Using Bayesian surprise to detect calcifications in mammogram images

Inês Domingues, Student Member, IEEE, and Jaime S. Cardoso, Senior Member, IEEE

Abstract— **Breast Cancer is still a serious health threat to women, both physically and psychologically. Fortunately, treatments involving complete breast removal are rarely needed today, as better treatment options are available. Mammography can show changes in the breast up to two years before a physician can feel them. Computer-aided detection and diagnosis is considered to be one of the most promising approaches that may improve the efficiency of mammography. Furthermore, there is a strong correlation between the presence of calcifications and the occurrence of breast cancer.**

In this paper we present a new technique to detect calcifications in mammogram images. The main objective is to support radiologists with automatic detection methods applied to medical images. Motivated by the fact that calcifications, when compared to the rest of the image, exhibit irregular characteristics, a technique based on Bayesian surprise is used.

Tests were performed using INBreast, a recent fully annotated database, composed of full field digital mammograms. Comparison both with a recently proposed state of the art method and other common image techniques showed the superiority of our method. False positives are, however, still an issue and further studies focused on their reduction while maintaining a high sensitivity are planned.

I. INTRODUCTION

Breast cancer is any form of malignant tumour which develops from breast cells. It is the most common cancer in women worldwide, and the leading cause of death from cancer in women, especially those between 40 and 55 years of age [1]. The only widely accepted imaging method used for routine breast cancer screening is mammography. Screening mammography is performed in asymptomatic population to detect early signs of breast cancer such as masses, calcifications, bilateral asymmetry and architectural distortion. Mammography reading is performed by radiologists who visually inspect mammograms. This is not an easy task and radiologists may get easily worn-out, missing vital clues while studying scans. For the calcifications case, they may be missed when they are covered by fundamental tissues of breast. Also, the location of calcifications in a region with a dense background is challenging [2]. In the light of the difficulties involved in manual screening, the search for automated screening of mammograms or computer aided detection and diagnosis (CAD) of breast cancer has been encouraged [1].

*This work is financed by National Funds through the FCT - Fundação para a Ciência e Tecnologia (Portuguese Foundation for Science and Technology) within projects PTDC/SAU-ENB/114951/2009 and SFRH/BD/70713/2010.

Inês Domingues and Jaime S. Cardoso are with INESC TEC - INESC Technology and Science and Faculty of Engineering, University of Porto, Portugal inesdomingues at gmail.com, jaime.cardoso at ieee.org

In this paper we address this need by proposing a Bayesian Surprise method for detection of calcifications in mammogram images.

II. RELATED WORK

The literature on calcification detection in mammograms is extremely extensive. Several review papers have already been published, from which we highlight [2], [3], [4], [5]. Here, only some selected recent works are reviewed.

A deep learning technique is used in [6]. After a local peak detection scheme, patches around the detected regions are manually classified as containing or not containing calcifications. A Discriminative Restricted Boltzmann Machine is then used to automatically learn calcification morphology and consequently classify new patches. Results using 9-fold cross validation on a private database of 33 mammograms reached an area under the ROC curve of 0.83.

Zhang *et al.* [7] propose a mathematical morphology and support vector machine (SVM) method. First, the contrast in the original mammogram is improved by gamma correction and two structural elements are used to enhance the calcifications. Next, the potential regions are extracted using a dual-threshold technique. Finally, an SVM classifier is used to reduce the number of false positives (FPs). The performance of the proposed method is evaluated using the MIAS database. The experimental results achieved a true positive rate of 94.85%, a FP rate of 7.82% and 0.53 FP calcifications per normal mammogram without calcifications.

Perhaps the most similar works to the one presented here are those that use novelty detection. The approach of Rose [8] is motivated by the fact that signs indicative of breast cancer are not found in pathology-free mammograms. The approach requires a model of what normal mammograms look like. Rose's thesis [8] presents two generative statistical models. The first treats mammographic appearance as a stationary texture. The second models the appearance of entire mammograms. Results in simulated calcifications achieved an area under the ROC curve of 0.92. The area under the ROC curve for images with simulated masses and calcifications was 0.75. When testing in images with real findings, the area dropped to 0.56 for images with calcifications only and to 0.53 for images with both masses and calcifications.

