AUTOMATIC LOCALIZATION OF REGIONS OF INTEREST (ROI) IN MAMMOGRAPHY IMAGES

GERMAN FELIPE TORRES VANEGAS

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERIAS FISICO-MECANICAS
ESCUELA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES
Bucaramanga
2016

AUTOMATIC LOCALIZATION OF REGIONS OF INTEREST (ROI) IN MAMMOGRAPHY IMAGES

GERMAN FELIPE TORRES VANEGAS

A Thesis Presented in Fulfillment of the Requirements for the Degree of Electronic Engineer

Advisor:

PhD. Said David Pertuz Arroyo

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERIAS FISICO-MECANICAS
ESCUELA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES
Bucaramanga
2016

DEDICATION

To my parents, Anibal and Aminta, for giving me all of them and for teaching me the value of patience and work well done. To my brothers, Julian and Carlos, for being models of excellence. Besides, I dedicate this work to Karol who was a great support at every moment.

CONTENT

LIST OF FIGURES

LIST OF TABLES

RESUMEN

Título:

Localización Automática de Regiones de Interés (ROI) en Imágenes Mamográficas.[1]

Autor: German Felipe Torres Vanegas[2]

Palabras Claves: mamografía, región de interés (ROI), sistema coordenado, características de textura, análisis de imagen.

En el análisis mamográfico, los radiólogos saben que la interpretación de diferentes patrones de imagen cambia con su localización en el seno. Sin embargo, la localización anatómica de características de imagen no es considerada por muchos de los sistemas de diagnóstico asistido por computador (CAD) propuestos en el estado del arte. De este modo, el rendimiento de tales sistemas puede ser incrementado al considerar regiones de interés (ROI) anatómicas en el análisis de imágenes mamográficas. Este trabajo introduce un sistema de detección de ROI que se ajusta automáticamente a la geometría y el tamaño del seno. El sistema propuesto localiza cinco diferentes ROI, el cual ha sido implementado y probado en mamogramas digitalizados. Adicionalmente, fue evaluado su rendimiento en el análisis de imagen mamográfica al estudiar las medidas de características computarizadas de textura en diferentes ROI del seno. Los experimentos fueron conducidos con 504 imágenes tomadas de la base de datos DDSM (Digital Database for Screening Mammography). Los resultados mostraron que las medidas de las características y su correlación con la presencia del cáncer cambian de una región a otra. Finalmente, el rendimiento en la tarea de la detección de imágenes cancerígenas, usando regresión logística, también fue evaluado. Los resultados revelaron que el sistema automático de detección de ROI propuesto supera en rendimiento el método tradicional de ROI cuadradas seleccionadas manualmente en la detección de imágenes cancerígenas.

[1] Trabajo de Grado modalidad en investigación

[2] Facultad de Ingenierías Físico Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y Telecomunicaciones. Director: PhD. Said David Pertuz Arroyo.

ABSTRACT

TITLE:

Automatic Localization of Regions of Interest (ROI) in Mammography Images.[3]

AUTHOR: German Felipe Torres Vanegas[4]

KEYWORDS: mammography, region of interest (ROI), coordinate system, texture features, image analysis

In mammographic analysis, radiologists know that the interpretation of different image patterns changes with their location in the breast. However, the anatomical location of image features is not considered by most of state of the art Computer Aided Diagnosis (CAD) systems. In this way, performance of such systems might be increased by analyzing mammographic images by considering anatomic regions of interest (ROI). This work introduces a ROI detection system that adjusts automatically to the geometry and the size of breast. The proposed system localizes five different anatomy-based ROIs and has been implemented and tested on digitized mammographic images. In addition, the performance in mammographic image analysis was assessed by studying the measures of computerized texture features in different breast ROIs. Experiments were conducted with 504 mammographic images taken from the Digital Database for Screening Mammography (DDSM). Results showed that measures of texture features and their correlation with the presence of cancer change in function of location of the different ROIs. Finally, the performance in the task of cancer image detection using logistic regression was also assessed. Results revealed that the proposed automatic ROI detection system outperforms the traditional method of manually selected squared ROIs in the detection of cancer images.

[3] Degree Project

[4] Faculty of Physics Mechanics Engineering. Electrical, Electronics Engineering and Telecommunications School. Director: PhD. Said David Pertuz Arroyo.

INTRODUCTION

Cancer is defined as an uncontrolled division of a group of cells in a particular location of the body. These extra cells can divide without interruption, forming masses commonly known as tumors [1]. In the United States, breast cancer is the second most common cancer after skin cancer [2]. In Colombia, breast cancer is the most common cancer in women. Each year more than 8686 women suffer from this disease and about 2649 of them die, which corresponds to a mortality rate two times higher than 30 years ago [3].

Currently, there are not effective ways to prevent the breast cancer since their causes are unknown. Nevertheless, detection and diagnostic in early stages is fundamental for increasing the probability of full recovery and for reducing the associated mortality rate [4]. Another approach to reduce breast cancer deaths is to identify women at higher risk (risk stratification) as candidates for preventive treatment [5], [6]. At present, screening mammography is one of the most widely used, effective, and low cost techniques for early detection and risk stratification of breast cancer. Unfortunately, due the human factor involved in the visual examination of mammograms, evaluations are susceptible to errors and subjectivity. For instance, previous studies have found error rates between 10% and 30% for detection of breast cancer in screening studies [7], [8]. Similarly, the assignment of density categories, one of the most important risk factors for breast cancer [9], [10], has shown various degrees of agreement among radiologists since previous studies have found a Cohen's © coefficient (a statistic which measures inter-rater agreement for categorical items) around 0.43-0.76 [11], [12]. In order to improve the accuracy of mammographic image analysis, researchers have performed several studies with the purpose of developing automatic computerized systems for aiding breast cancer detection and risk stratification.

One of the main challenges in the development of quantitative image analysis tools is incorporating human knowledge in computerized systems. For instance, when radiologists are examining a mammogram, they know that the interpretation of mammographic patterns change with their location. Specifically, it is well known that the likelihood of a given cancer type changes according to the anatomical breast region under analysis [13]. Experienced human readers can readily incorporate this information in their analysis. In contrast, most of the computer-aided-diagnosis (CAD) systems proposed in the literature do not consider such information. The hypothesis of this research is that adapting the analysis of the mammographic images according to anatomic regions of interest (ROI) can improve the performance of such systems. Previous studies conducted with ROIs manually selected by human readers have shown that, indeed, the performance of image analysis systems change as a function of the anatomical breast region [14]. This research proposes a system for fully-automated detection of anatomic breast ROIs. Experiments conducted on publicly available datasets show that the performance of mammography image analysis systems for cancer detection improves by means of the proposed ROI detection system.

Several researchers have addressed the use of ROIs for performing computerized breast image analysis tasks. Li et al. [14] investigated the effect of the size and location of the ROI in order to optimize their computerized method for identifying the presence of the BRCA1/BRCA2 gene, which is associated with an increased risk of developing breast cancer. In that study, squared ROIs were manually selected from the LCC (left cranio-caudal) view by human readers in five different regions: the central breast region immediately behind the nipple (retroareolar region), the center of the central breast region, the central left breast region (far away from nipple), the upper central breast region and the lower central breast region. Their results showed that there was a statistically significant decrease in the location was moved away from the central region behind the nipple

performance of the computerized texture features as the ROI. Other studies [15], [16], [17] have also proposed methods for measuring parenchymal tissue by analyzing the texture patterns of fibroglandular tissue in the retroareolar region with rectangular ROIs. The main limitation of the aforementioned works is the need for human interaction [14] and the use of fixed squared ROIs [14], [15], [16], [17]. On the one hand, human-annotated ROIs pose different limitations in terms of subjectivity, reproducibility, and scalability (i.e. the application of these methods to large datasets). On the other hand, fixed squared ROIs cannot adapt to a wide variety of breast shapes and sizes.

In this work a methodology for automatic ROI location for mammography image analysis is proposed. The presented method takes into account the breast geometry and automatically adjusts to diverse of shapes and sizes. In previous research, the most related work was carried out by Tahoces et al. [18]. In that work, they have analyzed mammographic parenchymal patterns in ROIs with different shapes. Specifically, Tahoces et al. [18] developed a method for the quantification of breast texture using different algorithms to classify mammograms into the four patterns described by Wolfe [9](N1, P1, P2 y Dy). In that work, ROIs are studied to localize the presence or absence of breast parenchymal tissue. Two types of different ROIs were automatically selected by computer as follows: (a) Squared ROIs of fixed size, and (b) ROIs of irregular shapes delimited by straight lines and the breast contour, located mainly in the medial zone of the breast. To the best of our knowledge, the problem of automatic detection of anatomically-oriented ROIs has not been tackled previously.

The goal of this research is to introduce a novel methodology for fully automated ROI detection for image analysis in mammograms. For this purpose, recent developments in the creation of anatomical coordinate systems in mammographic

images have been considered [19], [20], [21]. The proposed methodology is then used to assess the impact of the breast anatomy on the performance of mammography image analysis. For this purpose, logistic regression classifiers were applied at each ROI for breast cancer detection. Experiments were performed using a dataset of 504 images from the DDSM dataset. Results show that the performance of the classifiers changes with the location of the ROI.

The remainder of the paper is organized as follows. Firstly, the proposed methodology for automatic ROI detection is presented in chapter 2. Then, methods related to the image analysis based in ROIs (used in performed experiments) are described in chapter 3. Finally, experimental results and conclusions are summarized in chapters 4 and 5, respectively.

