Research Article

Independent Component Analysis to Detect Clustered Microcalcification Breast Cancers

R. Gallardo-Caballero, C. J. García-Orellana, A. García-Manso,
H. M. González-Velasco, and M. Macías-Macías

CAPI Research Group, University of Extremadura, Avenida de la Universidad s/n, 10003 Cáceres, Spain

Correspondence should be addressed to R. Gallardo-Caballero, rgallardo@unex.es

Received 27 October 2011; Accepted 8 December 2011

1. Introduction

Over a million women worldwide are diagnosed with breast cancer every year, accounting for a tenth of all new cancers and 23% of all female cancer cases. Breast cancer incidence rates vary considerably, with the highest rates in North America and the lowest in Africa and Asia. Around 332,670 new cases of breast cancer occurred in the European Union in 2008 [1] and an estimated 182,460 occur in the USA each year [2]. Although breast cancer is a typically female disease, it may also occur in males, but in a much smaller percentage (1% in the USA, 2010 [2]).

The best way to fight against breast cancer is early-stage diagnosis, for which mammography seems to be the most effective test [3] since lesions can be detected before they are even felt by the patient. Obviously, early-stage detection greatly increases the chances of cure. Unfortunately, even expert radiologists can miss a significant proportion of abnormalities [4]. Moreover, a major fraction of mammographic abnormalities are diagnosed as benign after biopsy [5], obliging these patients to undergo an invasive procedure and to have to wait for the corresponding result. While a second reading of the screening mammograms by a human reader can increase cancer detection rates [6, 7], this has not been proven for second readings of commercial computer-aided detection systems [8].

Many methods have been proposed in the last two decades to achieve a robust mammography-based computer-aided diagnosis (CAD) system. Although there are various types of mammographic abnormality, they can primarily be grouped into either masses or microcalcifications [9]. As noted by Horsch et al. in [10], in both variants, the literature tends to consist of unrepeatable studies which provide only a vague description of the mammogram datasets used. Indeed, a premise of the present work is that, currently, the only mammography database that is public and sufficiently large to allow for meaningful and reproducible evaluation of a CAD system is the Digital Database for Screening Mammography (DDSM) [11].

Restricting consideration to papers which provide a reproducible selection of mammograms, one finds the number of papers that address this problem to be really small [10]. One of these works is the study of Kurdziel et al. [12] which
uses the 50 images of BCRP_Calc_1 and reports a false positive per image rate (FPI) of 2.8 for a sensitivity value of 0.79. In a second reproducible study on the same dataset, Yuan and Shi [13] report increasing the sensitivity level to 0.9 but the FPI rate rises to a value of 15.4. Better performance values have been reported by nonreproducible studies such as that of Altrichter and Horváth [14](1.6 FPI at a sensitivity of 0.92) or Kang et al. [15] (0.7 FPI at a sensitivity of 0.85).

In the present study, we use an independent component analysis (ICA) feature extractor to diagnose at pixel level the BCRP_Calc_1 mammograms of the USF’s Digital Database for Screening Mammography, a reproducible and freely available set of mammograms, to develop a diagnosing system with competitive performance values.

The rest of the paper is organized as follows. Section 2 describes the main characteristics of the mammogram database and the feature extraction techniques used. Section 3 presents the main results, and Section 4 the main conclusions.

2. Materials and Methods

2.1. DDSM and Prototype Selection. The Digital Database for Screening Mammography [11] is a resource openly available to the mammographic image analysis research community. It contains a total of 2620 cases. Each case provides four screening views: mediolateral oblique (MLO) and craniocaudal (CC) projections of the left and right breasts. Therefore, the database has a total of 10,480 images. Mammograms are provided in raw mode and were collected using different scanner models depending on the medical institution where the case was diagnosed. This implies that the mammograms may have different grey level values, sampling rates, and even grey level mapping (linear or logarithmic), so that the first step consists in converting each region of the mammogram to optical density using the calibration equations provided [11].

