

# Normal breast identification in screening mammography: a study on 18 000 images

Sílvia Bessa\*, Inês Domingues\*<sup>†</sup>, Jaime S. Cardoso\*<sup>†</sup>,  
Pedro Passarinho<sup>‡</sup>, Pedro Cardoso<sup>‡</sup>, Vítor Rodrigues<sup>§</sup>, Fernando Lage<sup>¶</sup>

\*Faculdade de Engenharia, Universidade do Porto, Portugal

<sup>†</sup>INESC TEC, Porto, Portugal

<sup>‡</sup>Emílio Azevedo Campos, Portugal

<sup>§</sup>Universidade de Coimbra, Portugal

<sup>¶</sup>Liga Portuguesa Contra o Cancro, Portugal

**Abstract**—Through the years, several CAD systems have been developed to help radiologists in the hard task of detecting signs of cancer in the numerous screening mammograms. A more recent trend includes the development of pre-CAD systems aiming at identifying normal mammograms instead of detecting suspicious ones. Normal breasts are screened-out from the process, leaving radiologists more time to focus on more difficult cases.

In this work, a new approach for the identification of normal breasts is presented. Considering that even breasts with malignant findings are mostly constituted by normal tissue, the breast area is divided into blocks which are then compared pairwise. If all blocks are very similar, the breast is labelled as normal, and as suspicious otherwise. Features characterizing the pairwise block similarity and characterizing the intra-block pixel distribution are used to design a predictive method based on machine learning techniques. The proposed solution was applied on a real world screening setting composed by nearly 18000 mammograms. Results are similar to the more complex state of the art approaches by correctly identifying more than 20% of the normal mammograms. These results suggest the usefulness of the relative comparison instead of the absolute classification. When properly used, simple statistics can suffice to distinguish the clearly normal breasts.

## I. INTRODUCTION

Screening mammography is performed in the asymptomatic population to detect early signs of breast cancer such as masses, microcalcifications (MCs), bilateral asymmetry and architectural distortions (AD). Diagnostic mammography is performed on patients who have already demonstrated abnormal clinical findings [1]. Both screening and diagnostic mammography are performed by radiologists who visually inspect mammograms [1]. This is not an easy task: mammograms generally have low contrast. Mammograms show normal structures such as fat, fibroglandular tissue, breast ducts and nipples, as well as possible abnormalities. Although fat appears as black regions on mammograms, everything else appear as levels of white, making it hard to distinguish between normal and abnormal tissue [2]. Furthermore, due to the advised two year interval screening routines for women between 45 and 69 years old (taking Portugal as an example), there is a large number of mammograms to be analysed every day. As a consequence, during manual screening radiologists may get easily worn-out, missing vital clues while studying the scans.

Supporting these facts, studies have shown radiologists have an error rate of 10% to 30% for detection of cancer in

screening mammograms [1], [3]. As a result, a high proportion of women without cancer (false positive (FP) cases) undergo further unnecessary clinical evaluation or breast biopsy which can lead to needless anxiety. Moreover, in false negative (FN) cases, the best time interval for the treatment of cancer can be missed, thus potentially endangering the patient. In either case, the costs associated with errors of misclassification are considerable, including emotional and economic burden, or even loss of life. To overcome these limitations, double reading of mammograms has been advocated. The idea is to have two radiologists reading the same mammographic images. This has been shown to increase sensitivity (equivalent to decrease the number of FN). However, the workload and cost associated with double reading remain high, and still susceptible to human error.

The assumption made in this work is that, in screening, a substantial proportion of normal cases can be automatically detected, alleviating the human effort and giving the specialist more time to carefully evaluate more ambiguous cases. Since, in general, only 0.5% of mammograms present an anomaly [4], [5], pre-CAD systems have a high potential in screening programs. Although several Computer-Aided Detection and Diagnosis (CAD) systems have been developed and studied, little research has been done on the development of screening systems (also known as Pre-CADs or Normal mammogram detection or analysis).