1. AUTOMATIC ROI DETECTION

The proposed methodology for automatic ROI location is based on a modification of the breast coordinate system (ST system) introduced by Pertuz et al. [20], which adapts to the breast geometry from the breast contour. For the sake of clarity, the proposed methodology is described in three steps: breast detection (chapter 2.1), ST coordinate system (chapter 2.2) and ROI localization (chapter 2.3). The first one explains a method to detect the breast contour. The second, details briefly the breast coordinate system proposed previously. The last, explains the proposed method for extracting different ROIs for each mammogram.

1.1 BREAST DETECTION

Breast segmentation is arguably the first processing step in CAD algorithms. Notwithstanding, breast segmentation remains a challenging task, whose performance depends on whether the analyzed mammogram is a digitized (scanned) image or a digital image. In contrast to digital images that can be segmented by means of simple pre-calibration or thresholding methods [22], [23], algorithms for breast contour detection in digitized images are more complex. In this work, the method for the segmentation of digitized mammograms is performed in two steps: *Scanning artifacts removal* and *Breast contour detection*. Figure 1 shows an example of this process.

1.1.1 Scanning artifacts removal: Tape artifacts are markings left by tapes, or other shadows, that appear as horizontal or vertical running strips [24]. Since these artifacts are commonly defined by straight lines, the algorithm to remove them is based on the straight line Hough transform [25]. In few words, an edge map is created through the application of a canny edge operator [26]. Then this

map is collected in a two-dimensional, (ρ, θ)-line-parameterization histogram of Hough space. Tape artifacts are approximated by straight lines at each side of the image (between $\theta = -10°$ and $\theta = 10°$ for left and right sides, and between $\theta = 80°$ and $\theta = 100°$ for upper and lower sides). Their angle and location are image-space is defined by the (ρ, θ) location of the global maximum of the Hough-space histogram. This process is illustrated in Figure 1(b).

1.1.2 Breast contour detection: Due to the image noise of the scanning process, the task of detecting the breast contour in a digitized mammogram can not be accurately solved by a simple thresholding method. A statistical technique was used in this work in order to detect the breast contour. Bearing in mind the rich texture in the breast region can be treated as samples form a normal distribution. Therefore, the Anderson Darling test is then applied to test the hypothesis that the observations are samples form a normal distribution $N(\mu, \sigma^2)$ [27]. In particular the Anderson Darling (AD) value for a small window of pixels with n samples is defined as:

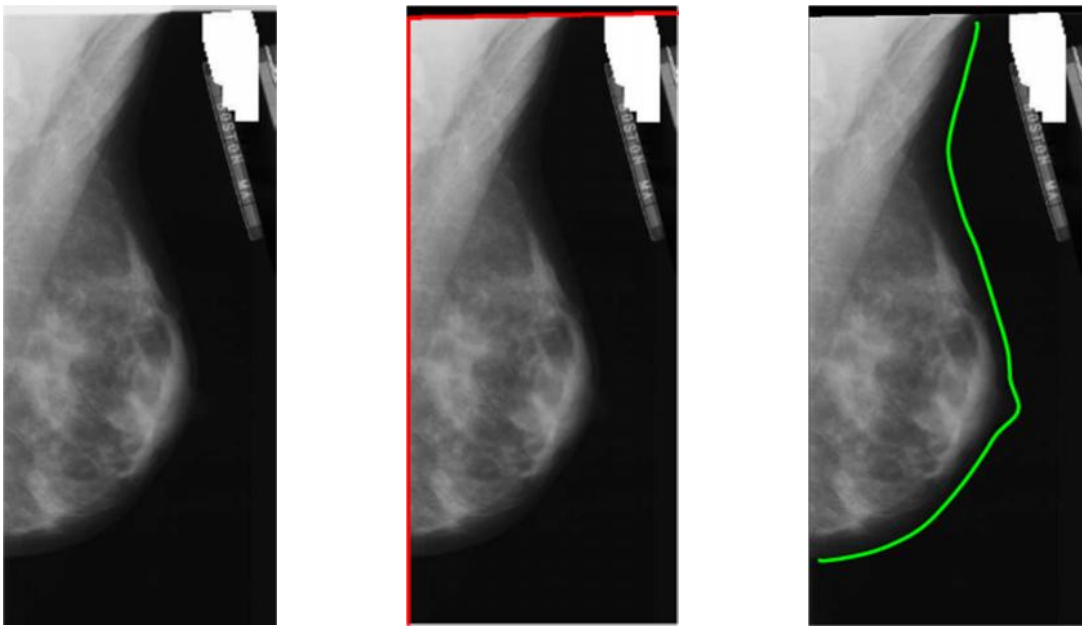
$$AD = -n - \sum_{i=1}^{n} \frac{2i-1}{n} [\ln F_0(x_i) + \ln (1 - F_0(x_{n+1-i}))] \tag{1}$$

where $\{x_i\}$ are sorted samples, $\{x_1 < \cdots < x_n\}$, and $F_0(x)$ is the cumulative normal distribution function. By applying the AD measure on every pixel (the sample set of a given pixel includes its $n-1$ nearest neighbors), an image of discrepancy is obtained which has the same size as the original mammogram. p-values are calculated on every pixel i according to Stephens et al. [28] as:

$$p(i) = \begin{cases} 1 - \exp[-13.43 + 101.1 \cdot AD(i) - 223.7 \cdot AD^2(i)], & if \quad 0 \le AD(i) < 0.2 \\ 1 - \exp[-8.31 + 42.8 \cdot AD(i) - 59.9 \cdot AD^2(i)], & if \quad 0.2 \le AD(i) < 0.34 \\ \exp[0.91 - 4.27 \cdot AD(i) - 1.38 \cdot AD^2(i)], & if \quad 0.34 \le AD(i) < 0.6 \\ \exp[1.29 - 5.71 \cdot AD(i) - 0.02 \cdot AD^2(i)], & if \quad 0.6 \le AD(i) < 13 \end{cases} \tag{2}$$

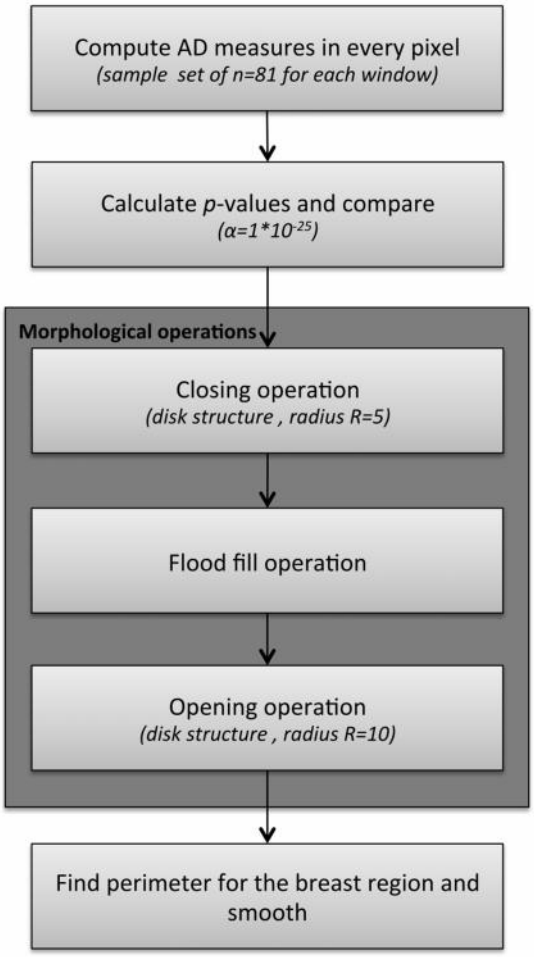
Results for the Anderson Darling test are obtained by comparing the p-values to a pre-defined confidence level α. Pixels in the background region lead to smaller p-values, and so, the hypothesis is rejected. In contrast, pixels in the breast region yield larger p-values, hence, the hypothesis is accepted. The obtained binary mask is post-processed by means of morphological operators in order to remove spurious artifacts. Finally, the found perimeter of the breast region is smoothed. All the steps for breast contour detection described previously are summarized in Figure 2. An example of the contour detected using the described method is shown in Figure 1(c).

Figure 1. Breast detection process for the normal case A_0022 right-MLO view from DDSM database. (a) Original image. (b) Artifacts detection using Hough transform. (c) Breast contour of the mammography image.



(a) **(b)** **(c)**

Figure 2. All steps in contour detection process.



1.2 ST COORDINATE SYSTEM

The proposed methodology for automatic ROI detection is based on a coordinate system that adapts to breast shape. Anatomy-based breast coordinate systems have been studied previously for the registration of mammography images [19], [20]. In this work, the system proposed by Pertuz at al., namely the ST coordinate system, was adapted for the automatic ROI detection. This coordinate system was chosen due to its flexibility and fast computation. The ST coordinate system can be interpreted as a mapping rule that assigns a coordinate pair (s, t), to each pixel in Cartesian coordinates (x, y). This mapping relies on the breast contour c in order

to adapt to the breast anatomy and shape. This process can be interpreted as a fully invertible, closed-form transform represented as:

$$(x, y) \xrightarrow{\mathcal{M}_c} (s, t) \xrightarrow{\mathcal{M}_c^{-1}} (x, y) \tag{3}$$

where $\mathcal{M}$ and $\mathcal{M}^{-1}$ are the *direct* and *inverse* mapping operators, respectively. The mapping operators transform an image from Cartesian to ST coordinates, and vice versa, such that $I(s,t) = \mathcal{M}_c\{I(x,y)\}$ and $I(x,y) = \mathcal{M}_c^{-1}\{I(s,t)\}$. As illustrated in Figure 3, the ST coordinate system (in Figure 3(c)) is generated from the breast contour as follows.

