Open Framework for Mammography-based Breast Cancer Risk Assessment

Said Pertuz[1,*], German F. Torres[2], Rulla Tamimi[3], Joni Kamarainen[2]
[1]*Universidad Industrial de Santander, Bucaramanga, Colombia*
[2]*Tampere University, Tampere, Finland*
[3]*Brigham and Women's Hospital, Boston, USA*
[*]spertuz@uis.edu.co

***Abstract*—In recent years, several studies have established a relationship between mammographic parenchymal patterns and breast cancer risk. However, there is a lack of publicly available data and software for objective comparison and clinical validation. This paper presents an open and adaptable implementation (OpenBreast v1.0) of a fully-automatic computerized framework for mammographic image analysis for breast cancer risk assessment. OpenBreast implements mammographic image analysis in four stages: breast segmentation, detection of region-of-interests, feature extraction and risk scoring. For each stage, we provide implementations of several state-of-the-art methods. The pipeline is tested on a set of 305 full-field digital mammography images corresponding to 84 patients (51 cases and 49 controls) from the breast cancer digital repository (BCDR). OpenBreast achieves a competitive AUC of 0.846 in breast cancer risk assessment. In addition, used jointly with widely accepted risk factors such as patient age and breast density, mammographic image analysis using OpenBreast shows a statistically significant improvement in performance with an AUC of 0.876 (p<0.001). Our framework will be made publicly available and it is easy to incorporate new methods.**

***Index Terms*—mammography, breast cancer, risk assessment, parenchymal analysis, texture analysis**

I. Introduction

Breast cancer is one of the leading causes of cancer deaths in women worldwide [1]. The identification of patients at high risk of developing breast cancer is of key importance to improve timely detection and prognosis of the disease. For this reason, substantial research efforts are devoted to the development of effective risk assessment methods based on different sources, such as genetic biomarkers, risk models, and image analysis [2]. Among existing strategies for risk assessment, we are particularly interested in the analysis of mammography images due to their wide adoption for screening programs worldwide, relatively low cost compared to alternative imaging modalities, and use as one of the primary screening tools. In this scope, the computerized analysis of mammography images has been widely studied for risk assessment. In the last decades, substantial experimental evidence have been found that consistently attributes texture patterns a significant role in breast cancer risk [3].

In the literature, researchers have investigated different tasks related to analysis of mammography images. In general, computerized mammographic analysis can be divided into four stages: i) breast segmentation, ii) region-of-interest detection, iii) feature extraction, and iv) risk scoring. Unfortunately, in the community there is no consensus about the best algorithms for each stage or how the analysis should be performed. One of the main challenges in this regard is *reproducibility*, which refers to the difficulty in replication of any prior art by independent researchers different from the original authors. This could be explained by the complexity of current computer-based approaches and the need for tuning method-specific parameters. It is clear that the difficulties in replication and extensive, independent validation are an important obstacle for reaching consensus and eventual adoption into clinical practice.

In order to tackle the aforementioned reproducibility problem, we present an open protocol and framework for breast cancer risk assessment using mammography images. Specifically, we implement and evaluate a set of state-of-the-art methods for each processing stage in breast cancer risk assessment based on parenchymal patterns. For evaluation, we compare different feature extraction methods and breast sampling strategies on a set of 305 full-field digital mammography images corresponding to 84 patients from the breast cancer digital repository (BCDR) [4]. In addition, we conduct experiments including established risk factors, such as patient age and breast density, showing that the incorporation of the best method combination produces a statistically significant improvement in performance with an AUC of 0.876 (p<0.001).

The main contribution of this work is the development of an open framework - OpenBreast v1.0 - for breast cancer risk assessment from mammography images. Combined with a publicly available data (BCDR in our case) OpenBreast facilitates fair comparisons and reproducible research. In addition, the technical contributions of our work are two-fold: firstly, we implemented state-of-the-art methods for each stage of mammography image analysis. Secondly, we performed an experimental comparison to determine the ROI-detection strategies and feature extraction methods that yield the best performance. In order to foster the development of novel methods and facilitate their evaluation for adoption into clinical practice, all the developed tools will be made publicly available.

