Segmentation and Classification of Breast Cancer Using Independent Component Analysis, Texture Features and Neural Networks

Lucio Flávio A. Campos[1,2], Emanuel C. M. Lemos[1], Luis C. O. Silva[2], Daniel D. Costa[2], Allan Kardec Barros[2]

[1]Instituto de Engenharia e Ciências da Computação – Universidade Estadual do Maranhão (UEMA)
Campus Paulo VI – São Luís – MA – Brazil

[2]Laboratório de Processamento da Informação Biológica– Universidade Federal do Maranhão (UFMA)

lucioflavio@engcomp.uema.br,emanuel.cml@gmail.com,
luisoliveirasilva@hotmail.com, danieldc@gmail.com, akbarros@ieee.org

Abstract. *We propose a method for segmentation and classification of breast cancer in digital mammograms using Independent Component Analysis (ICA), Texture Features and Multilayer Perceptron (MLP) Neural Networks. The method was tested for a mammogram set from MIAS database, resulting in 90.15% success rate, with 92% of specificity and 88.3% of sensitivity.*

1. Introduction

Breast cancer is the major cause of death by cancer in the female population [INCa, 2010]. It is know that the best prevention method is early diagnosis, which lessens the mortality and enhances the treatment. Therefore, a great effort has been made to improve the early diagnosis techniques. Among them, the most used is the mammogram, for it is low cost and easy access. However, mammogram has a high error value for medical diagnosis, ranging from 10 to 25%, resulting in a great number of false-positives diagnostics, which causes unneeded biopsies, or false-negatives, which delays the cancer diagnosis. The analysis of digital mammography is a complex cognitive task that includes various aspects of medical expertise and conclusive clinical findings [Newsread, Baute, 1992], [Meyer et al, 1989]. The visual task of clinical evaluation and diagnosis, based on digital mammography, consists of a number of different factors in multiple scales and levels of decomposition [Sickles, 1989], [Bocchi et al, 1997]. The fine-scale organization of the informational content on the mammogram is a key factor in the detection of breast cancer, as it represents the nature, structure and the quality of biological tissues, as they are projected on the mammogram [Homer, 1987], [Homer,1985]. Similar textural features are also present in rare clinical cases, where direct inference on probably benignancy or malignancy is much more complex.

These fine-scale structural details are realized as visual patterns in the image and they are often referred to as ‘‘texture’’ of the corresponding region. When the digital mammogram is obtained with adequate quality and resolution, these textural patterns can be identified, analyzed and classified with an aided of image processing and

computational vision algorithms, combined with artificial intelligence for features extraction.Those algorithms are able to decrease the error and make the mammograms more reliable [Bick, 1996].

The CAD (computer-aided diagnosis) systems can aid radiologists by providing a second opinion and may be used in the first stage of examination. For this to occur, it is important to develop many techniques to detect and recognize suspicious lesions and also to analyze and discriminate them.

Segmentation is an important task to identify regions suspicious of cancer in digital mammograms. After identified, each region must be classified as benign, malignant cancer or normal tissue.
The principal objective of image classification is to assign all pixels in the image to particular classes or themes, based on some features, for example. This type of classification is called pattern recognition.

Some methods of segmentation and classification of breast cancer in digital mammography have been reported.

[Campos et al, 2007] used independent component analysis (ICA) and neural network multilayer perceptron to classify mamogramas in 3 class: normal, benign and malignant, with 98,7% of successful. [Braz et al, 2007] classified the regions of interest of screening mammogram using spatial statistics, with performance of 98.24% to discriminate Mass from Non-Mass elements. [Dominguez, Nandi, 2007] use enhanced multilevel thresholding segmentation and region selection based on rank mammogram segmentation. According the authors, this method had a better performance, with 80% of sensitivity. [Campos et al, 2008] use independent component analysis, feature extraction and K - means cluster to segment digital mammography. The proposed method obtained 86.6% of success.

