UNIVERSIDAD **NACIONAL** DE COLOMBIA

Characterization of Architectural Distortion in Mammograms

Jorge Andres Alvarez Triana

Characterization of Architectural Distortion in mammograms

Jorge Andres Alvarez Triana

Tesis o trabajo de grado presentada(o) como requisito parcial para optar al título de:
Magister en Ingeniería Biomédica

Director(a):
Título (MD MSc PhD) Edgar Eduardo Romero Castro

Línea de Investigación:
Procesamiento Digital de ImÃ¡genes
Grupo de Investigación:
CIMALAB

Universidad Nacional de Colombia
Facultad de Medicina, Maestria en ingeniería Biomédica
Bogotá, Colombia
Año 2015

(Dedicatoria)

A Samus por llegar a mi vida accidentalmente y darme la paciencia y la fuerza para no desfallecer.

Agradecimientos

Quiero dar gracias principalmente a mi director de tesis Edgar Eduardo Romero Castro por su paciencia al guiarme y mostrarme la luz a su modo. A mi hija Samus por que no fue un impedimento para mi proyecto de vida si no que lo apoyo y me dio las fuerzas para seguir adelante. A Maria Yanira Gomez por estar en las buenas y en las malas a mi lado. A mi primita Charlens Alvarez por su apoyo y amistad durante este proceso. A mi amigo Jhonatan Tarquino por su apoyo, amistad y los almuerzos relajantes. A mis amigos del principio de la maestria Josue Ruano y Ricardo Trujillo, por hacer amena la estadia y compartir el peso de este trabajo. A Luisa Cardenas por su forma reconfortante de expresarse. Al grupo Cimalab por su colaboración y guía. A mis hermanos del alma Edward Salas, Jhonatan Quevedo y Francisco Paez que no dudaron de mi nunca. Y finalmente a mis padres Jorge Alvarez y Sandra Triana que desde la distancia me siguen apoyando y queriendo.

Resumen

En Mamografía suele evaluarse la presencia o posibilidad de cáncer de seno mediante signos como calcificaciones, masas, asimetría bilateral y distorsión arquitectural, este último de difícil identificación debido a la compleja distribución de los tejidos en la mama, siendo frecuentemente pasado por alto. La distorsión arquitectural se caracteriza por un patrón espiculado que define la malignidad de la lesión. Muchos de los trabajos realizados para caracterizar la distorsión arquitectural llevan la imagen entera a otro espacio en donde los patrones pueden ser discriminados. En esta tesis se presenta un método novedoso que utiliza información en el espacio de la imagen, en la cual primero se seleccionan manualmente las regiones de interés que son pre procesadas para mejorar sus detalles visuales. Después, la caracterización de la AD se realiza mediante la representación lineal de la saliencia en las regiones de interés (ROI) como un gráfo cuyos nodos corresponden a los píxeles a lo largo del borde de la ROI y cuyos arcos corresponden a las integrales de intensidad a lo largo de la ruta de conexión de cualquier par de nodos. Un conjunto de vectores propios obtenido de la matriz de adyacencia se utiliza para extraer coeficientes discriminantes que representen aquellos nodos con las líneas mas sobresalientes. Una reducción de dimensionalidad se logra adicionalmente mediante la selección del par de nodos con mayor contribución para cada uno de los vectores propios calculados. El conjunto de las líneas principales sobresalientes se ensambla como un vector de características que se introduce a una Maquína de soporte vectorial (SVM). Los resultados experimentales se realizan con dos bases de datos de referencia, el cojunto de datos MIAS y la base de datos DDSM, demostrando que el método propuesto tiene un buen desempeõ en términos de precisión El enfoque se evaluó con un conjunto de 246 ROI extraídas de la base DDSM (123 normales y 123 controles) y un conjunto de 38 ROI de la base de datos MINIMIAS (19 normales y 19 controles), respectivamente. Los resultados de la clasificación mostraron respectivamente, tanto para ambas bases de datos una precisión del 89,02 % y el 86,89 %, una sensibilidad del 85,37 % y el 84,21 %, y una tasa de especificidad del 92,68 % y 89,47 %.