## II. STATE OF THE ART

Sun et al. [4] propose a two-stage system for normal mammogram identification. After a pre-processing stage, four types of features are extracted from overlapping blocks of size  $512 \times 512$ , namely curvilinear features, texture features, Gabor features and Multi-resolution features. Each block is then classified by a Constrained Binary Decision Tree. If the result is not abnormal, the block is classified as normal. Otherwise, a second linear classifier is used to further distinguish between suspicious and normal blocks. As there is an overlap between blocks, each pixel is classified five times. A majority voting is used to make the final pixel decision. Finally, the full image is considered as suspicious if one or more pixels are abnormal, otherwise, the mammogram is classified as normal. Experiments made on Digital Database for Screening Mammography (DDSM) retrieved a True Positive rate (TPR) of 0.82 and a True Negative rate (TNR) of 0.75.

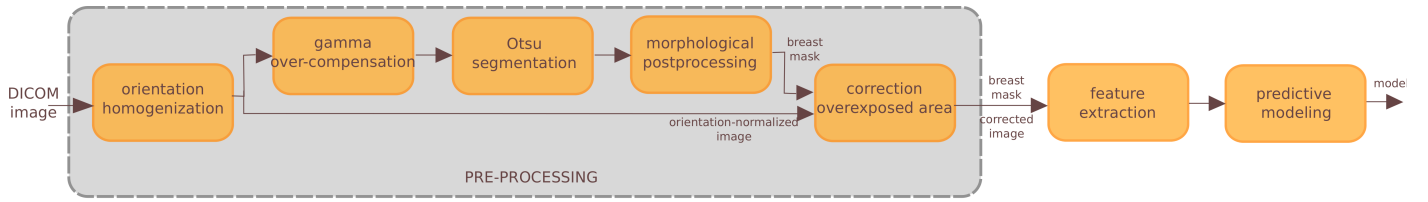


Fig. 1: Block diagram of the proposed methodology.

In [5], a traditional two-class classifier is compared with a one-class classifier. For the one class classifier, only the normal features are used for training. The assumption behind the use of the one class classifier approach is that the normal mammograms have less variability when compared to the abnormal mammograms which have a very wide variation of lesion type and shape. Classification was made per block and a majority voting approach was used to classify the full breast by combining CranioCaudal (CC) and MedioLateral Oblique (MLO) information. Results on the DDSM database show that separating the mammograms according to density reduces the False Negative rate (FNr) while keeping the False Positive rate (FPr) as low as possible. Moreover, the one-class SVM outperformed the two-class SVM. After settling the appropriate classifier to use, the team studied the feature extraction step. In [6], three different sets of textural features based on Local Binary Patterns (LBP) were assessed. GrayLevel Co-Occurrence Matrix (GLCM) features were studied in [7]. Finally, in [8], LBP and GLCM features were compared. It was observed using DDSM that GLCM is a good feature for fatty-tissue, while LBP behaves better in dense-tissue mammograms. Moreover, the sensitivity for each tissue type was improved when compared to the sensitivity of using all mammograms regardless of tissue type. Results of the recall rate varied from 97.34% to 99.75%.

The paper [9] presents a one-class classification pipeline for the classification of mammogram images into normal and abnormal. Because of the sparse distribution of abnormal mammograms, the two-class classification problem is reduced to a one-class outlier identification problem. Trace transform, which is a generalization of the Radon transform, is applied to extract the features. Using a private database, an accuracy rate of 92.48% is obtained with a Gaussian Mixture Model (GMM) classifier.

In the study [10], three combinations of wavelet and Fourier features, including Db2, Db4, and Bior 6.8 are tested. Classification into normal or suspicious using a private database is made with three classifiers, including Back-propagation Network, Linear Discriminant Analysis, and Naive Bayes. Best results are attained using Bior and Fourier Features and back-propagation Network: TP= 500, FP= 21, FN= 21, TN= 128, Sensitivity= 95.9%, Specificity= 85.9%, Accuracy= 93.7%.