II. Computerized Analysis of Mammography Images

This section presents the general processing pipeline and main methods for computerized analysis of mammography images. As illustrated in Fig. 1, the four main steps for mammographic image analysis are *breast segmentation*, *region of interest (ROI) detection*, *feature extraction* and *risk scoring*. This section briefly the main steps of the pipeline. Implementation details can be found in the supplemental material[1].

A. Breast segmentation

Breast segmentation is aimed at the detection of the breast region. Depending on the view and type of mammography image, there are several tasks associated to breast detection: *background detection*, *chest wall detection* and *nipple detection*.

The aim of the first task, background detection, is separating the air region of the mammogram from the rest of the image. In this work, we implemented two different methods for two types of mammography images. For *full-field digital mammography* (FFDM) images, we implemented the method proposed in [5]. In that work, background is detected by simply thresholding the input image based on the highest mode of the intensity histogram. Due to image noise and image artifacts that arise during the scanning process, thresholding methods do not perform well with *digitized screen-film mammography* (SFM) images. Therefore, for SFM images we implemented the statistical method proposed by Liu *et al.* [6]. Briefly, pixel intensities in the image are treated as a random variable and the Anderson-Darling test is utilized to identify pixels in the foreground and background regions.

For the detection of the chest wall (both in FFDM and SFM images) as well as for detecting scanning artifacts in SFM images, we have utilized a Hough-based line detector [7]. The line detector works by first applying an edge detector to the input image. Subsequently, each edge pixel is represented into a parametric Hough accumulator space. Lines are detected as local maxima of the space histogram. Finally, the nipple is detected as the furthers contour point from the chest wall [8].

B. ROI-detection

After successful detection of the main anatomical landmarks of the breast (breast contour, chest wall and nipple) in Section II-A, the aim of ROI-detection is to detect specific regions inside the breast for subsequent feature extraction. The most straightforward approach is to utilize the whole breast region [9]. In this case, the output of the breast segmentation is directly utilized without further post-processing. Alternatively, feature extraction can be performed within specific regions inside the breast. For instance, several researchers have utilized manually-selected ROIs in the region behind the nipple, namely the retroareolar (RA) region [10]. However, the need for manual annotations or user interaction poses important limitations in terms of reproducibility and scalability of image analysis methods. In order to avoid the need for manual interaction, four fully-automated ROI-detection methods were implemented for this work (see Fig. 2): full-breast [9] (Full), largest square within the breast [11] (SQ), RA region [8] (RA), and multiple ROI's following a lattice-based sampling [12] (Multi).

[1]https://sites.google.com/view/cvia/openbreast

C. Feature extraction

The aim of feature extraction is to describe the visual texture patterns in the ROI in terms of quantitative, reproducible measurements. In the case of mammography breast images the most popular types of features are typically those used in texture analysis in computer vision. According to their working principles, the feature extraction methods for mammography images can be classified into five main groups: *statistical features* (STA), *gray-level co-occurrence features* (GLC), *gray-level run-length features* (GLR), *gradient-based features* (GRA) and *spatial-frequency analysis* (SFA).

Statistical features aim to describe the properties of the histogram of gray-level pixel intensities. The motivation behind these methods is that, since pixel intensities in the mammogram are related to the radio-density of the breast tissue, statistical properties of the intensity histogram may be associated with cancer risk. In the literature of breast cancer risk assessment statistical features have been widely utilized mainly due to their simplicity, fast computation and good performance. Gray-level co-occurrence features aim to describe the statistical distribution of co-occurring image intensities at given pixel distances and orientations [8], [11], [12]. These features are usually computed from the gray-level co-occurrence matrix, as proposed by [14]. Gray-level run-length features aim to describe the length and distribution of consecutive gray-level values in the image (run-lengths). Gradient-based features estimate the change of pixel intensities in the horizontal and vertical directions of the image. This is often measured by the computation of low-order statistical moments on the magnitude of the image gradient [8]. Finally, Spatial-frequency analysis aims at the computation of descriptors based on texture properties in the spatial domain, frequency domain or both. In the first case, spatial-domain features can be computed from linear filters. In the second case, frequency-domain features are used to describe the frequency content of the input image and are often computed directly from the Fourier spectrum [10], [12]. In the third case, image analysis in both the spatial and frequency domain is either performed implicitly by applying feature extraction at multiple scales, or explicitly by means of filter banks, *e.g.* Gabor filters or Wavelet transform [8], [10].