Palabras clave: (Distorsión Arquitectural, cáncer de seno, carcinoma, mamografía, Centralidad de autovector).

Abstract

In mammography is usually evaluated the presence or possibility of breast cancer by using signs such as calcifications, masses, bilateral asymmetry and architectural distortion, being commonly overlooked. The architectural distortion is characterized by spiculated patterns that define the disease malignancy level, making it difficult to identify by such complex distribution of breast tissues. Most of existing methods characterize the architectural distortion by transforming the entire image to an alternative space, in which such complex patterns may be discriminated. In this thesis, we present a novel method that uses information in the image space, in which a region of interest is firstly selected to be preprocessed to enhance visual details. Afterward AD characterization is done by representing the linear saliency in mammography Regions of Interest (ROI) as a graph whose nodes correspond to those pixels along the ROI boundary and whose edges stand for the line intensity integrals along the path connecting any pair of nodes. A set of eigen-vectors from the adjacency matrix is then used to extract discriminant coefficients that represents those nodes with higher salient lines. A dimensionality reduction is further accomplished by selecting the pair of nodes with major contribution for each of the computed eigen-vectors. The set of main salient lines is then assembled as a feature vector that inputs a conventional Supported Vector Machine (SVM). Experimental results with two benchmark databases, the MIAS and DDSM database, demonstrate the proposed method performs well in terms of accuracy. The approach was evaluated with a set of 246 ROI extracted from the DDSM (123 normal and 123 AD) and a set of 38 ROI from the MINIMIAS collections (19 normal and 19 AD) respectively. The classification results showed respectively for both databases an accuracy rate of 89,02 % and 86,89 %, a sensitivity rate of 85.37 % and 84,21 %, and a specificity rate of 92.68 % and 89,47 %.

Keywords: Architectural Distortion, Breast cancer, Carcinoma, Mammography, Eigenvector centrality.

Content

List of Figures

List of Tables

1 Introduction

Breast cancer is a common public health problem, being the most frequently diagnosed cancer in women, representing a 16 % of all female cancers [48] and the second cause of mortality in female patients, exceeded only by lung cancer. According to the American Cancer Society [61], in United States for 2014 appeared approximately 232670 new cases of invasive breast cancer and 62570 new cases of carcinoma in situ (CIS). Additionally, about 40000 women die from breast cancer every year, which means that 1 of every 8 women develops the disease. According to the Liga Contra el Cancer [19], in Colombia for 2014 this type of cancer is the first cause of death with about 6500 new cases of invasive breast cancer and approximately 1600 women die from this disease every year. However, breast cancer is fully treatable when it is detected in early stages and mammography is still the more effective method for detecting early abnormalities, especially the screening mammography program in female patients without symptoms.

Some studies have reported that between 10 % and 25 % of breast cancers are not detected [32]. Such variability is presented because mammographic examination is particularly difficult to study, especially when radiological signs are hidden by the complex superposition of soft tissues. Recently, development of Computer Assisted Diagnosis Systems (CAD) in mammography has proved to be effective in finding calcifications and masses [9]. Nevertheless, it has been proved that CAD are not able to detect architectural distortions with adequate levels of accuracy (about 49 % of precision) [9, 42]. An architectural distortion (AD) lesion is any change of the spatial distribution of the breast connective tissue, in general with no visible mass. In such lesion, the typical retraction of tissues is variable and gradual, showing sometimes spiculations radiating from a point, sometimes accompanied of a sort of focal retraction and at some point, including also distortion at the parenchyma boundary. The set of radiological signs of an architectural distortion has been collected and formalized within the Breast Imaging Reporting and Database Systems (BI-RADS) standard [59]. Among the different lesions directly related with the presence of breast cancer, the architectural distortion has been reported as the most commonly missed abnormality in false-negative cases [36].