### III. PROPOSED METHODOLOGY

Typically the attempts to recognize normal cases in the screening setting divide the breast region in blocks, each block is individually classified as benign or suspicious. Only when all the blocks are marked as normal is the case marked as such.

In these approaches, it is necessary to learn a robust block classifier that works under the typical variability observed in this application. This is far from being an easy task, and researchers resort to complex descriptors [9] and complex decision models [10].

In here we follow a different approach. We argue that even in a mammogram with malignant findings, most of the breast tissue is normal. Therefore, we split the breast in different blocks and compare them pairwise. If all the blocks are very similar to each other, we assume to be in the presence of a benign case; if a high dissimilarity between two blocks is found we mark the case as suspicious. By comparing blocks pairwise inside the same image, the variability observed due to external causes will impact all blocks and tend to cancel out, hopefully simplifying the decision task. The breast labeling as benign or suspicious is based on the joint analysis of the MLO and CC views, and all operations are applied only inside the breast ROI. A block diagram of the main steps of the proposed methodology is presented in Figure 1.

#### A. Pre-Processing

Pre-processing is the first step of every traditional image processing pipeline. In the case of mammograms, typical pre-processing techniques include: noise reduction, image enhancement, background exclusion, orientation homogenization, pectoral muscle identification, among others [11], [12].

Concerning background exclusion and orientation homogenization, standard techniques were used. First, the image laterality information stored in the DICOM header is used to orientation homogenization: those images where the nipple faces left are vertically mirrored. The breast area is selected using Otsu's threshold applied not directly to the image but to a gamma over-compensated image to emphasize the differences between background and breast area (Figure 2b). The breast area is considered to be the largest blob in the resulting binary image. An opening morphological operation is at last applied to eliminate undesired connections between the breast ROI and background non-homogeneous areas. Finally, the pectoral muscle segmentation task is accomplished by an algorithm based on the shortest path on a graph [12]. Figure 2c show the resulting breast mask.

Although enhancement techniques usually figure in the pre-processing modules, the use of Full Field Digital Mammograms (FFDM) usually circumvents the need to use them. However, digital mammograms may present an overexposed area in the peripheral part of the breast characterized by lower intensities and contrast. This happens because the breast is pressed during the mammography exam resulting in non-uniform thickness across the breast, which is particularly

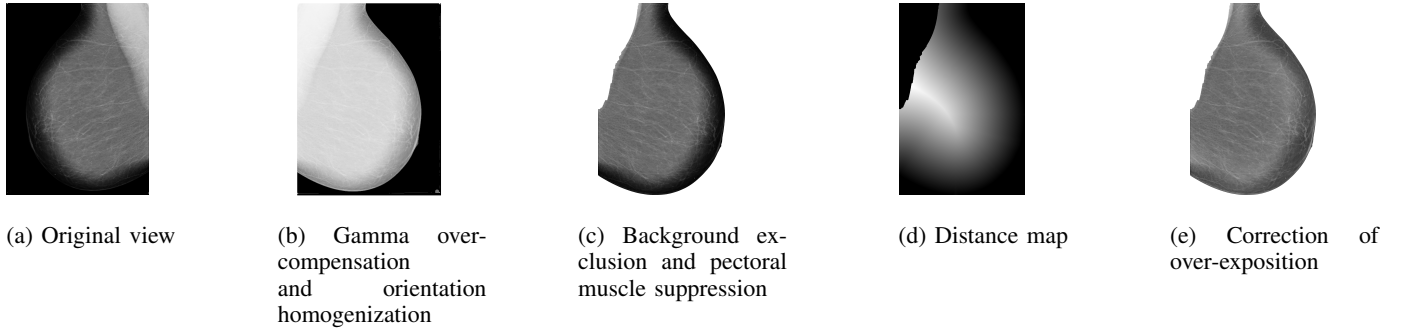


Fig. 2: Orientation homogenization, background exclusion, and pectoral muscle suppression examples for a right MLO view.