The s-coordinate is defined as the normalized surface distance, measured along the two parametric curves $\boldsymbol{\phi}^0$ and $\boldsymbol{\phi}^1$ from a reference point, so that $s \in [0,1]$. These two segments illustrated in Figure 3(b) can be approximated in parametric space by means of two N-degree parametric polynomials of s:

$$c = \begin{cases} \boldsymbol{\phi}^0(s) = \begin{bmatrix} x^0 \\ y^0 \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^{N} P_{0,n} s^n \\ \sum_{n=1}^{N} P_{1,n} s^n \end{bmatrix} \\ \boldsymbol{\phi}^1(s) = \begin{bmatrix} x^1 \\ y^1 \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^{N} P_{2,n} s^n \\ \sum_{n=1}^{N} P_{3,n} s^n \end{bmatrix} \end{cases} \tag{4}$$

where $P_{i,n}$ are constant coefficients for $i = 0,1,2,3$.

The t-coordinate is understood as the penetration distance within the tissue, measured from the breast's surface. Similarity as s, the domain of the t-coordinate is restricted to $t \in [0,1]$. Thus, the polynomials in (4), $\boldsymbol{\phi}^0$ and $\boldsymbol{\phi}^1$, correspond to the parametric curves for all points in the ST coordinate system with $t = 0$ and $t = 1$, respectively. In order to generate the parametric curves for all points at any t,

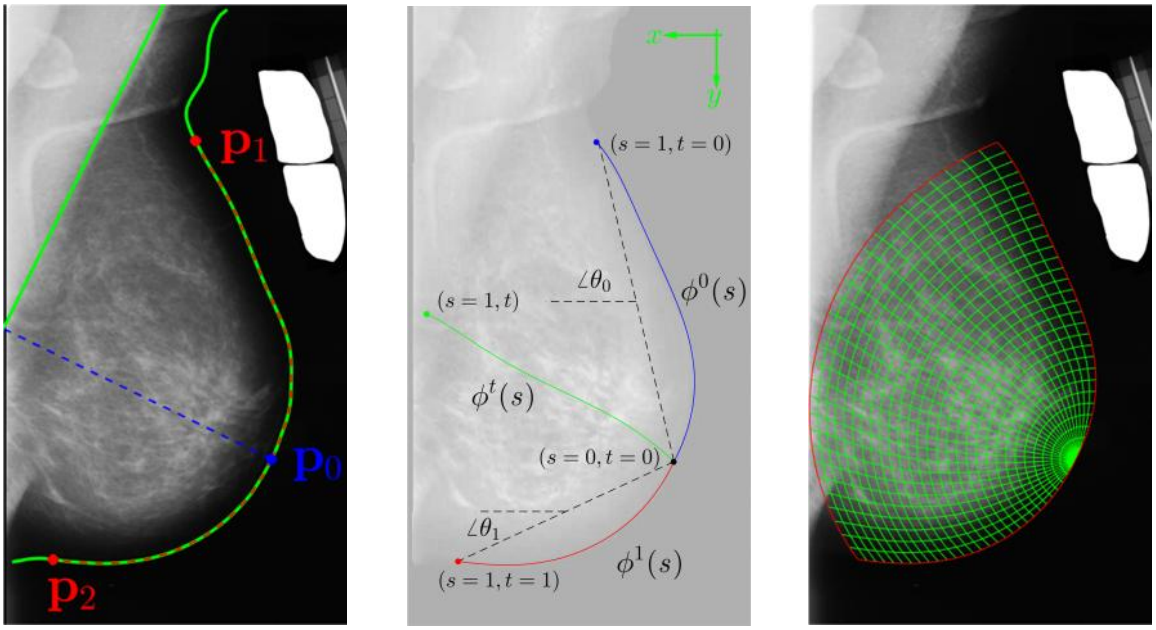
$\phi^t(s)$ (green curve in Figure 3(b), ϕ^0 and ϕ^1 are rotated and linearly combined. It is analytically expressed in (5), which relates each (x, y) pair with its corresponding (s, t) coordinates.

$$\phi^t(s) = \mathbf{R}(\theta(t)) \left[(t-1)\mathbf{R}(\theta_0)\phi^0(s) + t\mathbf{R}(\theta_1)\phi^1(s) \right] \tag{5}$$

where $\theta_0 = \arctan(y^0/x^0)\big|_{s=1}$ and $\theta_1 = \arctan(y^1/x^1)\big|_{s=1}$ are the angles of the extreme points of ϕ^0 and ϕ^1 respectively (as shown in Figure 3(b)), $\theta(t) = (\theta_1 - \theta_0)t + \theta_0$ is a rotation angle that goes from θ_0 to θ_1, and $\mathbf{R}(\theta)$ is the rotation matrix:

$$\mathbf{R}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{6}$$

Figure 3. ST coordinate system. (a) Contour, nipple and extreme points of the mammography image. (b) Generation of the ST coordinate system. (c) Coordinate system generated from de breast contour.



(a) **(b)** **(c)**

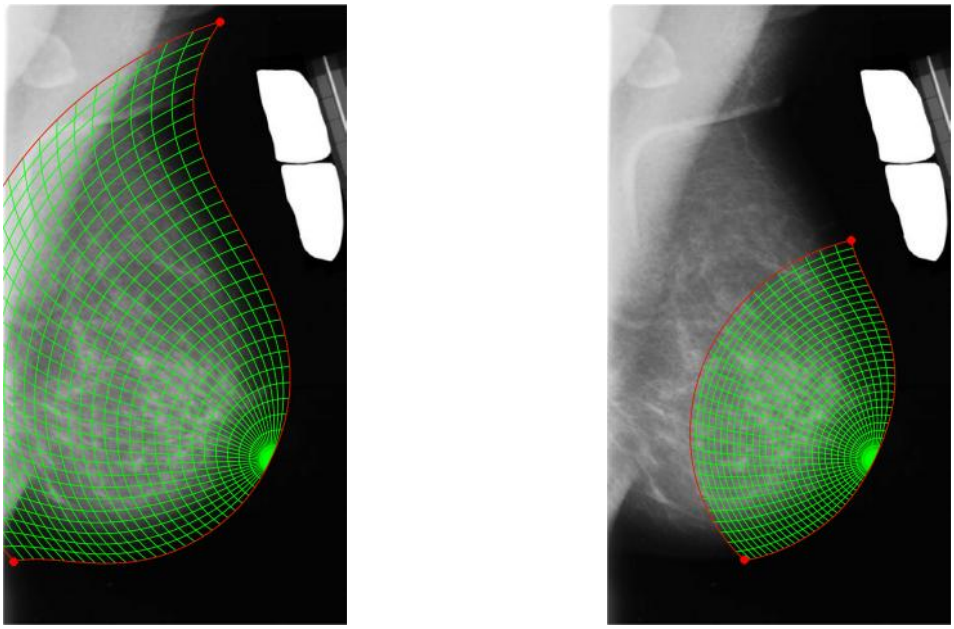
The mapping in (3) requires the specification of three reference points on the contour c: the origin, p_0, and the extreme points p_1 and p_2. The localization of these points in a mammogram example are shown in Figure 3(a). In the original work [20], these points were manually selected. In order to make the process fully-automatic, in this work the reference points are defined in two steps as follows. Firstly, following the work by Nasser et al. [21], p_0 is set at the nipple of the breast. Secondly, the two extreme points p_1 and p_2 are defined by performing a statistical analysis of their location on a subset of the full dataset used in this work. Each one of these two steps are detailed below.

A simple idea to localize the nipple is characterizing it as the farthest point from the pectoral line, which is identified using a previously validated algorithm based again on the straight line Hough transform [29]. In this case, a contrast enhancement and a Gaussian filter are applied before the canny edge operator. Contrast limited adaptive histogram equalization (CLAHE) algorithm is used for the contrast enhancement [30]. The resulting pectoral-boundary line may appear between $\theta = 10°$ and $\theta = 70°$, and it is defined by the global maximum of the Hough-space histogram. Figure 3(a) is showing the strategy used for the localization of the nipple point p_0 in a mammogram example. Notice that there are several methods in the state-of-the-art dealing with the nipple localization [31], [32]. The method used here was chosen due its simplicity and good results in the used dataset.

The location of the extreme points p_1 and p_2 is an important parameter of the ST coordinate system. As illustrated in Figure 4, these extreme points indicate what portion of the mammogram is considered as breast tissue. For instance, Figure 4(a) shows two extreme points whose location is too distant from the nipple. In this case, the resulting ST coordinate system includes information of the pectoral muscle and outside the image limits. In contrast, extreme points too close to the

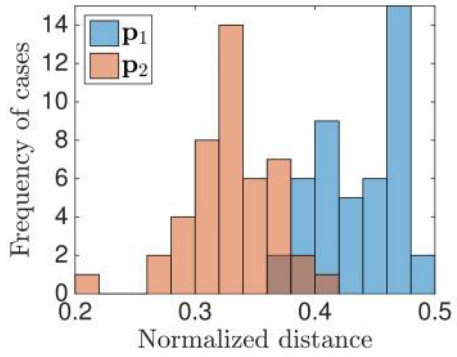
nipple (Figure 4(b)) will yield the exclusion of relevant information of the image. To situate the extreme points for the coordinate system, an experiment has been conducted with the aim of statistically describing the location of these points. For nearly 10% of the mammograms of our dataset (45 images), the extreme points were manually established onto the contour. By measuring the distance from the reference point p_0 (normalized over the entire length of the breast contour), p_1 and p_2 are located at normalized distances of $0.44(\pm 0.03)$ and $0.33\,(\pm 0.03)$, respectively. The histograms of the superior and inferior extreme points are shown in Figure 5. Based on these results, in this work, the superior and inferior points are situated as the points upon the breast contour with a normalized distance about of 0.44 and 0.33 respectively.