The methods introduced for detect AD seek connective tissue radiating from a particular point, characterizing orientations, intensities and width of the spicules. Several techniques have been used to detect areas with potential spicular pattern like linear detectors [69], steerable filters [58], oriented field analysis by phase portraits maps [56], others use spicule enhancement by image transformation (Gabor and Radon spaces) [57, 6, 12], or fractal fun-

ctions [28, 64]. Karssemeijer et al [34] detected stellate patterns, including spiculating masses and architectural distortion, using a multiscale method that classified the output of three-directional second order Gaussian derivative operators. Guo et al [28] applied five several methods that estimate the RoI Fractal Dimension. Ichikawa et al. [31] used a concentration index of linear structures obtained by the mean curvature of the image. Nemoto et al. [46] proposed the likelihood of spiculation and a modified point convergence index weighted by the likelihood to enhance architectural distortion. Biswas et al. [14] introduced a generati-ve model with bank of filters that learns "texton"patterns. The success of the previously mentioned techniques depends on the choice of parameters to construct linear detectors.

Furthermore, some methods combine spicular enhancement and orientation analysis. Ayres and Rangayyan [4] studied the phase portrait maps for capturing the radiating-out linear structure for AD identification, strategy that is based on Gabor filters that highlight linear patterns at certain orientations. Recent works have been reported promising results with this strategy [52], for instance, Banik et al. [12] characterize the AD using information from Gabor filters, phase portrait analysis, the fractal dimension, the entropy of the angular power spread and 14 Haralick's features. These approaches takes the detection problem of Architectural Distortion (AD) as a classification task. The methodology proposed by [4, 52] suffers a main limitation in the representation of radiating-out linear structure as a node-pattern by a phase portrait analysis because the spicules, in general, do not fully describe a well-defined converging pattern. In practice, the spicules mostly make an incomplete "star"shape with many missing parts, making the texture orientation field extremely noisy to deal with.

This thesis presents a new strategy to characterize the typical spiculated pattern of AD. The Main contribution of this approach is a novel method to compute a linear form of saliency based in a graph representation of the image spatial information. The graph is builded using the pixels of the ROI boundary as nodes and paths in which the intensity integrals (weights) are computed as vertices. The graph topology and vertices weights are summarized as an adjacency matrix, from which a set of eigenvectors are computed. Afterward, The adjacency matrix is then used to extract discriminant coefficients that represents locations with higher salient lines and a dimensionality reduction is accomplished by selecting the pair of nodes with major contribution for each of the computed eigenvectors.

1.1.   Breast cancer

Cancer cells growth abnormal instead of dying and divides forming new abnormal cells. This cells can also invade and growth in another tissues something that normal cell cannot do, often travels through blood or lymph system to other parts o f the body. If more lymph nodes are invaded more risk of metastasis in other tissues. Most breast masses are not malignant, but some may need sampled and analyzed under a microscope to confirm the presence of cancer cells. Most of their turn out to be fibrocystic change, this can cause breast swelling and pain.

Other breast tumors like as fibro adenomas or intraductal papillomas are abnormal but not cancerous and not invade other tissues. However still needs attention because women with these conditions have a higher risk of developing breast cancer.

1.1.1. Types of breast cancer

Ductal carcinoma in situ

Most common form of breast cancer, cancer cells are inside the milk ducts which transport the milk to nipple, but not have spread through the walls into the surrounding tissues.

Lobular carcinoma invasive

Its present in the lobules where milk is produced it can metastasize in other surrounding regions of surrounding tissue or the body. It's harder to detect than invasive ductal carcinoma.

Ductal invasive carcinoma

This is the most common invasive carcinoma. The cancer cells break through the walls of the ducts and growth in the fatty tissue of the breast, in this phase the cancer can metastasize to other regions of the body trough the lymph system or bloodstream.