thinner in the peripheral area [13]. In here, the correction of peripheral area intensities is accomplished by adaptive gamma intensity correction, with gamma values weighted by the pixels distances to the skin line. To do so, the Euclidean distance map is obtained by computing the distance  $d$  of each pixel to the skin line (Figure 2d). The gamma correction factor for each pixel intensity is then obtained as a sigmoid function of its distance to the skin line:  $\gamma = \frac{1}{1+k_1 \exp(-k_2 d)}$ , where  $k_1$  and  $k_2$  are constants controlling the strongest gamma correction and the decrease rate of the correction with the distance to the skin line. Therefore, pixels in the peripheral areas will be assigned lower gamma values, meaning that they will be more strongly corrected to brighter values. Figure 2e shows the mammogram from Figure 2 after the correction of the over-exposition.

### B. Feature Extraction

To assess the benignancy of a mammogram, the breast ROI is divided in overlapping blocks, both horizontally and vertically (see Figure 3). For each view, the distance between each pair of blocks histograms is calculated as a measure of similarity between them. Given  $h_i$  and  $h_j$ , the cumulative histograms of two distinct blocks  $i$  and  $j$ , normalized to have unitary area, the distance  $D(i, j)$  between them is calculated as follows:

$$D(i, j) = \sum_{n=1}^{\text{nbins}} |h_i(n) - h_j(n)|, \quad (1)$$

where nbins corresponds to the number of bins in the histogram (255 in our work). We also record for each horizontal block the median and the 95<sup>th</sup> and 5<sup>th</sup> percentiles. The difference between the 95<sup>th</sup> and 5<sup>th</sup> percentiles ( $p_{95} - p_5$ ) is a measure of the spread of the intensities, which is expected to be higher on non-homogeneous blocks such as the ones containing malignant findings.

### C. Predictive Modeling

In the design of a predictive model we considered both a fully data driven approach and a design driven by expert knowledge.

1) *Data Driven (DD) Methodology*: In a fully data driven methodology we inputted all the 15 + 15 + 6 + 6 features (CC pairwise blocks distance + MLO pairwise blocks distance + block medians + block intensity spread) to a machine

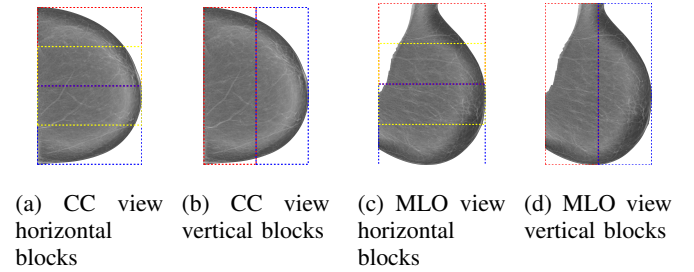


Fig. 3: Feature extraction methodology: breast area is divided in vertical and horizontal overlapping blocks.

learning algorithm armed with a feature selection technique. We evaluated a generative approach (a Gaussian model - GM, both in the heteroscedastic and the homoscedastic versions), a probabilistic discriminant model (logistic regression - LR) and a non-probabilistic model (SVM), both the two-class and the one-class classifier versions). For feature selection we considered the mRMR (minimum redundancy maximum relevance) method [14].

Since the benign breast identification is both an example of cost-sensitive learning [15] (the two possible types of errors have quite different costs) and class imbalance (one class occurs much more often than the other), the straightforward option of deciding for the class with the highest a posteriori probability does not lead to an acceptable performance.

The identification of normal breasts is intended to function as a “first look”, automatically screening out normal mammograms without further human analysis. Therefore, it is of major importance to keep the number of FN near to zero even if it compromises a higher percentage of normal breasts correctly marked as such. In the case of a FN, an early breast cancer can be missed jeopardizing the best interval time to treat it with success, which is exactly the opposite of the purpose of screening routines. As such, the design should be driven to an operating point where the error in the suspicious cases is *low enough*. Therefore, for all the three design options, we start by estimating the a posteriori probability for the classes. For the generative model, the a posteriori probability is computed using Bayes formula; the LR outputs directly the posteriori probability; for the SVM, the output is transformed

using a log-sigmoid to provide estimations for the a posteriori probabilities. Then the threshold decision can be chosen to set the operating point in a suitable value.