Concerning commercially available systems, a collection of selected independent assessments is presented in Table I. Differences in published results may be due to several factors, including system version, database and definitions used for True Positives (TPs), FPs and False Negatives (FNs). Baum *et al.* [9], considered as TP, only the automatic detections consistent with the histology result of a carcinoma,

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

TABLE I: Independent evaluation of commercially available systems.

System	Reference	Year	Sensitivity (%)	FPr
ImageChecker	[9]	2002	89	0.35
	[10]	2005	40	2.00
SecondLook	[11]	2005	95	0.65
	[12]	2013	49	0.22

whereas automatic marks in other areas were counted as FP. Likewise, only unmarked proven carcinomas were considered as FN. The results presented in [10] were obtained by using images with amorphous calcifications. Brem *et al.* [11] use cancer cases to compute sensitivity and normal images to calculate the FPr. In [12], only images with benign findings were used to calculate the FPr. Sensitivity was calculated by using both images with benign and malign findings. An image was considered as being correctly identified if at least one automatic detection was located within the ground truth region of the breast cancer, as opposed to the typical way of computing the True Positives (TPs), where all automatic detections are taken into consideration.

III. METHOLODOGY FOR CALCIFICATION DETECTION

The surprise caused by an observation is defined in a Bayesian sense as the change it brings to an observer's prior beliefs with respect to the phenomenon under consideration [13]. It can be mathematically defined as follows [14]: given a prior distribution $P(M)$ over a (discrete) space of models $\mathscr{M}$ describing a phenomenon, and the posterior distribution $P(M|D)$ after new data D is obtained for this phenomenon through an observation, the surprise incurred by D relative to the space $\mathscr{M}$ is given by the Kullback-Leibler divergence (K) between the prior and the posterior distribution,

$$S(D,\mathscr{M}) = K(P(M)||P(M|D)) = \sum_{M\in\mathscr{M}} P(M) log \frac{P(M)}{P(M|D)}. \quad (1)$$

Bayesian surprise can be used for images to explain the saliency of regions that, compared to the rest of the image, exhibit irregular characteristics [13]. This is particularly interesting for calcification detection since they usually correspond to bright spots in the mammogram image. The visual context of the region implies a prior distribution $P(M)$ over the model space before the region is observed. After the region is observed, a posterior distribution $P(M|D)$ is formed, where D is the data acquired from the observation of the region. The surprise incurred by D relative to the space $\mathscr{M}$ is then given by (1) [13]. This interpretation is demonstrated in Fig. 1, where the area surrounding a square region serves as its context.

The methodology for applying this technique to calcification detection is:

- for each patch of the image
  - compute the surprise induced by the patch in relation to its neighbourhood
    - if the surprise value exceeds a threshold
      - consider the region as a calcification

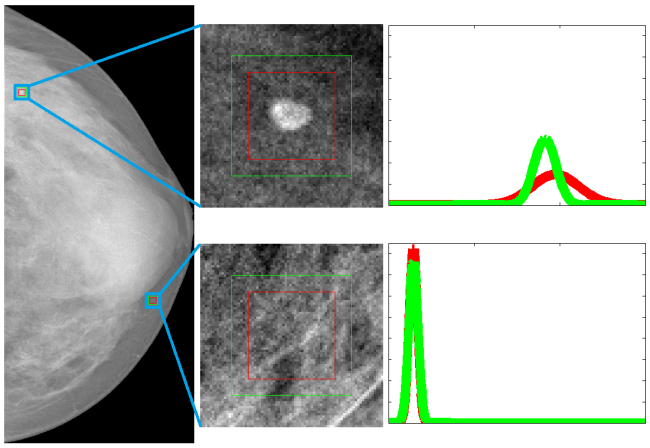

Fig. 1: Spatial interpretation of surprise for two regions (red squares) and their respective context (frames between red and green squares). Intensity distributions are also shown.