In this paper we propose to use the independent component analysis (ICA) to generate a data dependent filter bank for mammograms segmentation, and after generate a basis functions through the coefficients (features) extracted using ICA. Then, those features are used as input parameters to a Neural Network do the classification.

We divide this work as follows. Into section 2 we show the techniques for segmentation and classification the mammograms. In section 3 we present the results and discuss about the application of the techniques under study. Finally, section 4 presents some concluding remarks.

2. Methods

The block diagram of the proposed method is shown in Figure 1.

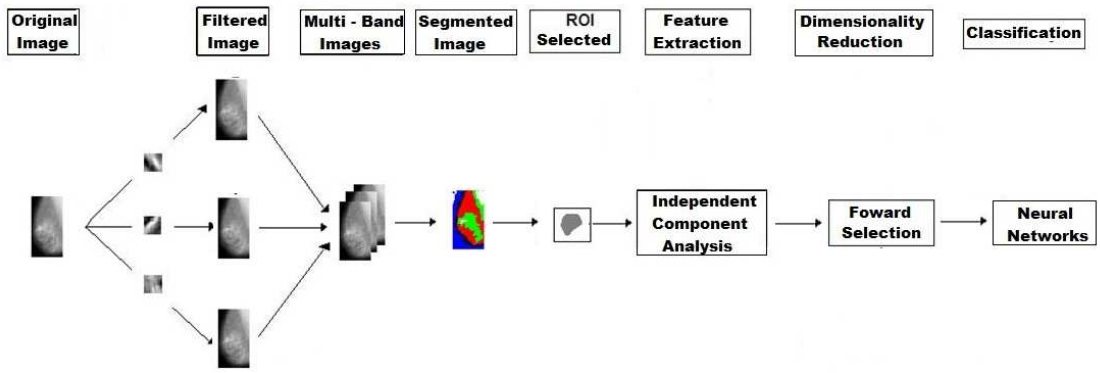

Figure 1: Block Diagram

The method was divided in two steps. Step one basically consists on filtering an original image, using ICA filter bank, feature extraction using non–linear operators, and segmentation using K-means clustering. The step two consists of the extraction of features in the ROI using ICA, reduction of insignificants features using the *forward-selection* technica and the classification of the tissue through neural networks.

2.1. Independent Component Analysis

Let us assume that an image x may be expressed as a linear combination of basis images $a_1, a_2, \cdots, a_n$ pondered by mutually statistically independent coefficients $s_1, s_2, \cdots, s_n$ [Hyvärinen, 1997], such that

$$x = a_1 s_1 + a_2 s_2 + \cdots + a_n s_n \tag{1}$$

In equation 1, only the variable x is known, and from that we estimate the coefficients a_i and the independent components s_i.

In the sequel, we represent an image as a column vector

$$x = \left[x_1, x_2, \cdots x_m \right]^T \tag{2}$$

by re-shaping the image matrix row-by-row into a single column. In this way we avoid using two-dimensional matrices in the ICA filtering, and an image may thus be represented using the standard ICA methodology. Mathematically, we express the model as,

$$x = \sum_{i=1}^{N} a_i s_i = As \tag{3}$$

Where x is the image data, the basis functions a_i, $i = 1, 2, \cdots, N$, are the columns of the $(M \times N)$ matrix A, and $s = \left[s_1, s_2, \cdots s_N \right]^T$ the independent components.

The image model described here is illustrated in Fig. 2.

Fig. 2: The ICA model.

2.2. FastICA Algorithm

The data matrix $\boldsymbol{X}$ is considered to be a linear combination of non-Gaussian (independent) components i.e., $\boldsymbol{X} = \boldsymbol{A}.\ \boldsymbol{S}$ where columns of $\boldsymbol{S}$ contain the independent components and $\boldsymbol{A}$ is a linear mixing matrix. In short ICA attempts to "un-mix" the data by estimating an un-mixing matrix $\boldsymbol{W}$, where $\boldsymbol{X}.\boldsymbol{W} = \boldsymbol{S}$.