Cases are categorized into four major groups: normal, cancer, benign, and benign without callback. In addition to the regular DDSM volumes, the DDSM website provides the so-called “DoD BCRP Mammography Datasets at USF” comprising four sets [16]. They were created for preliminary evaluation of the performance of CAD algorithms on a common dataset. Two of the four datasets focus on spiculated masses (BCRP_MASS_0 and BCRP_MASS_1) and the other two on clustered microcalcifications (BCRP_Calc_0 and BCRP_Calc_1). Each case included in the BCRP datasets contains at least one malignant lesion.

All cases in the DDSM database were annotated by experienced radiologists providing various BI-RADS parameters (density, assessment, and subtlety), a BI-RADS abnormality description, and proven pathology. For each abnormality identified, the radiologist draws free-form digital curves defining ground truth regions. Additionally, each case includes other information such as patient age, date of study, and digitization or digitizer’s brand.

The present work was carried out using the 100 cases (200 mammograms using the two MLO views of each case) selected in the BCRP_Calc subset. This subset contains uniquely mammograms sampled at 43.5 microns per pixel with two different linear scanners and a spectral resolution of 12 bits. The BCRP_Calc_0 (Calc_0) subset is used to adjust the model, and BCRP_Calc_1 (Calc_1) is used to provide system performance.

Our prototype database was constructed by labeling at pixel level not only microcalcifications but also other typical mammogram structures (eggshell and lucent-centred calcifications, calcified vessels, artifacts, etc.). This step was carried out with the assistance of an experienced radiologist who labeled pixels for the proposed types of structures on all mammograms of the set. We built an initial database of 4638 malignant microcalcification prototypes, 6794 isolated microcalcifications, and more than 100,000 prototypes of completely normal tissue or of some other benign abnormalities.

The first set of prototypes tends to provide erroneous classifications in certain healthy mammogram areas. We thus decided to add more prototypes on those areas with the aim of improving the performance of the pixel classification stage. New prototypes were included as normal pixels using a previously diagnosed mammogram. These new prototypes were included as single-pixel samples in contrast to the previous normal prototypes which were taken by covering small areas of normal tissue. The new normal prototypes seem to provide a broader description of the image space which should be considered as unsuspicious. Hence we decided to discard all previous normal prototypes and preserve only the new 53,334 single-pixel normal prototypes in the following studies.

2.2. Independent Components Analysis. Independent component analysis (ICA) [17] is a statistical and computational technique designed to reveal hidden factors that underlie sets of random variables, measurements, or signals. It defines a generative model for the observed multivariate data, typically given as a large sample database. In this model, it is assumed that the data are linear combinations of unknown latent variables, and the system whereby they combine is also unknown. It is also assumed that non-Gaussian latent variables are mutually independent and thus considered independent components of the observed data. These independent components, also called sources or factors, can be found by ICA. ICA-analyzed data may come from very different sources, including digital images, documents, databases, economic indicators, and psychometric measurements. In many cases, the measurements are given as sets of parallel signals or time series in problems known as blind source separation.

Thus, suppose that one has \( n \) signals. The objective is to express the signals registered by the sensors \( x_j \) as a linear combination of \( n \) sources \( s_j \), a priori unknown

\[
    x_j = \sum_{j=1}^{n} a_{i,j} s_j. \quad (1)
\]

The goal of ICA is to estimate the mixing matrix \( A = (a_{i,j}) \), in addition to the sources \( s_j \). This technique can be used for feature extraction since the components of \( X \) can
be regarded as features representing the objects as stated in Chapter 21 of [17].

2.3. ICA Feature Extraction and Classification. Many basic models in image processing express an image as a linear superposition of some features or basis functions, where the coefficients are different for each image. Some linear transformations widely used in image processing are the Fourier, Haar, and Gabor transforms. Following this model, one can represent a pixel and its neighbourhood in a mammographic image as a series expansion in these functions. The coefficients of this series expansion can then be used as features characterizing the pixel to compose a feature vector which may be used for its diagnosis. One can visualize this concept in Figure 1.