Inflammatory breast cancer

An uncommon type of breast cancer, it consists that makes the skin of the breast red and feel warm. The cancer blocks the lymph vessels in the skin. The affected breast becomes larger or firmer, itchy or tender. This type of cancer is often mistaken with for an infection called mastitis. Because there is no lump it may not show in mammography.

Paget disease of the nipple

The cancer starts in the ducts of nipple and spreads to skin of the nipple and then to areola in a dark circle around it. Skin of the nipple and areola often appears crusted, scaly, and red, with areas of bleeding or oozing.

Phylloides tumor

It develops in the connective tissue of the mammary gland. Usually this tumor is benign, but too can be malign in a minor percent.

Angiosarcoma

This type of cancer starts in cells that line in blood vessels or lymph vessels rarely occurs in breast. When occurs in breast is extremely difficult the treatment with radiotherapy[24].

1.2. Mammography

Most efficient way to control breast cancer is the early detection. In women less than 50 years old, doctors should do a breast exam at least once in a year to detect thin masses that may be suspicious of cancer. From this age, mammography is the ideal exam for early detection of cancer.[32, 36, 48, 17, 19]; in younger womans this technique is not used because the breast are more dense (fibrous) therefore image is not clear.

According to use mammography may be classified as screening exam and diagnostic exam. The screening mammography has prevention purposes, suggested for women over 40 years and offers the exam once every year. Colombia's ministry of health includes this exam in the mandatory health program for all women. Diagnostic exam is done when a palpable or visual abnormality sign is detected.

In mammography the X-ray beams traverse the parenchymal mammary gland tissues and the projections are drawn on a radiographic plate in which differences are seeing between normal and abnormal tissues. This is a reably method that allows detect non palpable small tumors. This are visible because are composed of dense tissue, an thus absorb most part of the incident X-rays, often seeing as bright regions in mammograms.

In the figure **1-1** we can see an example of a normal case mammogram. Is a gray scale image in which can be identified epitelial, adipose, dense and muscle tissues and a dark region corresponding to the region outside the breast. In the procedure, two parallel plates are used to compress the breast increasing image quality by reducing the thickness of tissue that x-rays must penetrate and reducing the scattering effect that reduce the image quality see figure **1-2**. The exam takes two pictures of each breast, one of an angled side view (medio lateral oblique, MLO) and a head to foot view (cranio caudal,CC) as seen in the figure **1-1**.

The tissues of the mammary gland describe a pattern in its normal state; abnormalities are identified by the alteration of these patterns. In the image the structures of the mammary gland are described by the gray level change which identifies some alterations [29]. Women
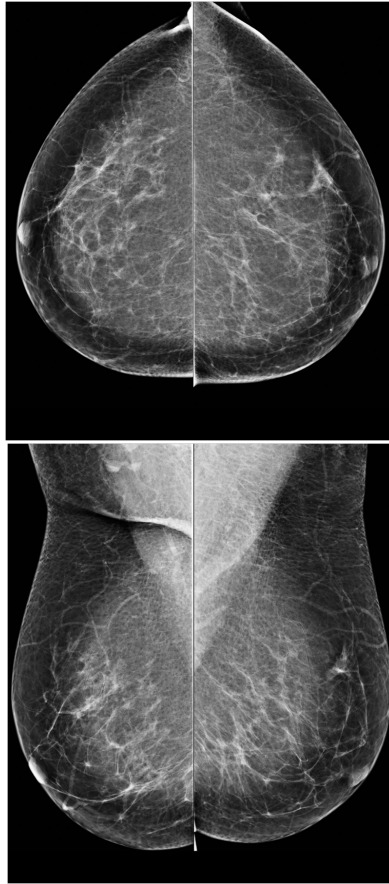
Figure 1-1: Example of a mammogram in a normal case: Left craneo-caudal view. Right medio Lateral oblique view
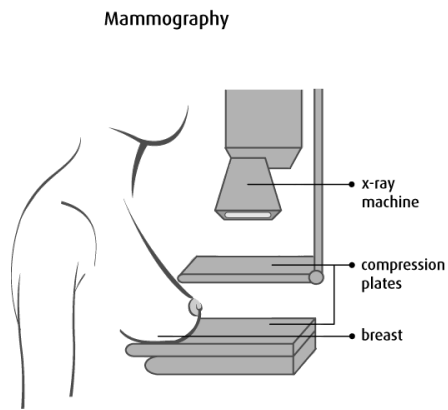
Figure 1-2: Procedure to take a x-ray in each breast. From www.cancer.ca consulted 15 june 2015.