Figure 4. Effect of the location of the extreme points in the ST mapping. (a) Too distant from the nipple. (b) Too close to the nipple.



(a) **(b)**

Figure 5. Histograms of normalized distance of the extreme points.



1.3 AUTOMATIC ROI LOCATION

Due to the fact that the ST coordinate system automatically adapts to the breast shape and size, different anatomically-oriented ROIs can be defined in breast space by placing limits on the s and t coordinates. Thus, a region $\mathcal{R}$ is defined as the portion of the breast between S_a and S_b in the s-axis, and between T_a and T_b in the t-axis, as illustrated in Figure 6(a). Specifically, for this work, five anatomically-oriented regions are defined, where the n-th ROI is given by:

$$\mathcal{R}_n = \{(x, y) \in \mathcal{M}_c^{-1}\{s, t\} | S_{a,n} \leq s < S_{b,n}, T_{a,n} \leq t < T_{b,n}\} \quad (7)$$

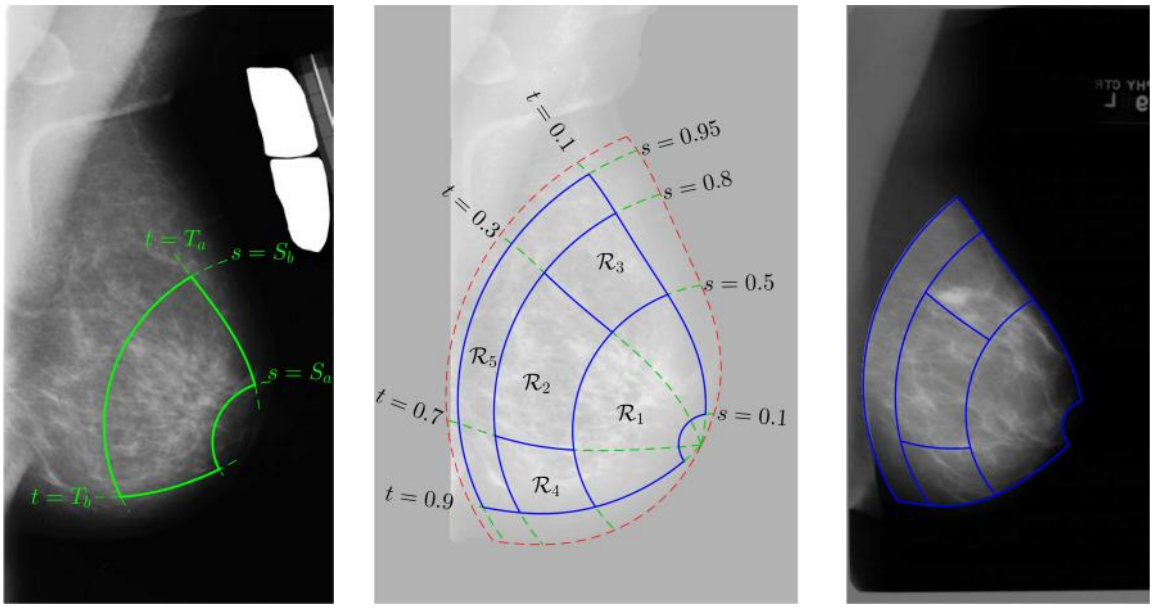
where $S_{a,n}$, $S_{b,n}$, $T_{a,n}$, and $T_{b,n}$ are constant limits for each ROI $\mathcal{R}_n$, for $n = 1,2, \ldots,5$. The specific limits used in this work are listed in Table 1 and illustrated in Figure 6(b). These limits were established in order to delimit five anatomically meaningful regions, as described below:

- $\mathcal{R}_1$: Named the retroareolar region is located immediately behind the nipple.
- $\mathcal{R}_2$: It is the middle-central region which is situated in the center of the breast.
- $\mathcal{R}_3$: This is denominated the upper-central region and it is placed in the superior part of the breast.
- $\mathcal{R}_4$: The lower-central region, it is placed in the inferior part of the breast.

- $\mathcal{R}_5$: This consists of the region located in the posterior part of the breast. It borders with the pectoral muscle, so it has been titled the pre-pectoral region.

For illustration, an example of the ROIs extracted with this methodology can be observed in Figure 6(c).

Figure 6. Automatic ROI detection. (a) Method for ROI location (b) Limits for ROIs. (b) ROIs extracted for the cancer case A_1122.



(a) **(b)** **(c)**

Table 1. Limits for all ROIs in coordinates (s, t)

$\mathcal{R}_n$	s-limits		t-limits	
	$S_{a,n}$	$S_{b,n}$	$T_{a,n}$	$T_{b,n}$
$\mathcal{R}_1$	0.10	0.50	0.10	0.90
$\mathcal{R}_2$	0.50	0.80	0.30	0.70
$\mathcal{R}_3$	0.50	0.80	0.10	0.30
$\mathcal{R}_4$	0.50	0.80	0.70	0.90
$\mathcal{R}_5$	0.80	0.95	0.10	0.90

2. ROI-BASED IMAGE ANALYSIS AND CANCER DETECTION

With the purpose of researching the usefulness of the proposed automatic ROI detection system, the visual information of the mammograms was studied through the comparison of computerized features extracted in each one of the five ROIs defined above. Also, the performance of a classification algorithm in the task of detecting cancer images was compared based on the features calculated using different ROIs. Receiver operating characteristic (ROC) analysis [33] was used to measure the performance of a classification algorithm employing the leave-one-out validation technique in its evaluation [34]. Before training the classification algorithm, a feature selection procedure was performed in order to reduce the dimensionality of data. This chapter explains the methods of features extraction, feature selection, and image classification used in this work.

2.1 FEATURES EXTRACTION

In image processing, features are computerized measures used in order to extract visual information of a given image. In mammography, breast tissues have different shape and composition, so these can be characterized using computer-extracted texture features. Based on an extensive review of the literature, five groups of features have been considered in this work: First-order statistics, GLRLM (gray level run-length methods), gradient features, Haralick, and Gabor texture features. In total, 101 features were implemented in this work for the image analysis.

2.1.1 First-order statistics: This type of features is based on the analysis of the gray level histogram. The first-order histogram $P(i)$ is defined as the percentage of pixels with gray level i in the region of analysis of the image. Based on the histogram, different moments can be computed, such as the mean,

skewness, kurtosis, etc. In total, 11 first order statistics have been considered. For further details about these features and their computation, the reader is referred to [14].

2.1.2 Gray level run-length method (GLRLM): The GLRLM is based on calculating the number of consecutive pixels having the same gray level value for each one of the four directions ($0°, 45°, 90°$, and $135°$) defined as the run length [35], [36]. For each direction of the runs, a gray level run length matrix is calculated, thus yielding to four matrices for each ROI in the image. Subsequently, these matrices are used for the computation of five descriptors, namely the *short run emphasis*, *long run emphasis*, *gray level nonuniformity*, *run length nonuniformity*, and *run percentages*. By computing the mean and the standard deviation of those descriptors, a total of 10 GLRLM features are obtained.

2.1.3 Gradient features: The gradient of an image measures the intensity change in every pixel of the image. Mathematically, as the image $I(x, y)$ is a two-variable function, the gradient is given by the derivatives in the horizontal and vertical directions. Gradients in each direction can be approximately computed through the convolution of the image with the *Sobel* operator and then combined to produce the magnitude of the gradient. A total of 6 features are calculated from the magnitude of the image gradient: mean, uniformity, standard deviation, smoothness, skewness, and entropy [37].

2.1.4 Haralick texture features: Haralick texture features are mathematical descriptors calculated on the *gray level co-occurrence matrix* (GLCM). The GLCM displays the distribution of co-occurring pixel values at a given offset. In total, 26 features are obtained by computing the mean and range of 13 descriptors

proposed by Haralick [38] calculated over the symmetric GLCM in the four directions ($0°, 45°, 90°$, and $135°$): *angular second moment*, *contrast*, *correlation*, *sum of squares*, *inverse difference moment*, *sum average*, *sum variance*, *sum entropy*, *entropy*, *difference variance*, *difference entropy*, *information measures of correlation 1*, *information measures of correlation 2*.

2.1.5 Gabor texture features: Gabor features are the result of measuring the energy of the image in different frequency bands and orientations [39]. This is accomplished by filtering the image with a filter bank. For the image analysis, 48 features were extracted by calculating the mean and standard deviation of the energy corresponding to the bank of 24 filters proposed by Manjunath for texture analysis [40].

2.2 FEATURE SELECTION

Due to the large number of image descriptors considered, feature selection was used in order to remove descriptors that were redundant or irrelevant, and avoid overfitting. Feature selection consists in selecting a relevant subset of features to construct a classification model for a particular task. The aim is reduce the dimensionality of the model without incurring much loss of information. The sequential forward feature selection method [41] is used in this work, with the misclassification rate of a logistic regression classifier as the criterion function and leave-one-out assessment. This feature selection was performed with the computed features in all ROIs.

2.3 IMAGE CLASSIFICATION

Breast cancer detection in mammographic images is one of the main tasks performed by CAD systems. This is accomplished by using classification

algorithms. One of the traditional methods provided in the pattern recognition theory is the binary logistic regression [42], [43]. Contextualizing to this problem, classifying an image I consists in assigning it to a class $\mathcal{C} = 1$ (cancer) or $\mathcal{C} = 0$ (normal), with the highest posterior probability $p(\mathcal{C}|\mathbf{x})$ given the set of features $\mathbf{x}$. The logistic regression model arises from the desire to model the difference between the logarithms of the class-conditional density functions via linear function of $\mathbf{x}$:

$$\log\left[\frac{p(\mathbf{x}|\mathcal{C}=\mathbf{0})p(\mathcal{C}=0)}{p(\mathbf{x}|\mathcal{C}=\mathbf{1})p(\mathcal{C}=1)}\right] = \mathbf{w}^T\mathbf{x} + b \tag{8}$$

where $\mathbf{w}$ and b are the parameters of the model.