Under this generative model the ICA, the measured in $\boldsymbol{X}$ will tend to be more Gaussian than the source components $\boldsymbol{S}$. Thus, in order to extract the independent components we search for an un-mixing matrix $\boldsymbol{W}$ that maximizes the non-gaussianity of the sources. In FastICA, non-gaussianity is measured using approximations to negentropy (J) which are more robust than kurtosis based measures and fast to compute [Marchini et al, 2004]. The approximation takes the form

$$J_{G(y)} = \left| E_y\{G(y)\} - E_v\{G(v)\} \right|^p \tag{4}$$

Where v is a standartized Gaussian randon variable, y is assumed to be normalized to unit variance, and the exponent p=1,2 typically (The notation J_g do not be confused with notation on for negentropy , J).

2.3. ICA Filter Bank

For generate the ICA filter bank, we use the fastICA algorithm applied to Regions on Interests (ROIs), containing textures of normal and abnormal tissues. We reduced the dimension of the ROIs at the Principal Components Analysis (PCA) step.

The generation of training data includes the following computations:

1. Select the ROI of an ensemble;
2. Select how many patches to take per image;
3. Select the given number of (m×m) sized patches from random locations;
4. Subtract from each patch its mean value;
5. Represent samples as columns, and store in a matrix consisting of training vectors.

The computations described above result in an ensemble of training vectors which are presented to the FastICA algorithm [Marchini, 2004], after additional preprocessing by PCA. After convergence of the algorithm, matrix W has been learned from the available training data. Matrix A is obtained by finding the inverse (or pseudo inverse) of W. As a last step, the columns A, are reshaped into size $(m \times m)$ to constitute the basis functions. These are now the impulse responses of our filter bank.

Each filter responds to specific frequency and orientation characteristics and specific texture properties. Thus, each filtered image will have high frequency in regions corresponding to textures which are tuned to the filter, and low energy corresponding to textures which are not tuned to that filter. If the banks consist of N filters, the result is N filtered images of the same size as the input image.

2.4. Features Extraction

Only filter outputs by default are not appropriate for identifying key texture features. The objective of the feature extraction using non-linear operators is to estimate the energy in a local region of the filter outputs [Jain, Farrokhnia, 2001], [Unser, Eden, 1990].

A number of feature extraction methods were proposed to extract useful information from the filter outputs, however, the most commonly used non-linearities are:

- Magnitude function; $F(y)=|Y|$
- Non-linear sigmoidal function; $F(y)=\tanh(Y)$
- Square function. $F(Y)=(Y)^2$

Where Y is a filtered image by ICA.

The feature image obtained F(Y) has the variance disparities converting of into mean values differences, an essential step if standard clustering are to be employed.

2.5. Clustering in the Feature Space

At the end of the feature extraction step we are left with a set of feature images extracted from the filter outputs. Pixels that belong to the same texture region have the same texture characteristics, and should be close to each other in the feature space. The final step in unsupervised texture segmentation is to cluster the pixels into a number of clusters representing the original texture regions. Labeling each cluster yields the segmented image.

Different approaches were taken for the clustering process [Jain, Farrokhnia, 2001]. In this paper we use the basic K-means clustering algorithm for simplicity.

K-means [Moore, 2001] is one of the most used unsupervised learning algorithms that solve the clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of **k** clusters fixed a priori.

The main idea is to define **k** centroids, one for each cluster. These centroids shoud be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other.

The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point we need to re-calculate **k** new centroids as barycenters of the clusters resulting from the previous step.

After we have these **k** new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the **k** centroids change their location step by step until no more changes are done. In other words centroids do not move any more.