Since feature extraction is the most essential stage in our pipeline, we implemented 32 different feature extraction methods including all of the aforementioned categories. A detailed description of each method, as well as their utilization in previous works related to mammography image analysis can be found in the supplemental material.

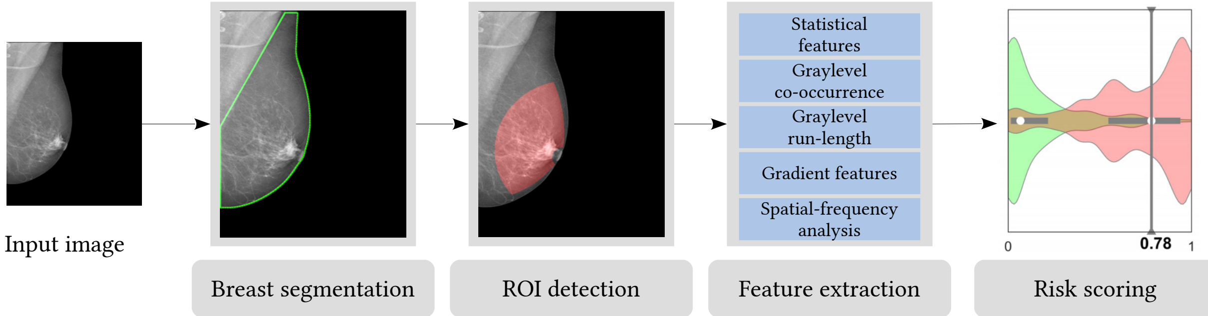


Fig. 1: Computerized risk assessment. In the first step, *breast segmentation*, the breast foreground region is segmented and pixel intensities are standardized. In the second step, *ROI detection*, a region of interest within the breast is identified for subsequent analysis. In the third step, *feature extraction* algorithms are applied to describe mammographic texture patterns. In the final step, a risk score r (x-axis) is used to determine the level of risk associated to the input image. The violin plots shown in red and green correspond to the scores of the high- and low-risk groups, respectively.

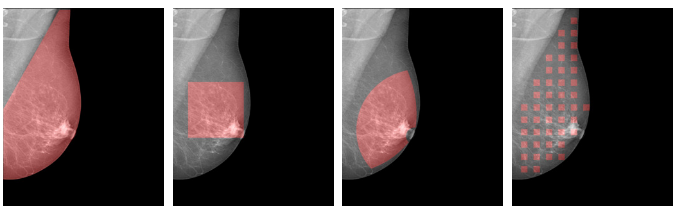


Fig. 2: Implemented ROI-detection methods. From left to right: full breast [9], largest square [11], retro-areolar region [13], and lattice-based sampling [12]

D. Risk scoring

Given a dataset of training mammography images for which the risk category is known (i.e., high risk or low risk), the aim of risk scoring is estimating the level of risk $r \in [0,1]$ of a test input image, where 0 represents the lowest risk and 1 is the highest risk. In this work, we use a logistic regression classifier with stepwise feature selection. Briefly, stepwise feature selection performs an exhaustive search over all available features and compares the performance of regression with and without the feature using t-test [15]. Features that yield a statistically significant improvement ($p < 0.05$) are included in the final model. For ease of interpretation and reproducibility, OpenBreast includes a visualization toolbox that shows the risk score of a test image with the probability density estimate, median and inter-quartile ranges of the scores of training images (right-hand-side in Fig. 1).