think that the sensitivity of mammography examination is superior to 95 %; in this percent some women thinks it fully precision. The precision of mammography really surrounds the 68 % to 92 %. And not all types of breast cancer are detected in analog or digital technology. Actually the trained radiologists have several problems with not palpable lesions by the superposition of prenchymal tissues and its opinion differs about the image and must use other methods such as the percutaneous biopsy and the MR image [30] to confirm the diagnosis.

1.3. Diagnosis in mammography

To solve the different opinions between radiologists was created the standard BIRADS (Breast Imaging Report and Database System) in 1992 published by the ACR (American College of Radiologist). Is divided into three chapters mammography, ultrasound and MRI. Lesions are evaluated numerical by category. The categories are organized such help radiologist makes a treatment and following, this allows a uniform and consensus report that be understood by multiple medics or hospitals. Has seven different kinds according to their staging, diagnostic characteristics and disease manage [59].

A potential abnormality on a mammogram may be look like a mass, calcifications, bilateral asymmetry or an architectural distortion. Warning signs are visible and palpable in some cases, the presence or potential of developing breast cancer.

1.3.1. Calcifications

Are deposits of calcium that can be found in ductal ectacy process, vascular level, cutaneous, and fibroadenomas. The mammary gland is correlated to milk production when an abnormality occurs this deposit of calcium appears and may be sign of breast cancer [59, 2, 63].Usually there are two types of calcifications:

1. **Usually benign calcifications:** Usually gross or rounded, with smooth edges easy to detect than malign ones see **1-3**. These can be clasified into:

   - Skin Calcification: Little deposits of calcium that radio lucid pathognomonic center. The atypical ones can be viewed skins own by tangential projections.

   - Vascular: Parallel linear and tubular calcifications clearly associated with blood vessels.

   - Gross or Popcorn-Fibroadenoma: They are produced by the involution of an adenoma.
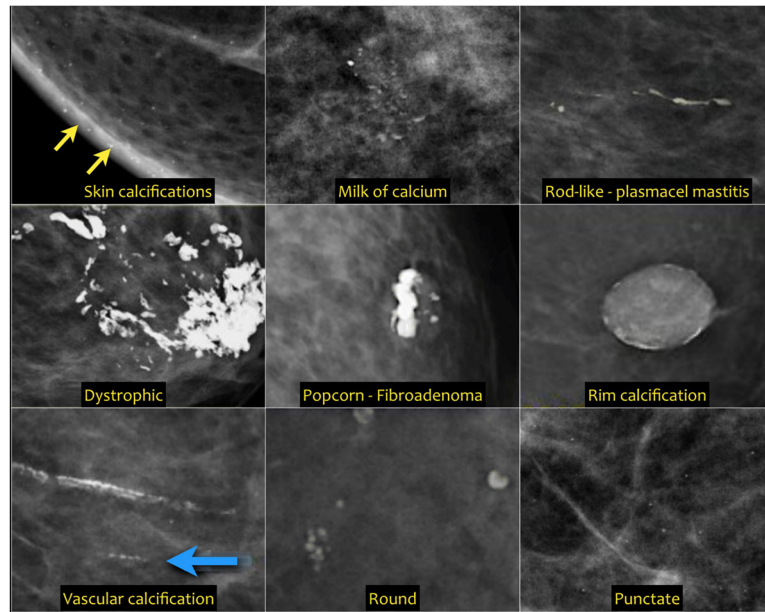
Figure 1-3: Usually bening calcifications on a mammogram taken from http://www.radiologyassistant.nl/en/p53b4082c92130/bi-rads-for-mammography-and-ultrasound-2013.html, consulted in 13 june 2015.