To estimate the ICA basis from images, one needs to collect samples (patches) from the images to model. The collected patches are used to build a data matrix \( \mathbf{X} \) which is the input to the FastICA [18] algorithm. In this algorithm, the data are first centred by subtracting the mean of each column of the data matrix \( \mathbf{X} \). The data matrix is then whitened by projecting the data onto its principal component directions using a prewhitening matrix (\( \mathbf{K} \)). The ICA algorithm then estimates an unmixing matrix \( \mathbf{W} \) so that \( \mathbf{X} \mathbf{W} = \mathbf{S} \), with \( \mathbf{S} \) being the estimated source matrix. Using the previous notation, with \( \mathbf{R} \) being the image associated with a previously centred window, one can obtain the feature vector \( \mathbf{F} \) that characterizes the window as \( \mathbf{F} = \mathbf{K} \mathbf{W} \mathbf{R} \).

The number of features to obtain and the size of the sampling window are selected a priori. Therefore the optimum ICA configuration for the problem must be determined using a supervised procedure. Each feature extractor requires a set of ICA matrices to be generated. We created them using samples centred on the previously defined prototypes. Thus, working at pixel level, we compute a wide range of ICA features (from 2 to 80) for different window sizes (from 3 to 81 pixels wide) over a set of more than 65,000 testing samples. The ICA matrix generation was carried out using the FastICA implementation available for the R statistical environment [19].

Once the ICA basis has been obtained, one can obtain a feature vector that characterizes the pixel under study and its neighbourhood by selecting the window size and the number of features to extract. We used an artificial neural network (ANN) to classify the feature vectors obtained, in particular, a classical feed-forward multilayer perceptron with a single hidden layer trained with a variation of the back-propagation algorithm named Resilient Back-Propagation (RPROP) [20] to obtain faster convergence. The number of neurons in the hidden layer is allowed to grow from 50 to 200 in order to improve the success rate of the classifier.

In this step, the objective is to optimize microcalcification detection in mammograms. Hence, for each combination of ICA window size and feature number, one needs to compute the success rate in microcalcification detection of each classifier. Due to the heavy demand on computational resources required to select the best ICA feature extractor and classifier, all subsequent batch calculations were carried out on a 48 quad-core distributed PC cluster.

2.4. Generation of Regions of Interest. To determine the location of a cancer defined by a microcalcification cluster in a mammogram, one needs to generate regions of interest (ROIs) in those areas where the number of microcalcifications is greater than 3 in an one-square-centimetre neighbourhood [21]. Therefore, once the microcalcification detection phase has been optimized, each mammogram is processed to obtain an image with the location of microcalcification candidates. This image is filtered to discard spurious microcalcifications (isolated pixels or artifacts). The resulting image is processed by an ROI builder system which groups neighbouring microcalcifications into disjoint zones of interest.

The first step of the ROI builder consists in an object detector that isolates microcalcifications and computes their sizes. The second step consists of grouping the objects by means of a density map. For each pixel of the mammogram, the number of microcalcifications in the previously defined neighbourhood area is computed. After that, the density map is filtered, retaining only those zones which have a microcalcification density value equal to or greater than 3. Figure 2(b) provides a sample of the regions of interest obtained with this procedure on a mammogram.

2.5. Computing Performance. The performance of mammographic CAD systems is evaluated by means of such parameters as sensitivity, specificity, and false positives per image (FPi). Sensitivity (2) or the true positive rate measures the proportion of actual positives which are correctly identified as such. And FPi measures the mean value of false positives (FP) that are generated for a test set (see (3))

\[
\text{sensitivity} = \frac{\text{TN}#}{\text{TN}# + \text{FN}#}, \quad (2) \\
\text{FPi} = \frac{\text{FP}#}{\text{image}#}. \quad (3)
\]
Figure 2: (a) Original DDSM mammogram with the expert radiologist’s mark surrounding the microcalcification cluster. (b) Mammogram showing the regions of interest generated by our system. The upper ROI accounts for the true abnormality.