The technique is repeated in a set of 10 scales, by extracting patches of sizes ranging from 12×12 to 372×372 pixels. The context region was defined as the square frame having the same area as the patch. In order to make the process less computationally expensive, instead of using a dense grid, patches with 25% overlap were extracted. As there are, generally, more small calcifications than larger ones, surprise threshold values were set according to a linear relationship: $Surp_{th} = m\times Patch_{width} + b$. Parameters were experimentally set to $m = 30.5$ and $b = 178$. If, for a particular image, the number of detected calcifications exceeds 300, b is iteratively increased by 100 until the point when fewer than 300 calcifications are detected.

A. False Positive reduction

The above described detection technique tends to return a high number of FPs. To reduce the number of FP detections, a second step is performed, where some simple features are initially extracted and used in a classification algorithm. The feature vector of size 8 consists of the following characteristics: intensity value of the image at the detected point; standard deviation, minimum value, 25^{th} percentile, median value, mean value, 75^{th} percentile and the maximum value of the image intensities in the 21×21 patch around the detected value. An SVM with the Radial Basis Function kernel was trained in randomly selected subset containing 75% of the images in the dataset. This process is repeated 40 times in order to achieve more stable results.

IV. EVALUATION

Most of the evaluation methodologies individually look at each lesion in the ground-truth (GT) and find a corresponding lesion on the detection results. There is, however,

a significant drawback with this approach, since it does not force a one-to-one relationship and several detected regions may be wrongly associated to the same GT lesion. Here, the approach in [15] is followed. The main steps are as follows:

- Dissimilarity Matrix computation: calculation of all the costs between the GT findings and the detected regions;
- Optimal assignment calculation: assignment of exactly one GT region to one and only one detected region in such a way that the total cost of the assignment is minimized (in the current implementation, the Hungarian Algorithm [16] is used).

For the calcification detection case, as there are small structures, they can be approximated by the corresponding centroid. In this way, the most natural dissimilarity metric to use is the distance. A saturation threshold is then applied to the dissimilarity matrix in order to minimize an incorrect assignment [15].

After the optimal assignment calculation, the number of FPs is automatically determined by observing the detected regions with no corresponding GT region. When dealing with clusters, a final stage is added where the FP detections are checked to see if they fall inside the GT cluster region. If they do, they are no longer considered FPs. The missed detection cases are assessed by setting an upper threshold in the allowed dissimilarity value.

V. RESULTS

The experiments presented in this section are made using the INBreast database [15]. All images were automatically pre-processed so that the nipple is facing the right side of the image; images have been automatically cropped in order to include the breast area only; and pectoral muscle has been removed. Images with BI-RADS class 1 have been manually screened out, resulting in a total of 343 mammograms to be analysed. Calcifications have areas ranging from 1 pixel^2 to more than 20 000 pixel^2. For comparison with our proposed Bayesian surprise detection method, other techniques were tested:

Fixed threshold: As calcifications are usually brighter than the remaining breast tissue, a first naive approach is to determine a fixed value (threshold) and classify all image pixels with intensity above that value as suspicious. The problem resides in the determination of a unique robust threshold value that works for all the images.

Outlier detection: Instead of using the same threshold for all images, it can be adjusted individually for each image. In this technique, all points with intensity higher than $q_3 + 1.5 \times (q_3 - q_1)$ where q_1 and q_3 are the 25^{th} and 75^{th} percentiles, are considered as outliers and thus as calcifications. The value 1.5 corresponds to approximately 99.3% coverage if the data (i.e. pixel intensities) are normally distributed [17].

Mathematical morphology: For comparison with the state of the art, the technique of Zhang *et al.* [7] based on mathematical morphology was reimplemented. The FP reduction part was not performed since, as will be shown in the experimental section, this method has a low sensitivity. Reducing the FPs would decrease the sensitivity even more.

From Table II it can be seen that the method with the worst sensitivity is the Fixed threshold, followed by Mathematical morphology, Outlier detection and the highest sensitivity is achieved with the Bayesian surprise. Concerning the FPs, the method with the worst behaviour is Mathematical morphology, followed by Fixed threshold, Bayesian surprise and Outlier detection.