- Rod form: Linear calcications that radiolucid center, can be ramied, generally high density and smooth edges. Sign of mastitis, ductal ectasia and secretor disease.

- Rounded: Smooth surface with a diameter range of 1mm to 1cm, and it center can be radio lucid.

- Eggshell or Rim calcification: Slim calcium deposits on a sphere surface, are associated to fat necrosis and cysts.

- Suture: Are occasioned by accumulation of calcium in the material used in surgical suture.

- Dystrophic: Irregular form, usually radio lucid center, is caused by radiotherapy.

2. **Mostly likely malignant:** These are irregular, small and hard to detect**1-4**, can be classified into:

- Pleomorphic: are granular or heterogeneous, more speculated than the amorphous one, but not are typical in benign or malign cases.

- Finest: are slim or ramied, irregular, discontinues and less than 0.5 mm of width. Represents obstruction of light in a conduit compromised by cancer.
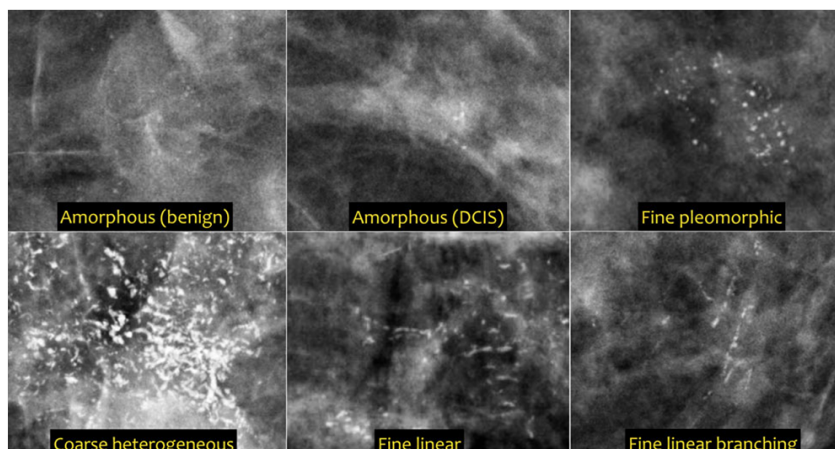
Figure 1-4: Mostly likely malignant calcifications taken from http://www.radiologyassistant.nl/en/p53b4082c92130/bi-rads-for-mammography-and-ultrasound-2013.html, consulted in 13 june 2015.

1.3.2. Masses

The masses are an important change that is observable in a mammography. Describe their size, shape, margins and calcifications associated, although there are cases in which the mass may occur without calcifications. Many signs can be seeing as masses like tumors, cysts, noncancerous fibroadenomas . Masses are space occupying lesions and are evident in two projections; if the mass is visualized only in one projection is called density [2, 59].

1. **Form:**
   Lumps are described by form as lobuled, oval,architectural distortion, irregular and circu- lar;see Fig**1-5**.

2. **Edge:**
   Is a visual characteristic that allows deline masses contours. However no all of it have their edges clearly limited or delined because of structure superposition, causing the partial denfition [2, 59]. They are classfied into circumscribed, microlobuled, darkest, espiculated and indistinct; see Fig**1-5**.