The criteria chosen in our implementation to define the possible outcomes of our system are as follows. Each generated possible ROI in a mammogram is tagged as true positive (TP) if it overlaps with the annotated region provided in the DDSM database (if present). Analogously, the ROI is tagged as false positive (FP) if there exists no overlap with an actual malignancy of the mammogram or the mammogram is healthy. On the contrary, if DDSM contains malignant ROIs for a mammogram and our system fails to find them, we compute all those ROIs as false negatives (FN). And finally, a successfully diagnosed healthy mammogram is considered a true negative (TN).

A graphical method known as free-response operating characteristic (FROC) is also used to evaluate the performance of CAD system. A computer-based detection algorithm attempting to locate a target in a digital image may provide more than one suspicious zone in an image, in most cases only one being the correct target while the others are false positives. FROC [22] is a variant of receiver operating characteristic analysis [23] needed to be able to take into account the FP data. For an FROC curve, the hit rate is expressed as a proportion, but unlike ROC curves the FP rate is expressed as the mean number per image. FROC curves provide a clearer picture of how a system is operating than a conventional ROC curve. However, no very useful measure of sensitivity can be derived from them.

3. Results and Discussion

Any combination of ICA window size and number of features to extract generates an input feature extraction configuration to be evaluated. To train the neural classifier, we defined a dataset of pixel locations for which ICA-extracted feature vectors should be obtained for each input configuration. The dataset was split into three disjoint subsets: learning, validation, and test according to pixel origin (CALC_0 or CALC_1). The validation subset is to avoid overtraining and improves the network generalization capabilities [24].

The Stuttgart neural network simulator (SNNS) [25] environment was used to generate and train the network. The high number of input configurations to optimize led us to use SNNS’s kernel function facilities to implement the network training procedures in standalone executables that could be run in a distributed Beowulf cluster [26].

To achieve an optimal network configuration, we made a sweep of the number of neurons in the hidden layer. We built and trained several network configurations with a number of hidden neurons ranging from 50 to 200 in steps of 50. Furthermore, each configuration was repeated four times, performing a random initialization of the neuron weights in each repetition, and retaining the classifier that provides the greatest success rate. Finally, the test subset, which contains input vectors extracted from CALC_1 prototypes, was used to provide performance results of the microcalcification detection subsystem.

The main objective of the study was to obtain a system that relies mainly on image processing features, but we also studied input configurations that included the patient’s age as an additional feature. The age of the patient was scaled linearly from the range [27, 91] to the range [-1, +1] for inclusion in the feature vector.

3.1. Detection Performance on Complete Mammograms. We trained 7011 neural classifiers with different input configurations and prepared 85 experiments to evaluate the performance on the CALC_1 subset. Each experiment provides different performance values depending on the network-analyzing function selected, so that the 85 experiments required repeating the diagnosis procedure 1705 times. These
Table 1: Performance results on the BCRP_CALC1 subset of the five system configurations that generate the lowest false positives per image and provide a sensitivity higher than 80%.

| Configuration | Analyzer | FPI | Sensitivity (%) |
|---------------|----------|-----|-----------------|
| ICA-D013C040  | WTA1     | 2.55| 81.8            |
| ICA-D013C040  | WTA0     | 3.08| 83.6            |
| ICA-D015C020+AGE | 402040 | 3.25| 81.4            |
| ICA-D015C024  | WTA1     | 3.40| 84.2            |
| ICA-D017C020+AGE | WTA3  | 3.44| 81.8            |

Table 2: Performance results on the BCRP_CALC1 subset of the five system configurations that generate the lowest false positives per image and provide a sensitivity higher than 90%.

| Configuration | Analyzer | FPI | Sensitivity (%) |
|---------------|----------|-----|-----------------|
| ICA-D017C020+AGE | WTA1  | 4.45| 91.8            |
| ICA-D017C020+AGE | WTA0  | 4.85| 91.3            |
| ICA-D019C020+AGE | WTA1  | 4.97| 91.5            |
| ICA-D015C010+AGE | WTA3  | 5.40| 91.9            |
| ICA-D015C010+AGE | WTA2  | 5.58| 92.0            |

The use of the patient’s age as a parameter seems to be significant for high sensitivity requirements, but of negligible effect at lower sensitivity levels. In either case, its use would appear to be acceptable because it is easily obtainable, whereas other proposals sometimes use expert-provided DDSM parameters such as breast density or even the assessment of the abnormality to provide a diagnosis.