Quite often, the performance of the models is analyzed with ROC curves (obtained by just varying the threshold decision), and measures extracted from it (like the area under the curve - AUC). However, the analysis of the complete curve is of no interest for us; only the operating points close to one of the endings of the curve are interesting.

Finally, the performance of the classification task is evaluated both in terms of breast and case classifications. A case is classified as normal if both breasts are normal or suspicious if otherwise. Once more, TNr and FNr are used as performance metrics, where negative stands for normal and positive for suspicious breasts or cases.

2) *Expert Driven (ED) Methodology* : Brighter areas may be malignant findings or be associated with areas of the breast with denser tissue, which have also been related to increased probability of breast cancer [16]. As such, we considered using information only from the view containing the block with the highest median value. The blocks histograms' distances of this view are computed and inputted to the machine learning classifier. Two options were considered: using the maximum distance or the variance between all distances of the selected view. The effect of the number of blocks was also considered. In either case, the learning method received a second feature corresponding to  $p_{95} - p_5$ , the difference between the 95<sup>th</sup> and 5<sup>th</sup> percentiles of the block with the highest median value. Since only two features are used in this methodology, no feature selection mechanism is needed. For the design and evaluation of the classifiers we followed the same protocol as in the data driven methodology.

#### IV. EXPERIMENTAL VALIDATION

##### A. Dataset

The mammograms used in this work were acquired by several Liga Portuguesa Contra o Cancro' units during the screening routines carried on at the district of Évora, Portugal. Images from screening cases between 2008 and 2013 were used. A total of 4 485 cases were collected, all having 2 images (CC and MLO views) of each breast. This sums to a total of 8 970 breast scans and 17 940 mammograms. All the mammograms are FFDM and were saved in the DICOM format.

This LPCC dataset includes normal mammograms and mammograms with malignant findings such as masses, MCs and AD. The frequency of normal and suspicious cases is in accordance with their usual distribution on the screening setting: only 27 cases (0.6%) of the cases are cancer, corresponding to 28 breasts with suspicious findings.

##### B. Results

When analyzing per breast, the best results for each learning method, for the DD methodology, are plotted in Figure 4. As can be observed, the performance of the GM and the LR methods is very similar: both can identify 16% of the normal breasts without misclassifying any suspicious breast. The SVM seems to be the most promising though, achieving

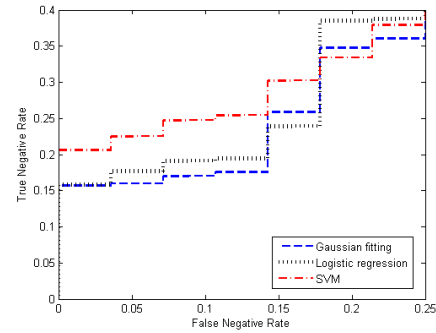


Fig. 4: Results for the GM, the LR and the SVM under the data driven methodology.

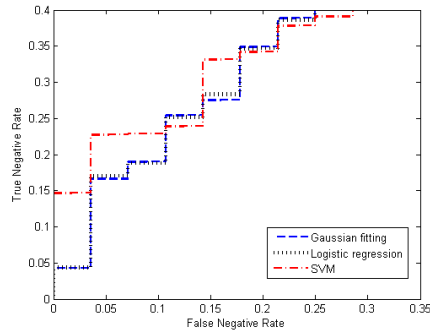
an automation rate of 21% without any FN (results obtained for the binary SVM with a polynomial kernel of second degree). However, accepting a small FNr does not increase significantly the TNr of all models.

