can also conclude that mammographic pattern (texture) analysis outperforms MD for breast cancer risk analysis, and linear combination of the two does not improve performance.

5 Conclusion and Discussions

In this paper, we presented a novel and effective framework for digital mammogram segmentation and breast cancer risk analysis. From this framework, the mammographic breast density can be estimated automatically, and has a strong correlation with the results from the ground truth (Pearson correlation coefficient of 0.8012). Automation is essential for use in breast screening, where throughput is high. Our method corresponds to the area based methods of estimating mammographic density which have, to date, shown the strongest relationship to cancer risk. Objective assessment of risk based on mammographic appearance has a texture component that has not been widely exploited in automated applications and commercial software, which tends to focus on the quantity of dense tissue within the breast rather than its pattern. We have developed an automated mammographic pattern analysis method which has achieved an AUC of 0.70 for discriminating between the contralateral mammograms of women with breast cancer and mammograms of women without

breast cancer. This texture analysis method was found to have more discriminatory power than the well-established MD approach (AUC 0.61). The advantages seem to come from the sampling that concentrated on the fatty region rather than the selection of DT-CWT feature descriptor, since sampling over the entire breast or the glandular region performed no better than the MD. Presumably, such a sampling strategy might also work well with other texture measures. In the literature, AUC values reported for breast cancer risk analysis are in the range of 0.55 to 0.65, dependent on the data used for evaluation. We therefore claim that our methodology and evaluation as reported in this paper make a significant contribution to the problem of estimating breast cancer risk. Future work will concentrate on improving segmentation accuracy by investigating the contribution from each part of the feature descriptor, refining the sampling strategy. We will explore the underlying reason that fatty region contributes more than the glandular region for breast cancer analysis in an extended version of the paper, based on a larger dataset.

References

1. McCormack, V., dos, I., Silva, S.: Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol. Biomarkers Prev. 15(6), 1159–1169 (2006)
2. Boyd, N., Martin, L., Bronskill, M., Yaffe, M., Duric, N., Minkin, S.: Breast tissue composition and susceptibility to breast cancer. J. Natl. Cancer Inst. 102(16), 1224–1237 (2010)
3. Byng, J., Boyd, N., Fishell, E., Jong, R., Yaffe, M.: The quantitative analysis of mammographic densities. Phys. Med. Biol. 39(10), 1629 (1994)
4. Keller, B., Nathan, D., Wang, Y., Zheng, Y., Gee, J., Conant, E., Kontos, D.: Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation. Phys. Med. Biol. 39(8), 4903–4917 (2012)
5. Alonzo-Proulx, O., Packard, N., Boone, J., Al-Mayah, A., Brock, K., Shen, S., Yaffe, M.: Validation of a method for measuring the volumetric breast density from digital mammograms. Phys. Med. Biol. 55(11), 3027–3044 (2010)
6. Jeffreys, M., Harvey, J., Highnam, R.: Comparing a New Volumetric Breast Density Method (Volpara^{TM}) to Cumulus. In: Martí, J., Oliver, A., Freixenet, J., Martí, R. (eds.) IWDM 2010. LNCS, vol. 6136, pp. 408–413. Springer, Heidelberg (2010)
7. Wolfe, J.: Breast patterns as an index of risk for developing breast cancer. Am. J. Roentgenol. 126(6), 1130–1137 (1976)
8. Gram, I., Funkhouser, E., Tabár, L.: "The Tabár classification of mammographic parenchymal patterns ". Eur. J. Radiol. 24(2), 131–136 (1997)
9. Nielsen, M., Karemore, G., Loog, M., Raundahl, J., Karssemeijer, N., Otten, J., Karsdal, M., Vachon, C., Christiansen, C.: A novel and automatic mammographic texture resemblance marker is an independent risk factor for breast cancer. Cancer Epidemiol. 35(4), 381–387 (2011)
10. Chen, X., Moschidis, E., Taylor, C., Astley, S.: A novel framework for fat, glandular tissue, pectoral muscle and nipple segmentation in full field digital mammograms. In: Fujita, H., Hara, T., Muramatsu, C. (eds.) IWDM 2014. LNCS, vol. 8539, pp. 201–208. Springer, Heidelberg (2014)
11. Berks, M., Chen, Z., Astley, S., Taylor, C.: Detecting and classifying linear structures in mammograms using random forests. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 510–524. Springer, Heidelberg (2011)