Through this assumption the posterior probability $p(\mathcal{C}=0|\mathbf{x})$ can be expressed as a logistic sigmoid function of $\mathbf{w}^T\mathbf{x} + b$:

$$p(\mathcal{C}=0|\mathbf{x}) = \sigma(\mathbf{w}^T\mathbf{x} + b) = \frac{1}{1 + e^{-(\mathbf{w}^T\mathbf{x}+b)}} \tag{9}$$

Thus, given a data set of P images for training, the log likelihood function assuming the logistic regression model becomes:

$$L(\mathbf{w}, b) = \sum_{i=1}^{P} \mathcal{C}_i \log\sigma(\mathbf{w}^T\mathbf{x}_i + b) + (1 - \mathcal{C}_i)\log\left(1 - \sigma(\mathbf{w}^T\mathbf{x}_i + b)\right) \tag{10}$$

Gradient ascent is applied to iteratively maximize the likelihood function and finding the parameters of the logistic regression model. In particular, the actualization equations for $\mathbf{w}$ and b are:

$$\mathbf{w}^{new} = \mathbf{w} + \eta \sum_{i=1}^{P} \left(\mathcal{C}_i - \sigma(\mathbf{w}^T \mathbf{x}_i + b) \right) \mathbf{x}_i \tag{11}$$

$$b^{new} = b + \eta \sum_{i=1}^{P} \left(\mathcal{C}_i - \sigma(\mathbf{w}^T \mathbf{x}_i + b) \right) \tag{12}$$

where η is the learning rate. It must be a scalar small enough to ensure convergence of the method.

3. EXPERIMENTAL RESULTS

In order to assess the usefulness of the proposed ROI-based methodology for mammographic image analysis, four types of experiments have been conducted:

- Analysis of mammographic maps.
- Statistical comparisons.
- Correlation analysis.
- Analysis of images classification.

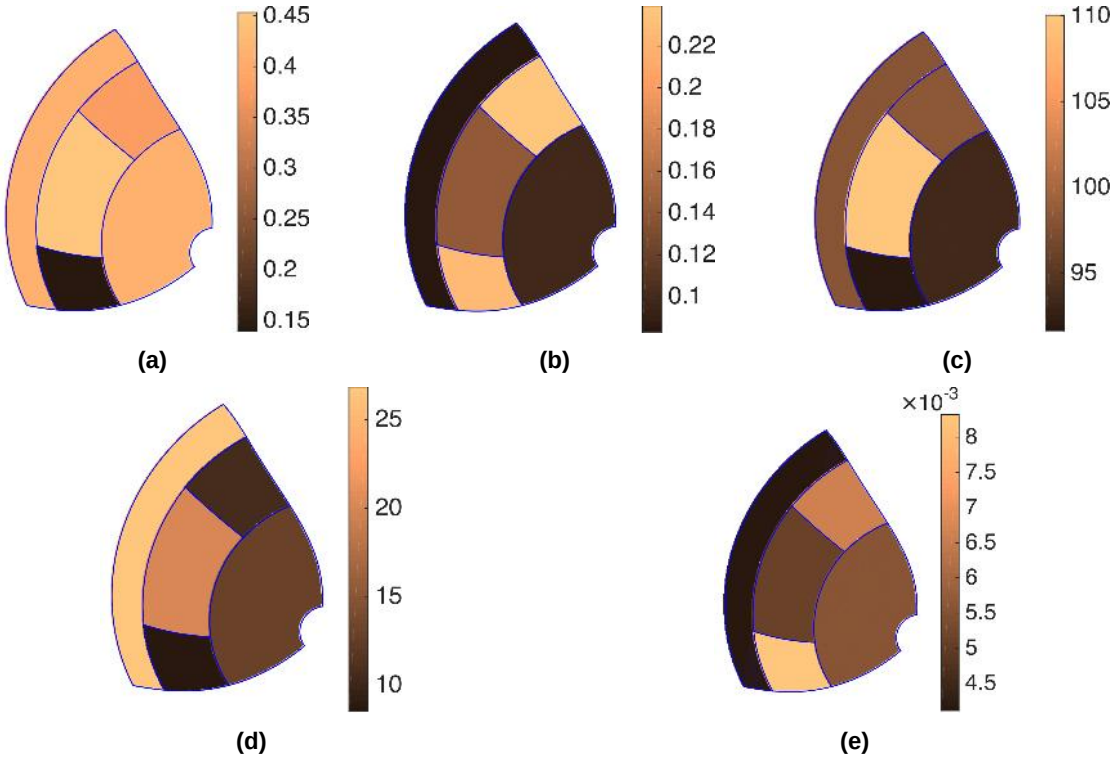
The image dataset used in this study was taken from the Digital Database for Screening Mammography (DDSM) [44]. The DDSM database contains approximately 2500 cases of mammograms from several medical institutions in the USA. Only images with MLO (medio-lateral oblique) view have been considered in this work, which were scaled by a factor of 8 and converted to 8-bits unsigned integer format to speed up the processing. For the first three experiments only cancer cases were considered, using 150 mammograms in the analysis. For the fourth one, 504 mammographic images were taken into account (150 cancer and 354 normal cases). Results for every experiment are summarized in this chapter.

3.1 ANALYSIS OF MAMMOGRAPHIC MAPS

Several mammographic maps were constructed as a function of the ROIs to illustrate the benefits of the proposed methodology as an analysis tool of mammographic patterns. With the ground truth of the cancer cases, two different maps were generated related to the presence of breast cancer. The first one (Figure 7(a)), consists in the map of cancer incidence (CI) and it represents the proportion of cancer masses that appear in each ROI [13]. It shows that the middle-central region is the highest cancer incidence zone ($CI > 45\%$) and the

lower-central region is the lowest incidence area $(CI < 15\%)$. The second map (Figure 7(b)), illustrates the average percentage of cancerous area in each ROI, which is understood as the area occupied by a cancerous tissue over the area of the respective ROI. The results for this map showed a range of variation between $23.9 \sim 8.3\ \%$, which is maximum in the upper-central region. On the other hand, mammographic maps were made with the average value at each ROI for three different features: the average of the gray level (Figure 7(c)), the contrast measured in the GLCM (Figure 7(d)), and the standard deviation of the energy of a Gabor filter output (Figure 7(e)). The two first are shown due their relationship with the density of fibroglandular tissues, and the third because it presents the highest correlation on average with the average percentage of cancerous area.

Figure 7. Mammographic maps. (a) Cancer incidence map. (b) Map of cancerous percent area. (c) Average of the gray level. (d) Average contrast. (e) Average of the deviation of the output energy of Gabor filter 21.



3.2 STATISTICAL COMPARISON

The statistical hypothesis test t-test was performed to determine if the measures of a feature are significantly different from one ROI to another $(p < 0.05)$. Specifically, each image feature is computed at each ROI and compared against the same feature computed at the remaining ROIs. Since five different ROIs were considered in this work, a total of 10 pairwise comparisons are performed for each feature ($\mathcal{R}_1$ vs $\mathcal{R}_2$, $\mathcal{R}_1$ vs. $\mathcal{R}_3$, $\mathcal{R}_1$ vs. $\mathcal{R}_4$, etc.). For instance, Table 2 shows the p-values corresponding to pairwise comparisons of the contrast (GCLM) feature computed in each ROI. This table reveals that this feature had statistically significant differences among all the ROIs considered. In contrast, Table 3 shows only two significant differences among ROIs for the balance 2 (first-order statistics) feature.

Table 2. p-values of pairwise t-test comparisons for contrast (GLCM) feature

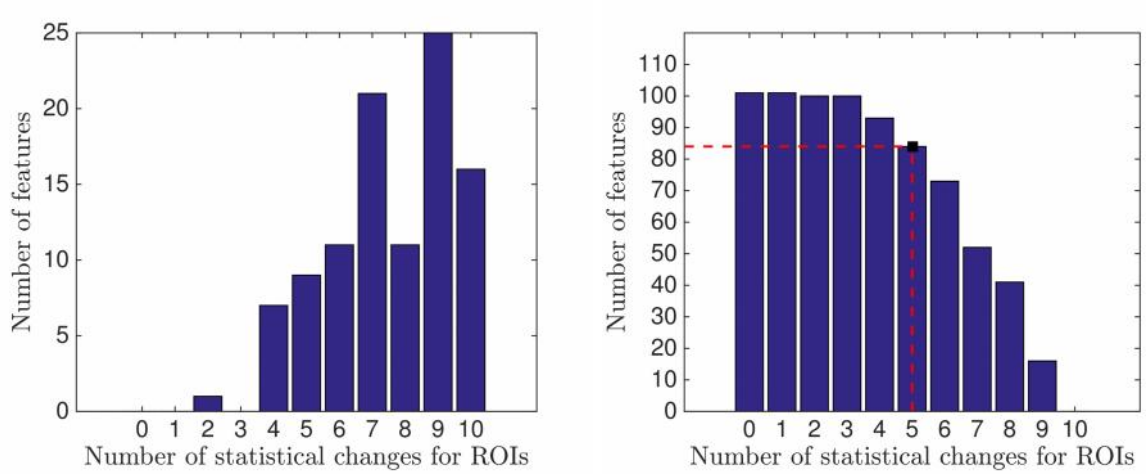
p-values	$\mathcal{R}_2$	$\mathcal{R}_3$	$\mathcal{R}_4$	$\mathcal{R}_5$
$\mathcal{R}_1$	$5.56 \cdot 10^{-59}$	$9.55 \cdot 10^{-106}$	$2.42 \cdot 10^{-113}$	$7.14 \cdot 10^{-40}$
$\mathcal{R}_2$	–	$2.03 \cdot 10^{-92}$	$2.22 \cdot 10^{-119}$	$3.41 \cdot 10^{-10}$
$\mathcal{R}_3$		–	$1.73 \cdot 10^{-66}$	$1.10 \cdot 10^{-97}$
$\mathcal{R}_4$			–	$1.59 \cdot 10^{-115}$