Finally, this algorithm aims at minimizing an *objective function*, in this case a squared error function. The objective function is:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \tag{5}$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, is an indicator of the distance of the *n* data points from their respective cluster

centers. When clustering is done, each pixel is labeled with its respective cluster, finally producing the segmented image.

2.6. Multilayer Perceptron Neural Networks

The Multilayer Perceptron (MLP), a feed-forward back-propagation network, is the most frequently used neural network technique in pattern recognition [Duda, Hart, 1973], [Bishop, 1999]. Briefly, MLPs are supervised learning classifiers that consist of an input layer, an output layer, and one or more hidden layers that extract useful information during learning and assign modifiable weighting coefficients to components of the input layers. In the first (forward) pass, weights assigned to the input units and the nodes in the hidden layers and between the nodes in the hidden layer and the output, determine the output. The output is compared with the target output. An error signal is then back propagated and the connection weights are adjusted correspondingly. During training, MLPs construct a multidimensional space, defined by the activation of the hidden nodes, so that the three classes (malignant, benign and normal tissue) are as separable as possible. The separating surface adapts to the data.

2.7. Selection of Most Significant Features

Our main objective to identify the effectiveness of a feature or a combination of features when applied to a neural network. Thus, the choice of features to be extracted is important.

Forward selection is a method to find the “best” combination of features (variables) by starting with a single feature, and increasing the number of used features, step by step [Funukaga, 1990]. In this approach, one adds features to the model one at a time. At each step, each feature that is not already in the model is tested for inclusion in the model. The most significant of these feature is added to the model, so long as P-value is below some pre-selected level.

2.8. Evaluation of Classification Methods

Sensitivity and specificity are the most widely used statistics to describe a diagnostic test. Sensitivity is the proportion of true positives that are correctly identified by the test and is defined by $S = TP/(TP+FN)$. Specificity is the proportion of true negatives that are correctly identified by the test and is defined by $TN/(TN+FP)$. Where ***FN*** is false-negative, ***FP*** is false-positive, ***TN*** is true negative and ***TP*** is true positive diagnosis.

3. Experimental Results and Discussion

Here are describe the results obtained using the method proposed in the previous section.

3.1. Mammogram Database

The database used in this work is the *Mammographic Institute Society Analysis* (MIAS) [Sucking et al, 1994]. The mammograms have a size of 1024x1024 pixels, and resolution of 200 micron. This database is composed of 332 mammograms of right and left breast, from 161 patients, where 53 were diagnosed as being malignant, 69 benign and 206 normal. The abnormalities are classified by the kind of found abnormality

(calcification, circumscribed masses, architectural distortions asymmetries, and other ill-defined masses).

For the segmentation**,** to generate the ICA filter bank, we selected 150 ROI from this database: 25 benign, 25 malignant and 100 normal tissues of the mammograms. The ROIs was found through of ***xy*** images coordinates of centre of abnormality contained in file list of MIAS database. To the normal tissue, only pectoral muscle wasn't considered as a possible ROI, but tissue and fat tissue were.

3.2. ICA Application, in Segmentation of mammograms

For generate the matrix *x* of equation 3, were randomly selected patches of size 12 x 12 of the 95 ROIs, and each patches generate de vector of size 1 x 144. It was generate 30000 training vectors that were used as input to fastICA algorithm.

In order to reduce the dimensionality of the filters, we made a preprocessing using principal component analysis. ICA was then performed successively on the 8 most significant filters.

Each row of matrix A, of size 1 x 144, was reshaped in windows, of size 12 x 12. Each window was used as a filter, localized in spatial frequency, and orientation.

3.3. Extracting Features

For feature extraction of filtered images by ICA, we used the magnitude function, combined with a square function:

$$F(Y) = (|\, Y \,|)^2 \tag{6}$$

In order to carry out the tests, we selected 60 mammograms: 35 malignant and 25 benign diagnoses. These mammograms that do not used to generate the ICA filter bank.