TABLE II: Calcification detection results

Method	Sensitivity (%)	FP
Fixed threshold	34.5	164
Outlier detection	45.8	60
Mathematical morphology	40.3	225
Bayesian surprise	60.3	108

A comparative plot of the performances in several operating points can be seen in Fig. 2. The different behaviours were accessed by: (1) in the Fixed threshold method, the threshold was varied between 1.0 and 0.8; (2) for the Outlier detection technique the computed individual threshold was varied by adding a constant between -0.50 and 0.50; (3) the Mathematical morphology method has two parameters, α and β which were varied in the interval between 0.0 and 1.0; (4) finally, in the Bayesian surprise, the parameter b was changed between -2000 and 2000. The curves show that Bayesian surprise has a poor performance for small sensitivity values but it rapidly performs better than all the other tested methods as sensitivity increases. Illustrative examples are presented in Fig. 3.

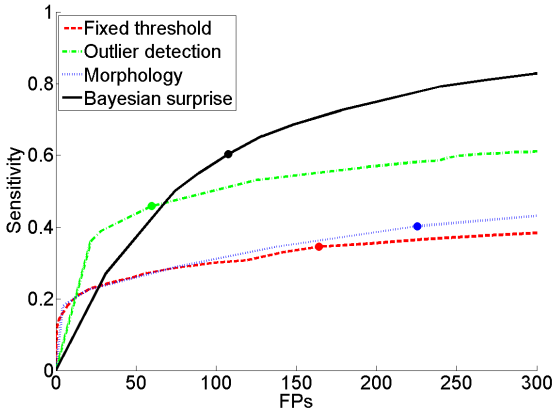

Fig. 2: Comparative plot of the performances of the detection methods. Circles correspond to the operating points in Table II.

To test the FP reduction, an "over-detection" was first performed leading to a Sensitivity of 83.9% with an average number of 355 FP detections per image. After the reduction of FPs, the sensitivity of the Bayesian surprise method decreased to 48.8% (standard deviation $= 0.143$) and FP to 87 (standard deviation $= 46$). Some examples are shown in Fig. 4.

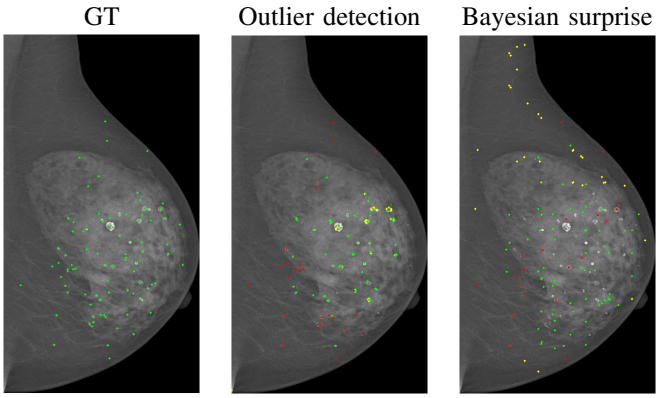

Fig. 3: Examples of Calcification detection. Green dots correspond to TPs, yellow to FPs and red to FNs.

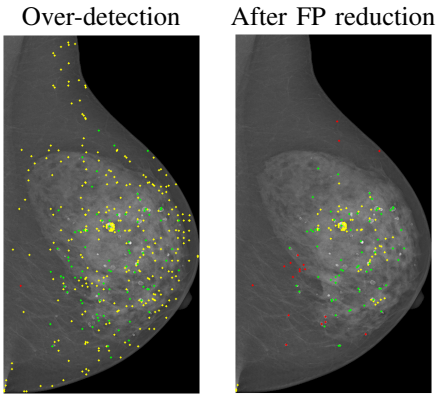

Fig. 4: Examples of Calcification detection with FP reduction. Green dots correspond to TPs, yellow to FPs and red to FNs.

Although a direct comparison with the results in Table I is not possible, due to differences in the databases, GT specification methodology, evaluation technique, etc., the results for the Bayesian Surprise method compare well in terms of sensitivity with the commercially available approaches. As illustrated with the examples of Fig. 4, the number of FPs for the Bayesian Surprise method is overestimated. This is a direct consequence of the sparse sampling. A dense grid would generate connected areas and thus reduce the number of FP. Other techniques for FP reduction that do not have such a strong impact in the sensitivity will be tested in the future.