Table 3. p-values of pairwise t-test comparisons for balance 2 (first-order statistics) feature

p-values	$\mathcal{R}_2$	$\mathcal{R}_3$	$\mathcal{R}_4$	$\mathcal{R}_5$
$\mathcal{R}_1$	0.72	0.86	0.28	0.65
$\mathcal{R}_2$	–	$\mathbf{6.40 \cdot 10^{-4}}$	**0.01**	0.27
$\mathcal{R}_3$		–	0.09	0.64
$\mathcal{R}_4$			–	0.52

The histogram shown in Figure 8(a) resumes the number of features that obtain 0 to 10 statistical differences. From these results, is found that all the features present at least 2 statistical variations in ROIs. Conversely, there are 16 features that exhibit significant changes of measurement from one ROI to another. Moreover, 84 features present more than 5 significant differences between ROIs (as is shown in Figure 8(b)) and all features presents more or equal than 2 significant differences between ROIs. So that in general, measures of features change depending with the ROI where they are calculated.

Figure 8. Summary of t-test comparisons. (a) Histogram of t-test comparisons. (b) Inverse cumulative histogram of t-test comparisons



(a) **(b)**

3.3 CORRELATION ANALYSIS

Additionally, the correlation between features and the cancer incidence, as measured by the percentage of cancerous area, was calculated for each ROI. Absolute values of correlation coefficients are summarized for each ROI in the box plot of Figure 9. Due the diversity of features taken into account, the variation range is considerably high in all ROIs. However, in average, features computed in the upper central region ($\mathcal{R}_3$) show a higher correlation to the cancer incidence.

Figure 10 shows the percentage of features that obtained the highest correlation coefficient in the specific ROI. The maximum coefficient value for each ROI and the feature associated are listed in Table 4. Results reveal that most of features present maximum correlation in the upper-central region (79% of features) and the highest correlation coefficients obtained for each ROI are Gabor features. Besides, in Table 5 are expressed the correlation coefficients in every ROIs for the features named in Table 4 in order to show the variation for each ROI. It shows that the best correlation coefficient was 0.451 in the upper-central region with the feature GAB19_std. This table also reveals that, depending of the ROI location, a feature can be more associated with the presence of a cancer.

Figure 9. Box plot of correlation coefficients for each ROI.

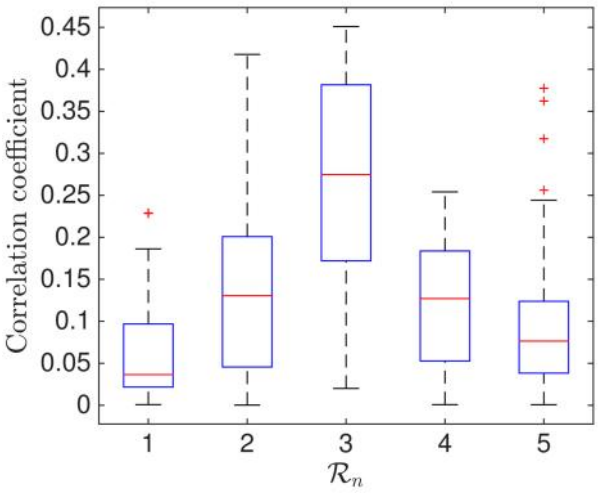


Figure 10. Pie chart of percentage of features that obtain maximum correlation in each ROI

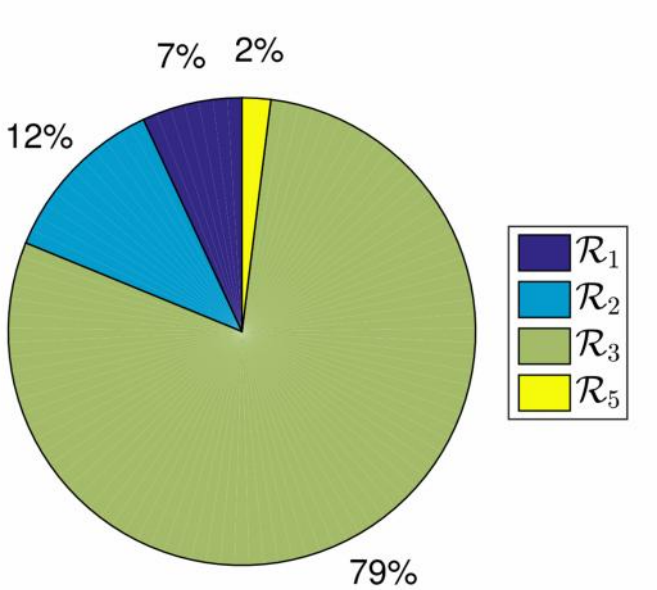


Table 4. Maximum correlation obtained for each ROI.

ROI	Maximum absolute value of correlation		
	Value	Feature	Description
$\mathcal{R}_1$	0.2290	GAB02_avg	Average output energy of Gabor filter 2
$\mathcal{R}_2$	0.4178	GAB21_std	Deviation output energy of Gabor filter 21
$\mathcal{R}_3$	0.4510	GAB19_std	Deviation output energy of Gabor filter 19
$\mathcal{R}_4$	0.2541	GAB24_std	Deviation output energy of Gabor filter 24
$\mathcal{R}_5$	0.3774	GAB23_std	Deviation output energy of Gabor filter 23

Table 5. Correlation coefficients in all ROIs for features shown in Table 4.

Features	Absolute value of correlation coefficients				
	$\mathcal{R}_1$	$\mathcal{R}_2$	$\mathcal{R}_3$	$\mathcal{R}_4$	$\mathcal{R}_5$
GAB02_avg	**0.2290**	0.0706	0.2193	0.0504	0.1263
GAB21_std	0.0023	**0.4177**	0.3909	0.1290	0.3173
GAB19_std	0.0169	0.0274	**0.4510**	0.1342	0.0614
GAB24_std	0.0008	0.2657	0.2870	**0.2541**	0.0826
GAB23_std	0.0322	0.2445	0.4154	0.1149	**0.3774**

3.4 ANALYSIS OF IMAGES CLASSIFICATION

Results of the feature selection process described in chapter 3.2 yielded 12 features in the model, which are specified in Table 6. In the performance analysis for the task of cancer detection, ROC curves and areas under the ROC curve (A_z values) for the binary logistic regression classifier with the selected features in the five defined ROIs are shown in Figure 11. It shows the best performance is found in the retroareolar region ($\mathcal{R}_1$) although without large difference compared to the rest of regions.

In order to compare the proposed automatic ROI detection methodology with the traditional approach (manually selected squared ROIs), ROC analysis for image classification experiment was performed again with features computed in a square ROI for each image of the dataset. For the traditional method, square ROIs were manually localized in the retroareolar region and their sizes were varied depending of the breast size according to the work by Wei et al. [16]. For the dataset, the size of established square ROIs were 128x128, 112x112, 96x96, and 64x64 pixels for 101, 201, 183, and 19 cases, respectively. So, the proposed method revealed a better performance based on the comparison results presented in Figure 12.

Figure 11. ROC curves and A_z values for image cancer detection experiment in all automatic ROIs.

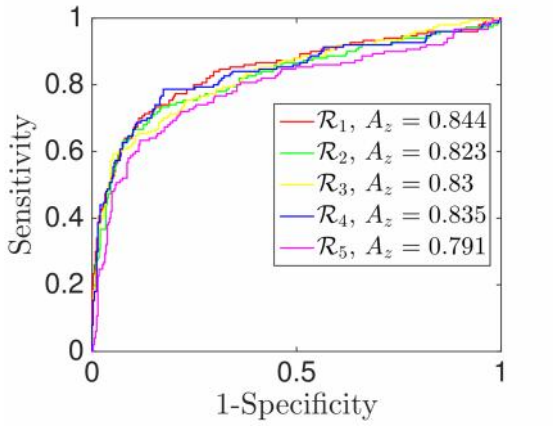


Figure 12. Comparison of ROC curves and A_z values for image cancer detection experiment in the retroareolar region

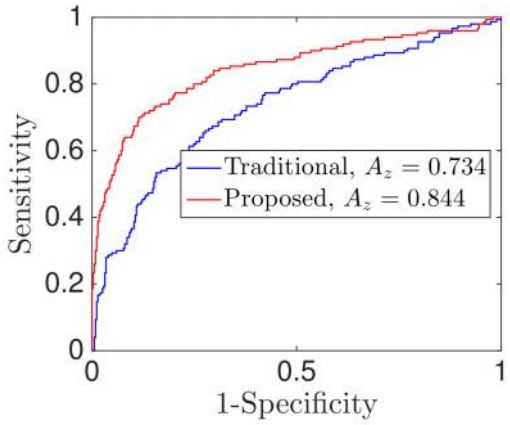


Table 6. Description of selected features.