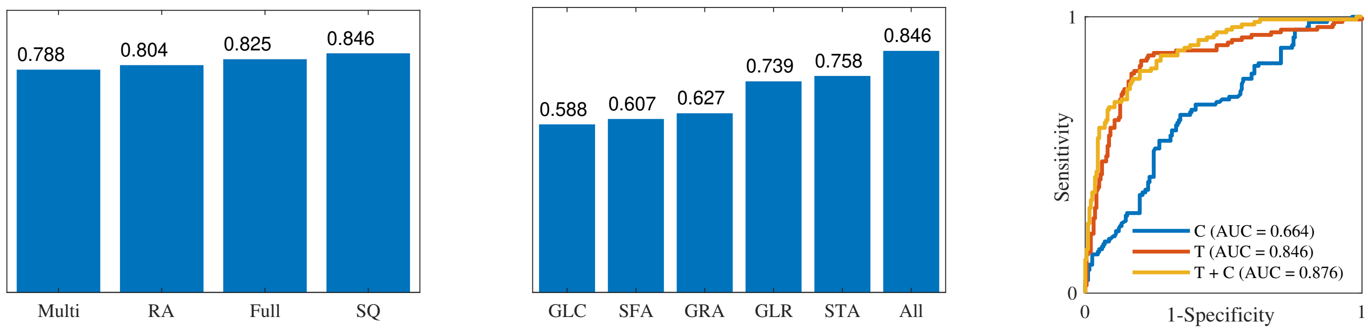
III. Performance Analysis

This section presents the results of the performance analysis using OpenBreast. The main goal is to perform an extensive experimental assessment of the methods implemented for each stage of the pipeline and analyze their effect on the overall performance of risk assessment. Our experiments differ from the previous works on texture-based breast cancer risk assessment (see [3] for details) in the sense that we experiment all combinations of the state-of-the-art methods for each processing stage and are able to find the best combination as evaluated under the same experimental conditions.

For the experiments, we used a dataset of 305 full-field digital mammography images from the Breast Cancer Digital Repository (BCDR [4]) corresponding to 84 patients: 39 cancer cases with clinically-proven malignancies (age 34-78, median: 60) and 45 healthy controls (age 34-82, median: 61). Specifically, we utilized sub-sets D01, N01 and N02 which are the ones that correspond to digital mammograms [4]. The images comprise both craneo-caudal (CC) and mediolateral oblique (MLO) views. Further details on the acquisition protocols and imaging data for this dataset are found in [4].

For the results reported in this work, experiments were performed following the pipeline shown in Fig. 1: automatic breast segmentation is followed by ROI detection, feature extraction and risk scoring, as described in section II. The whole pipeline with stratified 5-fold cross validation was repeated independently to evaluate each ROI-detection strategy with each group of feature extraction algorithms (20 different combinations). Performance was measured in terms of the area under the ROC curve (AUC) and differences were assessed using DeLong's test. The main results are summarized in Fig. 3. Additional results with the performance of individual features and correlation analysis can be found in the supplemental material.

As shown in Fig. 3a, when all feature extraction methods are considered, the best results were obtained using a squared ROI, yielding an AUC of 0.846. However, differences were only statistically significant when comparing feature extraction using a squared ROI vs. using multiple ROIs ($p = 0.047$). Fig. 3b shows the performance of risk assessment for different feature extraction strategies. In this case, a squared ROI was used since this method yielded the best overall performance. As expected, the best results were obtained when all feature extraction methods were included. If features from only one category are considered, then the best performance was ob-



(a) (b) (c)

Fig. 3: Obtained results. (a) AUCs for different ROI-detection methods using all available features. (b) AUC for different feature extraction methods using squared (SQ) ROI. (c) Performance of risk assessment using clinical data (C), texture analysis with OpenBreast (T), and texture analysis combined with clinical data (T+C).

tained using statistical features.

The inclusion of clinical data (i.e., patient age and breast density) yielded an improvement in performance with an AUC of 0.876 (Fig. 3c). This improvement was statistically significant compared to risk assessment using clinical data only (AUC = 0.664, $p < 0.001$) and texture analysis only (AUC = 0.846, $p = 0.038$). This suggests that texture analysis using OpenBreast has the potential to improve the performance of risk assessment in clinical practice.

IV. Conclusions

This paper presented an open computational framework and protocol for breast cancer risk estimation based on the fully-automated, computerized analysis of mammography images. The framework includes four main stages: breast segmentation, region-of-interest (ROI) detection, feature extraction and risk scoring. Experiments on a dataset of 84 patients (39 cases and 45 controls) from the Breast Cancer Digital Repository confirmed that parenchymal texture analysis can be successfully utilized as an imaging biomarker for breast cancer risk assessment, yielding a maximum AUC of 0.846 when the analysis is performed in the largest circumscribed square region withing the breast. Performed experiments also suggested that parenchymal analysis can be used jointly with clinical risk factors (breast density and age) yielding a statistically significant improvement in performance from an AUC = 0.664, using clinical data only, to an AUC = 0.876 using both clinical data and computerized texture analysis ($p < 0.001$).