4. Conclusions

Although the computer-aided detection of breast cancer is a fairly well-established field of research, the complexity of mammographic images has led to there being really few reproducible studies of diagnosis based on microcalcification cluster detection. Independent Component Analysis seems to constitute a very suitable approach to the efficient detection of clustered microcalcifications, improving published reproducible performance results by a factor of from 8.9% (at a sensitivity higher than 79%) to more than 71% (for sensitivities higher than 90%).

The relative small area required for the ICA windows that provide the best microcalcification cluster detection seems to suggest that microcalcification cluster detection is a highly local task. The patient’s age appears in the third configuration, but its presence was not required to reach the highest success rates, thus allowing a system to be used that is completely based on image-extracted parameters.

The analyzer column in Tables 1 and 2 specifies the analyzing function used to classify the output of the neural network. Here, 402040 stands for the “402040” network activation rule [25] with parameters $l = 0.45$ and $h = 0.55$, and WTA stands for winner takes all, meaning that the classification depends on the neural network unit with the highest output and the value of the parameter $h$. The value of the parameter $h$ is zero for WTA0, 0.5 for WTA1, 0.55 for WTA2, and 0.6 for WTA3.

Increasing the sensitivity value requirement to 90%, one observes a rise in the mean number of false positives per image generated. The lowest FPI value obtained by a system was 4.45 at a sensitivity of 91.8% (see Table 2), a significantly lower value than the 15.4 reported by Yuan and Shi [13]. It is noteworthy that the age of the patient seems to play a major role when high sensitivity values are required. Nevertheless, again both the size of the ICA sampling window and the number of ICA features to use remain at low values, reducing the computational demand.

Figure 3 shows the FROC curves for the systems that provide the lowest FPI values at 80% and 90% sensitivity levels and working on the CALC1 subset with the 21 predefined analyzing functions. The blue trace shows the FROC performance of the system with the lowest FPI at a sensitivity level higher than 80%, and the green trace shows the same but for a sensitivity level higher than 90%.

Acknowledgments

This work was partially supported by the “Junta de Extremadura” and FEDER through projects PRJ08A092, PDT09A036, PDT09A005, and GR10018.
References

[1] J. Ferlay, D. M. Parkin, and E. Steliarova-Foucher, “Estimates of cancer incidence and mortality in Europe in 2008,” European Journal of Cancer, vol. 46, no. 4, pp. 765–781, 2010.

[2] American Cancer Society, “Cancer facts & figures,” Tech. Rep., American Cancer Society, Atlanta, Ga, USA, 2010.

[3] O. Peart, Mammography and Breast Imaging. Just the Facts, McGraw-Hill, 1st edition, 2005.

[4] R. L. Birdwell, D. M. Ikeda, K. F. O’Shaughnessy, and E. A. Sickles, “Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection,” Radiology, vol. 219, no. 1, pp. 192–202, 2001.

[5] F. M. Hall, J. M. Storella, D. Z. Silverstone, and G. Wyshak, “Nonpalpable breast lesions: recommendations for biopsy based on suspicion of carcinoma at mammography,” Radiology, vol. 167, no. 2, pp. 353–358, 1988.

[6] E. L. Thurfler, K. A. Lernevall, and A. A. S. Taube, “Benefit of independent double reading in a population-based mammography screening program,” Radiology, vol. 191, no. 1, pp. 241–244, 1994.

[7] R. M. L. Warren and S. W. Duffy, “Comparison of single reading with double reading of mammograms, and change in effectiveness with experience,” British Journal of Radiology, vol. 68, no. 813, pp. 958–962, 1995.