VI. CONCLUSIONS

A methodology not yet used for calcification detection is suggested in the present work. Results, when compared with other state of the approaches, shown a high sensitivity. This comes at the cost of a large amount of false positives. False positive reduction is one of the envisioned future works.

Another contribution of this paper is to provide baseline detection results over the INBreast database. To the best of our knowledge, no other work has yet been made on the detection of calcifications in this database. Future proposals can thus be validated by comparison with the results presented here.

REFERENCES

[1] K. Ganesan, U. R. Acharya, C. K. Chua, L. C. Min, K. T. Abraham, and K. Ng, "Computer-aided breast cancer detection using mammograms: a review," *IEEE Reviews in Biomedical Engineering*, vol. 6, pp. 77–98, 2013.

[2] A. Jalalian, S. B. T. Mashohor, H. R. Mahmud, M. I. B. Saripan, A. R. B. Ramli, and B. Karasfi, "Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review," *Clinical Imaging*, vol. 37, no. 3, pp. 420–426, 2013.

[3] M. Rizzi, D. Matteo, and B. Castagnolo, "Review: Health care CAD systems for breast microcalcification cluster detection," *Journal of Medical and Biological Engineering*, vol. 32, no. 3, 2012.

[4] I. El Naqa and Y. Yang, "Techniques in the detection of microcalcification clusters in digital mammograms," in *Medical Imaging Systems Technology*. World Scientific, 2005, pp. 15–36.

[5] H. D. Cheng, X. Cai, X. Chen, L. Hu, and X. Lou, "Computer-aided detection and classification of microcalcifications in mammograms: a survey," *Pattern Recognition*, vol. 36, no. 12, pp. 2967–2991, 2003.

[6] S.-Y. Shin, S. Lee, and I. D. Yun, "Classification based microcalcification detection using discriminative restricted boltzmann machine in digitized mammograms," in *SPIE Medical Imaging*, vol. 115, 2014, p. 151.

[7] E. Zhang, F. Wang, Y. Li, and X. Bai, "Automatic detection of microcalcifications using mathematical morphology and a support vector machine," *Bio-medical materials and engineering*, vol. 24, no. 1, pp. 53–59, 2014.

[8] C. J. Rose, "Statistical models of mammographic texture and appearance," Doctor of Philosophy in the Faculty of Medical and Human Sciences, University of Manchester, 2005.

[9] F. Baum, U. Fischer, S. Obenauer, and E. Grabbe, "Computer-aided detection in direct digital full-field mammography: initial results," *European radiology*, vol. 12, no. 12, pp. 3015–3017, 2002.

[10] M. S. Soo, E. L. Rosen, J. Q. Xia, S. Ghate, and J. A. Baker, "Computer-aided detection of amorphous calcifications," *American Journal of Roentgenology*, vol. 184, no. 3, pp. 887–892, 2005.

[11] R. F. Brem, J. W. Hoffmeister, J. A. Rapelyea, G. Zisman, K. Mohtashemi, G. Jindal, D. M. P, and S. K. Rogers, "Impact of breast density on computer-aided detection for breast cancer," *American Journal of Roentgenology*, vol. 184, no. 2, pp. 439–444, 2005.

[12] M. Lobbes, M. Smidt, K. Keymeulen, R. Girometti, C. Zuiani, R. Beets-Tan, J. Wildberger, and C. Boetes, "Malignant lesions on mammography: accuracy of two different computer-aided detection systems," *Clinical Imaging*, vol. 37, no. 2, pp. 283–288, 2013.

[13] I. Gkioulekas, G. Evangelopoulos, and P. Maragos, "Spatial bayesian surprise for image saliency and quality assessment," in *17th IEEE International Conference on Image Processing*, 2010, pp. 1081–1084.

[14] L. Itti and P. F. Baldi, "Bayesian surprise attracts human attention," in *Neural Information Processing Systems (NIPS)*, 2005, pp. 547–554.

[15] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: towards a full field digital mammographic database," *Academic radiology*, vol. 19, pp. 236–248, 2012.

[16] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1–2, pp. 83–97, 1955.

[17] P. F. Velleman and D. C. Hoaglin, *Applications, basics, and computing of exploratory data analysis*. Boston, Mass.: Duxbury Press, 1981.