If the 60 mammograms, the algorithm detected 186 regions suspicious. Based in MIAS Database, can affirm that 186 regions founded, 60 regions have cancer, and 126 haven't. Figure 3 shows a comparison between the original mammograms and the segmented mammograms using K-means clustering, with 8 filters.

Fig.3: Comparison between: Original mammograms and mammogram segmented. The circumscribed regions indicate suspicious regions.

3.3. Classification of Regions Suspicious

The 186 suspicious regions, the ROI were manually selected, containing a lesion, in the case of abnormal mammograms. The ROI's was found using the informations through of *xy* images coordinates of centre of abnormality, contained in file list of MIAS

database, and the segmentation response. In the case of regions where not founded cancer, was extracted regions based only in segmentation response.

If the tissues had different sizes, it was rescaled each ROI. Therefore, they were resized to 24x24 pixels. Figure 4 exemplifies the ROI selection of a mammogram.

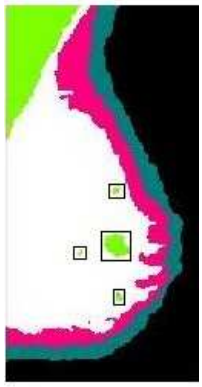

Fig. 4: Chosen tissue of mammogram

3.4. ICA Application for Classification of Mammograms

X of Equation 2 was represented using the chosen ROIs.

The images with ROI were rescaled and transformed into a one-dimensional vector

$$P = P_x \, x \, \boldsymbol{P_y} \quad \textbf{(6)}$$

Where ***Px*** is a rows and **Py** is a columns of ***P*** and ***P*** has dimension 1x 576.

Each sample represents one row of the mixture matrix. The matrix **X** is represented by the samples into the dimension of **P**, that is, 1x576. Thus, each row of the matrix **A** corresponds to a ROI, and each column corresponds to an attributed weight to a base image, i.e., an input parameter to the neural network [Christoyanni, 2002].

Using the FastICA algorithm and the matrix ***X***, we obtain the basis function matrix **A,** which contains the features of each sample.

3.5. Multilayer Perceptron Neural Networks

Using the *forward-selection* algorithm, 21 basis functions were selected as being the most significant features. The chosen features (a_i) are the input to the MLP Network. The best result was obtained using one input layer (21 neurons), one hidden layer (2 neurons) and one output layer (2 Neurons).

In order to realize the tests, we divided a sample into 186 ROI: 93 for training and 93 for tests.

4. Results

Table I shows the performance of application of the ICA technique and MLP network to each regions suspicious.

Table I: Classification of cancer and not cancer in regions suspicious. The lines and columns show the database and algorithm classification, respectively.

	Cancer	Not Cancer
Cancer	**53**	**10**
Not Cancer	**7**	**116**
Success (%)	**88.3**	**92**

Based on Table I, the method obtained a success rate of 90.15%, for discriminating normal (not cancer) and abnormal (cancer) tissues. The specificity was 92 %, and sensitivity was 88.3%. The method obtained 53 true positives, 116 true negatives. 10 false positives and 7 false negatives.

5. Conclusions

The presented results demonstrated that Independent Component Analysis, texture features and MLP Neural Networks were good techniques to Segment and classify Digital Mammograms, discriminating cancer and not cancer tissue.

Furthermore, the best performance was found using 8 ICA filters segmented those mammograms, and after MLP Neural Networks to classification, with a success rate of 90.15%. It can decrease the number of unneeded biopsies and late cancer diagnosis.

Nevertheless, there is the need to perform tests with a larger database and more complex cases in order to obtain a more precise behavior pattern.

References

Braz , G. Jr., E. C. Silva, A. C. Paiva and A. C. Silva, “Breast Tissues Classification Based on the Application of Geostatistical Features and Wavelet Transform”, In: International Special Topic Conference on Information Technology Applications in Biomedicine, ITAB 2007, 6th, 227-230, 2007.