[8] P. Taylor and H. W. W. Potts, “Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate,” European Journal of Cancer, vol. 44, no. 6, pp. 798–807, 2008.

[9] D. Kopans, Breast Imaging, LWW Medical Book Collection; Lippincott Williams & Wilkins, 2007.

[10] A. Horsch, A. Hapfelmeier, and M. Elter, “Needs assessment for next generation computer-aided mammography reference image databases and evaluation studies,” International Journal of Computer Assisted Radiology and Surgery, vol. 6, pp. 749–767, 2011.

[11] M. Heath, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer, “The digital database for screening mammography,” in Proceedings of the 5th International Workshop on Digital Mammography, M. J. Yaffe, Ed., pp. 212–218, Medical Physics, 2001.

[12] M. Kurdziel, K. Boryczko, and D. A. Yuen, “Detecting clusters of microcalcifications in high-resolution mammograms using support vector machines,” Bio-Algorithms and Med-Systems, vol. 3, pp. 11–21, 2007.

[13] X. Yuan and P. Shi, “Microcalcification detection based on localized texture comparison,” in Proceedings of the International Conference on Image Processing (ICIP ’04), vol. 5, pp. 2953–2956, October 2004.

[14] M. Altrichter and G. Horváth, “The refinement of microcalcification cluster assessment by joint analysis of MLO and CC views,” in Proceedings of the International Workshop on Digital Mammography (IWDM ’06), S. M. Astley, M. Brady, C. Rose, and R. Zwiggelaar, Eds., vol. 4046 of Lecture Notes in Computer Science, pp. 509–516, Springer, 2005.

[15] H. K. Kang, S. M. Kim, N. N. Thanh, Y. M. Ro, and W. Kim, “Adaptive microcalcification detection in computer aided diagnosis,” in Proceedings of the International Conference on Computational Science, M. Bubak, G. D. van Albada, P. M. A. Sloot, and J. Dongarra, Eds., vol. 3039 of Lecture Notes in Computer Science, pp. 1110–1117, Springer, 2004.

[16] “DoD BCRP Mammography Datasets at USF,” 2000, http://marathon.csee.usf.edu/Mammography/DDSM/BCRP/bcrp.html.

[17] A. Hyvarinen and E. Oja, “Independent component analysis: algorithms and applications,” Neural Networks, vol. 13, no. 4-5, pp. 411–430, 2000.

[18] A. Hyvarinen, “Fast and robust fixed-point algorithms for independent component analysis,” IEEE Transactions on Neural Networks, vol. 10, no. 3, pp. 626–634, 1999.

[19] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2008.

[20] M. Riedmiller and H. Braun, “Direct adaptive method for faster backpropagation learning: the RPROP algorithm,” in Proceedings of the IEEE International Conference on Neural Networks, pp. 586–591, April 1993.

[21] E. A. Sickles, “Mammographic features of 300 consecutive nonpalpable breast cancers,” American Journal of Roentgenology, vol. 146, no. 4, pp. 661–663, 1986.

[22] J. P. Egan, G. Z. Greenberg, and A. I. Schulman, “Operating characteristics, signal detectability, and the method of free response,” Journal of the Acoustical Society of America, vol. 33, pp. 993–1007, 1961.

[23] J. P. Egan, Signal Decision Theory and Roc Analysis, Academic Press, 1975.

[24] C. M. Bishop, Pattern Recognition and Machine Learning, Springer, Berlin, Germany, 2006.

[25] A. Zell, N. Mache, R. Hübner et al., “SNNS (stuttgart neural network simulator),” in Neural Network Simulation Environments, J. Skrzypek, Ed., vol. 254 of The Springer International Series in Engineering and Computer Science, Kluwer Academic Publishers, Norwell, Mass, USA, 1994.

[26] T. Sterling, D. J. Becker, D. Savarese, J. E. Dorband, U. A. Ranawake, and C. V. Packer, “A parallel workstation for scientific computation,” in Proceedings of the 24th International Conference on Parallel Processing, pp. 11–14, CRC Press, 1995.