In order to foster future research and allow for objective comparison of different approaches in breast cancer risk assessment based on parenchymal analysis, this work will be released to the public as OpenBreast. Future work will be aimed at the clinical validation of the OpenBreast toolbox and its extensions by incorporating novel feature extraction methods and more data.

References

[1] World Health Organization, “Cancer fact sheet n 297,” http://www.who.int/mediacentre/factsheets/fs297/en/, 2015, accessed: 02-12-2017.

[2] M. L. Giger, N. Karssemeijer, and J. Schnabel, “Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer,” *Annual Review of Biomedical Engineering*, vol. 15, pp. 327–357, 2013.

[3] A. Gastounioti, E. F. Conant, and D. Kontos, “Beyond breast density: a review on the advancing role of parenchymal texture analysis in breast cancer risk assessment,” *Breast Cancer Research*, vol. 18, no. 91, 2016.

[4] D. C. Moura and M. A. G. Lopez, “An evaluation of image descriptors combined with clinical data for breast cancer diagnosis,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 8, no. 4, pp. 561–574, 2013.

[5] B. M. Keller, D. L. Nathan, Y. Wang, Y. Zheng, J. C. Gee, E. F. Conant, and D. Kontos, “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,” *Medical Physics*, vol. 39, no. 8, pp. 4903–4917, 2012.

[6] L. Liu, J. Wang, and T. Wang, “Breast and pectoral muscle contours detection based on goodness of fit measure,” in *Int. Conference on Bioinformatics and Biomedical Engineering*, May 2011, pp. 1–4.

[7] D. Ballard, “Generalizing the hough transform to detect arbitrary shapes,” *Pattern Recognition*, vol. 13, no. 2, pp. 111 – 122, 1981.

[8] G. F. Torres and S. Pertuz, “Automatic detection of the retroareolar region in x-ray mammography images,” in *Latinoamerican Congress on Biomedical Engineering*, 2016, pp. 157–160.

[9] M. Nielsen, C. M. Vachon, C. G. Scott, K. Chernoff, G. Karemore, N. Karssemeijer, M. Lillholm, and M. A. Karsdal, “Mammographic texture resemblance generalizes as an independent risk factor for breast cancer,” *Breast Cancer Research*, vol. 16, no. 2, p. R37, Apr 2014.

[10] A. Manduca, M. J. Carston, J. J. Heine, C. G. Scott, V. S. Pakratz, K. R. Brandt, T. A. Sellers, C. M. Vachon, and J. R. Cerhan, “Texture features from mammographic images and risk of breast cancer,” *Cancer Epidemiology Biomarkers & Prevention*, no. 3, pp. 837–847, 2009.

[11] J. Wei, H.-P. Chan, Y.-T. Wu, C. Zhou, M. A. Helvie, A. Tsodikov, L. M. Hadjiiski, and B. Sahiner, “Association of computerized mammographic parenchymal pattern measure with breast cancer risk: A pilot case-control study,” *Radiology*, vol. 260, no. 1, pp. 42–49, 2011.

[12] Y. Zheng, B. M. Keller, S. Ray, Y. Wang, E. F. Conant, and D. Kontos, “Parenchymal texture analysis in digital mammography: a fully automated pipeline for breast cancer risk assessment,” *Medical Physics*, vol. 42, no. 7, pp. 4149–4159, 2015.

[13] G. Torres-Mejía, B. De Stavola, D. S. Allen, J. J. Pérez-Gavilán, J. M. Ferreira, I. S. Fentiman, and I. dos Santos Silva, “Mammographic features and subsequent risk of breast cancer: A comparison of qualitative and quantitative evaluations in the guernsey prospective studies,” *Cancer Epidemiology and Prevention Biomarkers*, vol. 14, no. 5, pp. 1052–1059, 2005.

[14] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, Nov 1973.

[15] N. R. Draper and H. Smith, *Applied Regression Analysis*. Wiley-Interscience, 1998.