Proceeding similarly for the ED methodology we obtained the results in Figure 5. The best results were obtained when using the maximum distance instead of the variance between all distances of the selected view. For both block division strategies, namely using only horizontal blocks (Figure 5a) or both horizontal and vertical blocks (Figure 5b), SVM seems again to be the best option, providing better operating points. Using horizontal blocks only, it enables to automate 15% of the normal breasts, without misclassifying any suspicious breast; an automation rate of 23% is obtained when accepting a FNr 3.5%. When using vertical and horizontal blocks, the results appear to be better. Although the automation rate is nearly the same without misclassifying any suspicious breast, 15%, when accepting a FNr of 3.5%, the automation rate increases to 30%. Once again, these results were obtained for the binary SVM with a polynomial kernel of second degree. The division in vertical blocks only provided poor results (data not shown).

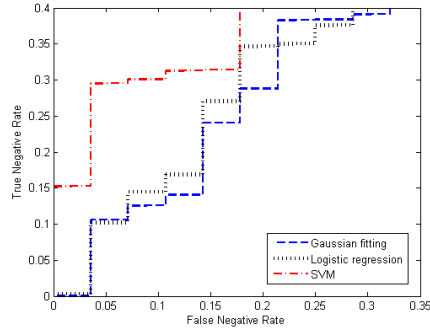
Comparing the DD and ED methodologies, the ED models seem to have better results. Although the performance of the GM and LR is better for the DD methodology (even matching the operating point of the SVM ED methodology when no FN is accepted), the SVM results were generally better in both methodologies. Therefore, comparing the SVM results for both methodologies, the DD methodology has a better performance for a FNr=0% (21% against the 15% achieved with the ED one). On the contrary, when a FNr of 3.5% is accepted, the ED methodology has a promising result of 30% of automation rate, while the increase in the TNr for the DD methodology is not significant.

It is now instructive to analyze the suspicious breasts that are misclassified when one starts increasing the automation rate. Figure 6 shows the suspicious breast that is misclassified as normal when trying to increase the automation rate of the ED models. This breast contains MCs, tiny deposits of calcium that appear as small bright spots on mammograms. In fact, it was expected that cases of MCs could fail to be correctly automatized due to their small influence in histogram related features such as ours. Furthermore, while a mass is





(a) Horizontal blocks only.



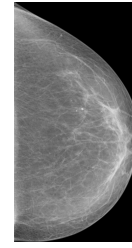
(b) Vertical and horizontal blocks.

Fig. 5: Results for the GM, the LR and the SVM under the expert driven methodology.

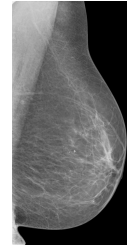
by definition a space-occupying lesion seen in at least two different projections, MCs may appear only in one view. As a consequence, regardless of the selection of a particular view to decide upon the benignancy of the breast, masses are expected to impact both views, while MCs could be discarded in selecting a particular view as we do. Therefore, our methodology must be combined with an approach to detect these high spatial frequencies corresponding to MCs to safely increase the automation rate.

Another important evaluator of our pre-CAD methodology is the performance regarding BIRADS 1 breasts. While BIRADS 2 breasts are benign but contain non-cancerous findings, BIRADS 1 breasts look all the same (they are symmetrical) with no masses, distorted structures, or suspicious calcifications found. Therefore, it is important to unravel why our methodology is failing to classify part of the BIRADS 1 breasts as benign. There are three main explanations for that: (i) when selected, the CC view may contain muscle, which introduces dissimilarities that are not associated with masses or other lesions (Figure 7a); (ii) the breast parenchymal tissue is unevenly distributed through the breast, which causes dissimilarities among breast blocks to be significant although no lesion is present (Figure 7b); and (iii) when selected, the MLO view has a defective suppression of the pectoral muscle, which has the same effect of (i) (Figure 7c).