Feature abbreviation	Category	Description	Feature abbreviation	Category	Description
GLCM06_avg	Haralick	Average of sum-average measure in all directions	GAB02_std	Gabor	Deviation output energy of Gabor filter 2
GLCM06_rang	Haralick	Range of sum-average measure in all directions	GAB03_avg	Gabor	Average output energy of Gabor filter 3
GLCM12_rang	Haralick	Range of information measure of correlation 1 in all directions	GAB06_std	Gabor	Deviation output energy of Gabor filter 6
GLRL02_std	GLRLM	Deviation of long-runs emphasis measure in all directions	GRAD06	Gradient	Entropy of magnitude gradient
GLRL03_std	GLRLM	Deviation of gray level nonuniformity measure in all directions	FOS09	First-order statistics	Balance 2
GLRL04_avg	GLRLM	Average of run-length nonuniformity measure in all directions	FOS10	First-order statistics	Skewness

4. DISCUSSION AND CONCLUSIONS

This research presents a novel automatic ROI detection system that localizes five distinct ROIs in the breast. It is characterized as a fully-automatic method that takes into account the breast anatomy and adapts to the breast size and shape. This is achieved by the application of a mammographic coordinate system, specifically the ST coordinate system, which only depends on the breast contour.

The application of this system allowed incorporating breast anatomical information in mammography image analysis. For instance, the presence of cancer was studied as a function of the location of malignant tissue. For the used image dataset, it was found that the zone of higher cancer incidence is the middle-central region, but these lesions occupy a bigger percentage of area in the upper-central region.

A large set of texture features was analyzed in the proposed ROIs. It was proved both visually and statistically that, in general, computerized texture features measured at different ROIs show different values. Specifically, most of the performed pairwise comparisons using t-test resulted in statistically significant differences when features are measured at different ROIs.

The relationship between texture features and cancer tissues was also studied. This was assessed through the calculation of the Pearson's correlation coefficient of each feature with the percentage of cancer area for each ROI. Results evidenced that the association of cancer with image-based features is more straightforward in the upper-central region, inasmuch as most of features showed a

higher correlation in that ROI. This finding has a clinical sense for radiologists, since the evaluation of a suspicious tissue is an easier task in the upper-central region than, for instance, in the retroareolar region, because the later contains a greater presence of fibroglandular and parenchymal tissues. By analyzing the features that yielded maximum correlation with cancer tissue for each ROI, different Gabor texture features have been noted and their values change in all ROIs as well. These outcomes suggest that cancer tissues have different orientation and spatial frequency components depending on their location.

Image classification experiments showed that the ROI with the highest performance in the detection of cancer images is the retroareolar region, which is in agreement with previous findings [14]. Notwithstanding, the difference in performance was not large compared with the other ROIs. Interestingly enough, given an A_z value of 0.844 in the retroareolar region, with only around 40% of cancer cases being located in this ROI, it is feasible that the classifier recognizes cancers that are situated in other ROIs. In consequence, this suggests that there must exist some type of pattern in tissues that predispose the cancer appearance.

Comparing the performance of two algorithms trained with features in the retroareolar region, the first with the proposed automatic localization method and the other with a manual square ROIs technique, the proposed method outperforms the classical method of manually selected ROIs for image analysis. This may be because in the square traditional method a fewer area of the retroareolar region is taken into account.

In spite of the automatic ROI detection system was tested in a high number of digitized mammograms, specifically for images from the DDSM database, it does

not work for the complete database, mainly due difficulties in the obtained breast contour by the algorithm used for its detection. However, the method for breast segmentation can be replaced for a most robust algorithm which enables applying the methodology in all the database, or even by algorithms that allow the use in full-field digital mammography (FFDM) images. This is possible because the methodology for ROI detection is based in the ST coordinate system, which is independent to the breast contour detection technique.

Finally, the findings in this work suggest that future work related to mass detection, risk stratification, or segmentation of fibroglandular tissue should take into account the localization of patterns and the proposed ROI detection system could be useful in this scope.

5. REFERENCES

[1] SCHULZ, Wolfgang. Molecular Biology of Human Cancers. Dordrecht: Springer Science & Business Media, 2005.

[2] NATIONAL CANCER INSTITUTE. Breast Cancer-for patients [Online]. <http://www.cancer.gov/types/breast> [cited 25 november 2015]

[3] LIGA CONTRA EL CANCER. Cancer de Seno: Magnitud del Problema [Online]. <http://www.ligacontraelcancer.com.co/cancer-de-seno/magnitud-del-problema/> [cited 25 november 2015]

[4] TANG, Jinshan, RANGAYYAN, R.M, XU, Jun, EL NAQA, I. and YANG, Yongyi. Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography. In: IEEE Transactions on Information Technology in Biomedicine. March, 2009. vol. 13, no. 2, p. 236-251.

[5] GAIL, Mitchell, COSTANTINO, Joseph, BRYANT, John, CROYLE, Robert, FREEDMAN, Laurence, HELZLSOUER, Kathy and VOGEL, Victor. Weighing the risks and benefits of tamoxifen treatment for preventing breast cancer. In: Journal of the National Cancer Institute. 1999. vol. 91, no. 21, p. 1829-1846.

[6] ROCKHILL, Beverly, SPIEGELMAN, Donna, BYRNE, Celia. HUNTER, David and COLDITZ, Graham. Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. In: Journal of the National Cancer Institute. 2001. vol. 93, no. 5, p. 358-366.

[7] KERLIKOWSKE, Karla, CARNEY, Patricia, GELLER, Berta, MANDELSON, Margaret, TAPLIN, Stephen, MALVIN, Kathy, ERNSTER, Virginia, URBAN, Nicole, CUTTER, Gary, ROSENBERG, Robert and BALLARD-BARBASH, Rachel. Performance of Screening Mammography among Women with and

without a First-Degree Relative with Breast Cancer. In: Annals of Internal Medicine. December, 2000. vol. 133, no. 11, p. 855-863.

[8] BIRD, R. E., WALLACE, T. W. and YANKASKAS, B. C.. Analysis of cancers missed at screening mammography. In: Radiology. September, 1992. vol. 184, no. 3, p. 613-617.

[9] WOLFE, John. Breast patterns as an index of risk for developing breast cancer. In: American Journal of Roentgenology. 1976. vol. 126, no. 6, p. 1130-1137.

[10] VACHON, Celine, BRANDT, Kathleen, GHOSH, Kartik, SCOTT, Christopher, MALONEY, Shaun, CARSTON, Michael, PANKRATZ, V. Shane and SELLERS, Thomas. Mammographic breast density as a general marker of breast cancer risk. In: Cancer Epidemiology Biomarkers & Prevention. 2006. vol. 15, no. 12, p. 11.

[11] BERG, Wendie, CAMPASSI, Cristina, LANGENBERG, Patricia and SEXTON, Mary. Breast Imaging Reporting and Data System: inter-and intraobserver variability in feature analysis and final assessment. In: American Journal of Roentgenology. 2000. vol. 174, no. 6, pp. 1769-1777.

[12] MARTIN, Katherine, HELVIE, Mark, ZHOU, Chuan, ROUBIDOUX, Marilyn, BAILEY, Janet, PARAMAGUL, Chintana, BLANE, Caroline, KLEIN, Katherine, SONNAD, Seema and CHAN, Heang-Ping. Mammographic Density Measured with Quantitative Computer-aided Method: Comparison with Radiologists' Estimates and BI-RADS Categories 1. In: Radiology. 2006. vol. 240, no. 3, p. 656-665.

[13] CANCER RESEARCH UK. Breast cancer incidence (invasive) statistics [Online]. <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive> [cited 13

april 2016].

[14] LI, Hui, GIGER, Maryellen, HUO, Zhimin, OLOPADE, Olufunmilayo, LAN, Li, WEBER, Barbara and BONTA, Ioana. Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: Effect of ROI size and location. In: Medical Physics. March, 2004. vol. 31, no. 3, p. 549-555.

[15] WANG, Xingwei, LEDERMAN, Dror, TAN, Jun, WANG, Xiao and ZHENG, Bin. Computerized prediction of risk for developing breast cancer based on bilateral mammographic breast tissue asymmetry. In: Medical engineering & physics. 2001. vol. 33, no. 8, p. 934-942.

[16] WEI, Jun, CHAN, Heang-Ping, WU, Yi-Ta, ZHOU, Chuan, HELVIE, Mark, TSODIKOV, Alexander, HADJIISKI, Lubomir and SAHINER, Berkman. Association of computerized mammographic parenchymal pattern measure with breast cancer risk: a pilot case-control study. In: Radiology. 2011. vol. 260, no. 1, p. 42-49.

[17] HUO, Zhimin, GIGER, Maryellen, OLOPADE, Olufunmilayo, WOLVERTON, Dulcy, WEBER, Barbara, METZ, Charles, ZHONG, Weiming and CUMMINGS, Shelly. Computerized analysis of digitized mammograms of BRCA1 and BRCA2 gene mutation carriers 1. In: Radiology. 2002. vol. 225, no. 2, p. 519-526.

[18] TAHOCES, P. G., CORREA, J., SOUTOS, M., GOMEZ, L. and VIDAL, J. J.. Computer-assisted diagnosis: the classification of mammographic breast parenchymal patterns. In: Physics in Medicine and Biology. 1995. vol. 40, no. 1, p. 103.

[19] BRANDT, S., KAREMORE, G., KARSSEMEIJER, N. and NIELSEN, M.. An Anatomically Oriented Breast Coordinate System for Mammogram Analysis.

In: IEEE Transactions on Medical Imaging. October, 2011 vol. 30, no. 10, p. 1841-1851.