Bick U., M. Giger, R. Schmidt, R. Nishikawa, D. Wolverton and K. Doi. “Computer-aided breast cancer detection in screening mammography”, In: Digital Mammogr'9Chicago, Il (1996), pp. 97–103.

Bishop, C.M.*:* “Neural Networks for Pattern Recognition”. Oxford University Press, New York (1999)

Bocchi L, G. Coppini, R. De Dominicis and G. Valli, *Tissue characterization from X-ray images*, *Med Eng Phys* 1997;19(4); pp. 336–342.

Campos, L. F. A. ; Barros, A. K. ; Silva, A. C. “Independent Component Analysis and Neural Networks Applied for Classification of Malignant, Benign and Normal Tissue in Digital Mammography”, In: Special Issue - Methods of Information in Medicine, v. 46, p. 212-215, 2007.

Campos, L. F. A. ;Costa, D. D.; Barros, A. K. “Segmentation on Breast Cancer Using Texture Features and Independent Component Analysis”, In: Bioinspired Cognitive Systems, BICS 2008.

Christoyianni I., Koutras A., Kokkinakis G., “Computer aided diagnosis of breast cancer in digitized mammograms”, In: Comp. Med. Imag. & Graph., 26:309-319, 2002.

Domínguez, A. Rojas. Nandi,A. K. “Detection of masses in mammograms using enhanced multilevel thresholding segmentation and region selection based on rank”, In: Proceedings of the fifth conference on Proceeding of the Fifth IASTED International Conference: biomedical engineering. 2007

Duda, R.O., Hart, P.E.: “Pattern Classification and Scene Analysis”, In: Wiley-Interscience Publication, New York (1973)

Fukunaga, K.: “Introduction to Statistical Pattern Recognition”. 2nd ed. London: AcademicPress. 1990.

Homer M. J, “Imaging features and management of characteristically benign and probably benign lesions”, In: RadiolClin N Am 1987; 25 (5); pp. 939–951.

Homer M. J, “Breast imaging: Pitfalls, controversies and some practical thoughts”, In: Radiol Clin N Am **23** (1985) (3), pp. 459–472.

Hyvärinen A. and E. Oja. “A fast fixed-point algorithm for independent component analysis”, In: Neural Computation, 9(7):1483-1492, 1997.

INCa, Internet site address: http://www.inca.gov.br accessed in 04/12/2010.

Jain, A.K, F. Farrokhnia, “Unsupervised texture segmentation using Gabor filters”, In: Pattern Recognition 24 (12) (1991) 1167–1186. 2001

Marchini J. L, Heaton C, Ripley B D. “FastICA algorithms to perform ICA and Projection Pursuit”. (2004) Available at http://www.stats.ox.ac.uk/~marchini/software.html

Meyer J. E. , E. Amin, K.K. Lindfors, J.C. Lipman, P.C. Stomper and D. *Genest*, “Medulary carcinoma of the breast: Mammographic and US appearance”, In: Radiology, 1989; 79-82

Moore A.: “*K-means and Hierarchical Clustering -Tutorial Slides”*, 2001. Internet site address: http://www2.cs.cmu.edu/~awm/tutorials/kmeans.html

Newstead G. M, P.B. Baute and H.K. Toth, “Invasive lobular and ductal carcinoma: Mammographic findings and stage at diagnosis”, In: Radiology 1992; 184; pp. 623–627.

Sickles E. A.: “Breast masses: mammographic evaluation”, In: *Radiology* 1989; 173; pp. 297–303.

Suckling, J et al (1994): “The Mammographic Image Analysis Society Digital Mammogram Database”, In: Exerpta Medica. International Congress Series 1069 pp375-378.

Unser M, M. Eden, “Nonlinear operators for improving texture segmentation based on features extracted by spatial filtering”, In: IEEE Trans. Syst. Man Cybern. 20 (1990) 804–815.