The analysis so far has been done per breast. However, in the clinical set, the presence of early signs of breast cancer in a

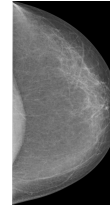


(a) CC view

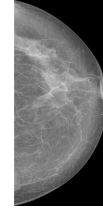


(b) MLO view

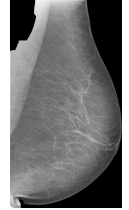
Fig. 6: Example of a breast containing microcalcifications misclassified as normal.



(a) CC view contains muscle



(b) Breast density is unevenly distributed



(c) Pectoral muscle is sub-segmented in MLO view

Fig. 7: Examples of a BIRADS 1 breasts misclassified as suspicious.

patient is judged by case (the case classification is the higher BIRADS classification between the two breasts). Therefore, the performance of the ED SVM model in terms of case classification is also an important result. Focusing on the same two operating points discussed before, 8% of the cases are correctly identified as normal without any FN; accepting a FNr of 3.8% (one suspicious case is misclassified), the case automation rate increases to 19%. Comparing the case and breast automation rates, the lower case automation rates were expected, particularly in the case of FNr=0%. Even so, the TNr of 19% obtained with the misclassification of only one suspicious case is a satisfactory result.

In a final experiment, we added to our database the malignant cases of the INBreast database [17]. This allowed us to increase the number of malignant cases, and to assess in more detail the robustness of the proposed methodology. Figure 8 shows the results of the SVM ED model for the LPCC database and the LPCC plus the 50 malignant cases of the INBreast database. As can be seen, the overall results are better for the LPCC database. Nevertheless, for TNr higher than 15% and lower than 20%, the FNr is similar for both databases: 3.51% and 3.84% for LPCC and LPCC+INBreast databases, respectively. This suggests that the proposed methodology may be extended to different databases, while keeping acceptable TNr and FNr. However, it is worth noticing that the INBreast database has a high incidence of cases with MCs (more than half of the malignant cases contain MCs findings alone) which, as previously discussed, is still a limitation of the

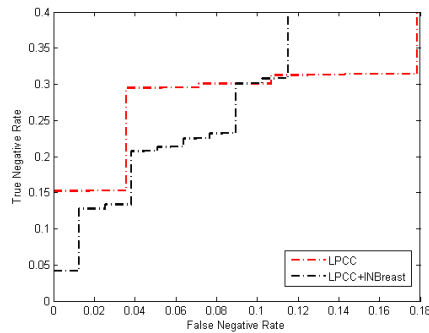


Fig. 8: Results of the best predictive model (expert driven binary-SVM with vertical and horizontal blocks) after adding the malignant cases of INbreast database.

proposed methodology. In addition, the INBreast database contains mammograms other than screening ones. Although normal identification should in theory work as well in all types of mammograms, follow-ups and diagnosis mammograms can contain specially hard cases. In some cases, multiple masses are found in the same mammogram, or a unique mass occupies most of the breast area. Therefore, comparing blocks of the breast area may result in small dissimilarities, resulting in the breast misclassification as normal. Other cases include particularly difficult subtle signs of bilateral asymmetries and AD to which our methodology is not well appropriated.

## V. CONCLUSION

In the proposed methodology, the identification of normal breasts is formulated as problem of similarity between blocks of the breast area. Information from both mammographic views is used: the view selected to classify the breast is the one containing the brighter block. This decision is based on the idea that malignant findings are usually associated with the increase of local intensity as well as denser areas, which although difficult to classify, are the ones with higher probability of developing breast cancer. Contrarily to traditional approaches, which classify blocks individually and combine their classifications to label the breast, we use the similarity between blocks to decide upon the normality of the breast. In this way, variability caused by external influences tend to impact all blocks and cancel out in the comparison.

The proposed methodology was tested in a real world screening setting comprised by nearly 18 000 images and with a true incidence of suspicious breasts. The obtained results are analogous to the more complex state of the art approaches, suggesting that simple descriptors might be used instead. Since the diagnosis of suspicious cases closer to the decision threshold are dominated by the presence of MCs and AD, we will, in future work, address these signs of breast cancer. Additionally, we aim to use information from the collateral breast to help deciding the dubious cases.