[20] PERTUZ, S, JULIA, C. and PUIG, D.. A Novel Mammography Image Representation Framework with Application to Image Registration. In: International Conference on Pattern Recognition (22nd : 2014: Stockholm). p. 3292-3297

[21] ABDEL-NASSER, Mohamed, MORENO, Antonio and PUIG, Domenec. Temporal mammogram image registration using optimized curvilinear coordinates. In: Computer Methods and Programs in Biomedicine. April, 2016. vol. 127, p. 1-14.

[22] MUSTRA, Mario and GRGIC, Mislav. Robust automatic breast and pectoral muscle segmentation from scanned mammograms. In: Signal Processing. October, 2013. vol. 93, no. 10, p. 2817-2827.

[23] KELLER, Brad, NATHAN, Diane, WANG, Yan, ZHENG, Yuanjie, GEE, James, CONANT, Emily and KONTOS, Despina. Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation. In: Medical Physics. August, 2012. vol. 39, no. 8, p. 4903-4917.

[24] KANTA, Indra, NAG, Sanjay and BANDYOPADHYAY, Samir. Accurate Breast Contour Detection Algorithms in Digital Mammogram. In: International Journal of Computer Applications. July, 2011 vol. 25, no. 5, p. 1-13.

[25] BALLARD, Dana. Generalizing the Hough transform to detect arbitrary shapes. In: Pattern recognition. 1981. vol. 13, no. 2, p. 111-122.

[26] CANNY, John. A computational approach to edge detection. In: Pattern Analysis and Machine Intelligence, IEEE Transactions on. 1986. no. 6, p. 679.

[27] LIU, Li, WANG, Jian and WANG, Tianhui. Breast and Pectoral Muscle Contours Detection Based on Goodness of Fit Measure. In: International Conference on Bioinformatics and Biomedical Engineering (5th: 2011: Wuhan). p. 1-4.

[28] STEPHENS, Michael. Tests based on EDF statistics. In: Goodness-of-fit Techniques. 1986. vol. 68, p. 97-193.

[29] KARSSEMEIJER, Nico. Automated classification of parenchymal patterns in mammograms. In: Physics in medicine and biology. 1998. vol. 43, no. 2, p. 365.

[30] ZUIDERVELD, Karel. Contrast limited adaptive histogram equalization. In: Graphics gems IV. San diego: Academic Rress Professional, 1994. p. 474-485.

[31] IGLESIAS, Juan and KARSSEMEIJER, Nico. Robust initial detection of landmarks in film-screen mammograms using multiple FFDM atlases. In: Medical Imaging, IEEE Transactions on. 2009. vol. 28, no. 11, p. 1815-1824.

[32] CHANDRASEKHAR, Ramachandran and ATTIKIOUZEL, Yianni. A simple method for automatically locating the nipple on mammograms. In: Medical Imaging, IEEE Transactions on. 1997. vol. 16, no. 5, p. 483-494.

[33] METZ, Charles. ROC methodology in radiologic imaging. In: Investigative radiology. 1986. vol. 21, no. 9, p. 720-733.

[34] REFAEILZADEH, Payam, TANG, Lei and LIU, Huan. Cross-validation. In: Encyclopedia of database systems. US: Springer, 2009, p. 532-538.

[35] KITANOVSKI, I., JANKULOVSKI, B., DIMITROVSKI, I. and LOSKOVSKA, S. Comparison of feature extraction algorithms for mammography images. In:

International Congress on Image and Signal Processing (4th: 2011: Shanghai). p. 888-892.

[36] GALLOWAY, Mary. Texture analysis using gray level run lengths. In: Computer graphics and image processing. 1975. vol. 4, no. 2, p. 172-179.

[37] AL-SHAMLAN, H. and EL-ZAART, A.. Feature extraction values for breast cancer mammography image. In: International Conference on Bioinformatics and Biomedical Technology (2010: Chengdu). p. 335-340.

[38] HARALICK, Robert, SHANMUGAM, Karthikeyan and DINSTEIN, I. H.. Textural features for image classification. In: Systems, Man and Cybernetics, IEEE Transactions on. 1973. no. 6, p. 610-621.

[39] ZHANG, Gang and MA, Zong-Min. Texture feature extraction and description using Gabor wavelet in content-based medical image retrieval. In: Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition (2007: Beihing). p. 169-173 .

[40] MANJUNATH, Bangalore and MA, Wei-Ying. Texture features for browsing and retrieval of image data. In: Pattern Analysis and Machine Intelligence, IEEE Transactions on. 1996. vol. 18, no. 8, p. 837-842.

[41] DEVIJVER, Pierre and KITTLER, Josef. Pattern recognition: A statistical approach. London: Prentice-Hall, 1982.

[42] FRIEDMAN, Jerome, HASTIE, Trevor and TIBSHIRANI, Robert. The elements of statistical learning. Berlin: Springer series in statistics Springer, 2001.

[43] BISHOP, Christopher. Pattern Recognition and Machine Learning. Springer, 2006.

[44] HEATH, Michael, BOWYER, Kevin, KOPANS, Daniel, MOORE, Richard and KEGELMEYER, W. Philip. The digital database for screening mammography. In: Proceedings of the fifth international workshop on digital mammography. (2000). p. 212-218.

6. BIBLIOGRAPHY

ABDEL-NASSER, Mohamed, MORENO, Antonio and PUIG, Domenec. Temporal mammogram image registration using optimized curvilinear coordinates. In: Computer Methods and Programs in Biomedicine. April, 2016. vol. 127, p. 1-14.

AL-SHAMLAN, H. and EL-ZAART, A.. Feature extraction values for breast cancer mammography image. In: International Conference on Bioinformatics and Biomedical Technology (2010: Chengdu). p. 335-340.

BALLARD, Dana. Generalizing the Hough transform to detect arbitrary shapes. In: Pattern recognition. 1981. vol. 13, no. 2, p. 111-122.

BISHOP, Christopher. Pattern Recognition and Machine Learning. Springer, 2006.

BRANDT, S., KAREMORE, G., KARSSEMEIJER, N. and NIELSEN, M.. An Anatomically Oriented Breast Coordinate System for Mammogram Analysis. In: IEEE Transactions on Medical Imaging. October, 2011 vol. 30, no. 10, p. 1841-1851.

CANNY, John. A computational approach to edge detection. In: Pattern Analysis and Machine Intelligence, IEEE Transactions on. 1986. no. 6, p. 679.

DEVIJVER, Pierre and KITTLER, Josef. Pattern recognition: A statistical approach. London: Prentice-Hall, 1982.

FRIEDMAN, Jerome, HASTIE, Trevor and TIBSHIRANI, Robert. The elements of statistical learning. Berlin: Springer series in statistics Springer, 2001.

GALLOWAY, Mary. Texture analysis using gray level run lengths. In: Computer graphics and image processing. 1975. vol. 4, no. 2, p. 172-179.

HARALICK, Robert, SHANMUGAM, Karthikeyan and DINSTEIN, I. H.. Textural features for image classification. In: Systems, Man and Cybernetics, IEEE Transactions on. 1973. no. 6, p. 610-621.

HEATH, Michael, BOWYER, Kevin, KOPANS, Daniel, MOORE, Richard and KEGELMEYER, W. Philip. The digital database for screening mammography. In: Proceedings of the fifth international workshop on digital mammography. (2000). p. 212-218.

KARSSEMEIJER, Nico. Automated classification of parenchymal patterns in mammograms. In: Physics in medicine and biology. 1998. vol. 43, no. 2, p. 365.

KITANOVSKI, I., JANKULOVSKI, B., DIMITROVSKI, I. and LOSKOVSKA, S. Comparison of feature extraction algorithms for mammography images. In: International Congress on Image and Signal Processing (4th: 2011: Shanghai). p. 888-892.

LI, Hui, GIGER, Maryellen, HUO, Zhimin, OLOPADE, Olufunmilayo, LAN, Li, WEBER, Barbara and BONTA, Ioana. Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: Effect of ROI size and location. In: Medical Physics. March, 2004. vol. 31, no. 3, p. 549-555.

LIU, Li, WANG, Jian and WANG, Tianhui. Breast and Pectoral Muscle Contours Detection Based on Goodness of Fit Measure. In: International Conference on Bioinformatics and Biomedical Engineering (5th: 2011: Wuhan). p. 1-4.

MANJUNATH, Bangalore and MA, Wei-Ying. Texture features for browsing and retrieval of image data. In: Pattern Analysis and Machine Intelligence, IEEE Transactions on. 1996. vol. 18, no. 8, p. 837-842.

METZ, Charles. ROC methodology in radiologic imaging. In: Investigative radiology. 1986. vol. 21, no. 9, p. 720-733.

PERTUZ, S, JULIA, C. and PUIG, D.. A Novel Mammography Image Representation Framework with Application to Image Registration. In: International Conference on Pattern Recognition (22nd : 2014: Stockholm). p. 3292-3297

REFAEILZADEH, Payam, TANG, Lei and LIU, Huan. Cross-validation. In: Encyclopedia of database systems. US: Springer, 2009, p. 532-538.

STEPHENS, Michael. Tests based on EDF statistics. In: Goodness-of-fit Techniques. 1986. vol. 68, p. 97-193.

WEI, Jun, CHAN, Heang-Ping, WU, Yi-Ta, ZHOU, Chuan, HELVIE, Mark, TSODIKOV, Alexander, HADJIISKI, Lubomir and SAHINER, Berkman. Association of computerized mammographic parenchymal pattern measure with breast cancer risk: a pilot case-control study. In: Radiology. 2011. vol. 260, no. 1, p. 42-49.

ZUIDERVELD, Karel. Contrast limited adaptive histogram equalization. In: Graphics gems IV. San diego: Academic Rress Professional, 1994. p. 474-485.