## ACKNOWLEDGMENT

This work was funded by the Portuguese Innovation Agency (ADI) through project QREN reference 33928 “BCS”.

We would like to thank Liga Portuguesa Contra o Cancro for providing us with the data used in this study.

## 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] B. Hela, M. Hela, H. Kamel, B. Sana, and M. Najla, “Breast cancer detection: a review on mammograms analysis techniques,” in *10th International Multi-Conference on Systems, Signals & Devices (SSD)*. 2013, pp. 1–6, IEEE.
- [3] M. P. Sampat, M. K. Markey, and A. C. Bovik, “Computer-aided detection and diagnosis in mammography,” in *Handbook of image and video processing*, vol. 2.1, pp. 1195–1217. Academic Press, 2005.
- [4] Y. Sun, C. F. Babbs, and E. J. Delp, “A two-stage classifier system for normal mammogram identification,” in *Computational Imaging II*, 2004, vol. 112.
- [5] M. Y. Elshinawy, A.-H. A. Badawy, W. W. Abdelmageed, and M. F. Chouikha, “Comparing one-class and two-class SVM classifiers for normal mammogram detection,” in *IEEE 39th Applied Imagery Pattern Recognition Workshop (AIPR)*, 2010, pp. 1–7.
- [6] M. Y. Elshinawy, W. W. Abdelmageed, A.-H. A. Badawy, and M. F. Chouikha, “Pre-CAD system for normal mammogram detection using local binary pattern features,” in *23rd IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, 2010, pp. 352–357.
- [7] M. Y. Elshinawy, A. H. A. Badawy, W. W. Abdelmageed, and M. F. Chouikha, “Normal mammogram detection using density information and texture features,” in *SiiM*, 2011.
- [8] M. Elshinawy, A. Badawy, W. Abdelmageed, and M. Chouikha, “Effect of breast density in selecting features for normal mammogram detection,” in *International Symposium on Biomedical Imaging (ISBI)*, 2011, pp. 141–147.
- [9] K. Ganesan, U. R. Acharya, C. K. Chua, C. M. Lim, and K. T. Abraham, “One-class classification of mammograms using trace transform functionals,” *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 2, pp. 304–311, 2014.
- [10] L. Zhang and W. Zhang, “A comparison of different pattern recognition methods with entropy based feature reduction in early breast cancer classification,” *European Scientific Journal*, vol. 3, pp. 303–312, 2014.
- [11] H. Wang, J.-B. Li, L. Wu, and H. Gao, “Mammography visual enhancement in CAD-based breast cancer diagnosis,” *Clinical Imaging*, vol. 37, no. 2, pp. 273–282, 2013.
- [12] J. S. Cardoso, I. Domingues, I. Amaral, I. Moreira, P. Passarinho, J. S. Comba, R. Correia, and M. J. Cardoso, “Pectoral muscle detection in mammograms based on polar coordinates and the shortest path,” in *32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2010, pp. 4781–4784.
- [13] Meritxell Tortajada, Arnau Oliver, Robert Martí, Sergi Ganau, Lidia Tortajada, Melcior Sentís, Jordi Freixenet, and Reyer Zwiggelaar, “Breast peripheral area correction in digital mammograms,” *Computers in biology and medicine*, vol. 50, pp. 32–40, 2014.
- [14] H. Peng, Fulmi Long, and C. Ding, “Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, Aug 2005.
- [15] Charles Elkan, “The foundations of cost-sensitive learning,” in *Proceedings of the 17th International Joint Conference on Artificial Intelligence - Volume 2*, 2001, IJCAI’01, pp. 973–978.
- [16] Rachel F Brem, Jeffrey W Hoffmeister, Jocelyn A Rapelyea, Gilat Zisman, Kevin Mohtashemi, Guarav Jindal, Martin P DiSimio, and Steven 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.
- [17] 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.