3. **Density:**
   Is defined as X-ray attenuation of the mass compared with the same volume of fibro glandular tissue. Usually the cancer is the same volume or greater than the fibro glandular tissue and don't contains fat. They are classied into high density, low density, same density and radiolucid.
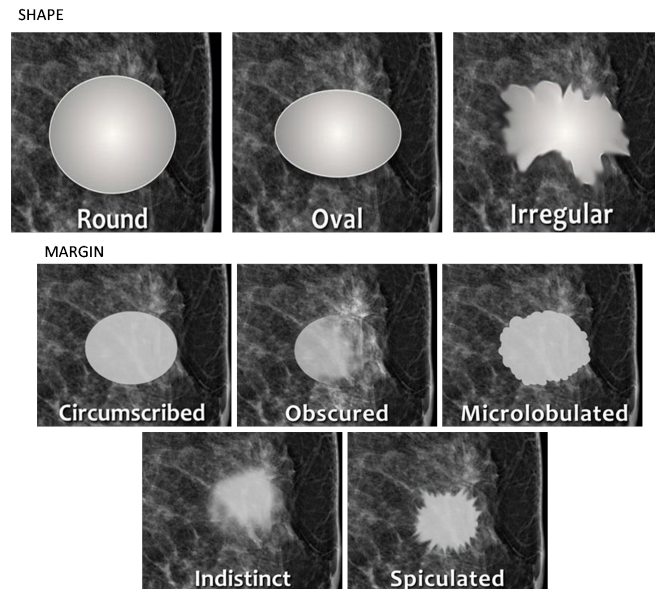
Figure 1-5: Classification of masses by shape and margins. Taken from http://www.radiologyassistant.nl/en/p53b4082c92130/bi-rads-for-mammography-and-ultrasound-2013.html ,consulted in 13 june 2015.

1.3.3. Bilateral Asymmetry

Is the presence of glandular tissue in a part of the breast which is not present in their contralateral part see **1-6**. It can be seen as an opacity in the mammogram in two projections, but not having the same signs as a mass. Requieres posterior exams to discard the presence of a mass or an architectural distortion[2, 59].

1.4. Architectural Distortion

The term Architectural Distortion (AD) is used, when the fibroglandular tissue has unusual changes, with no definite mass visible. This includes thin straight lines or spiculations radiating from a focal point, and focal retraction, distortion or straightening at the edges of the parenchyma see figure**1-7**. Is the third sign most common in mammograph and can be an associating finding[59]. It has been associated with breast malignancy in 1/2 to 2/3 of the cases in which it is present, in general with no visible and palpable mass. Spicular pattern is due to intrusion of cancer cells in the surrounding connective tissue, but also some previous biopsy or surgery scar can generate this pattern. Could appear in the initial stages of a breast tumor [38, 2] and may closely confuse the appearance with the normal breast tissue overlapped in the projection used in the mammography. Because breast cancer alters normal structure of parenchymal tissue even in its early stages the architectural distortion is considered an early detection sign.
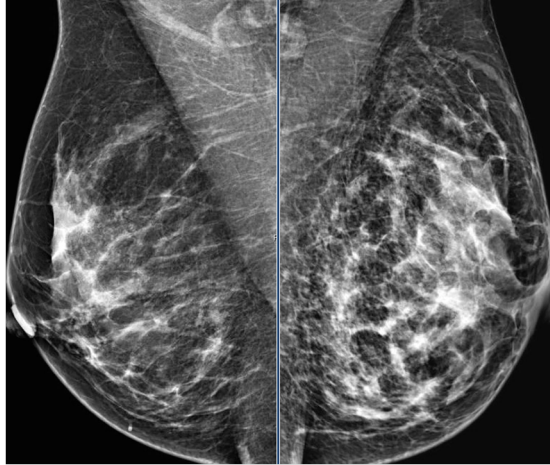
Figure 1-6: Bilateral asymmetry taken from http://www.radiologyassistant.nl/en/p53b4082c92130/bi-rads-for-mammography-and-ultrasound-2013.html, consulted in 13 june 2015.
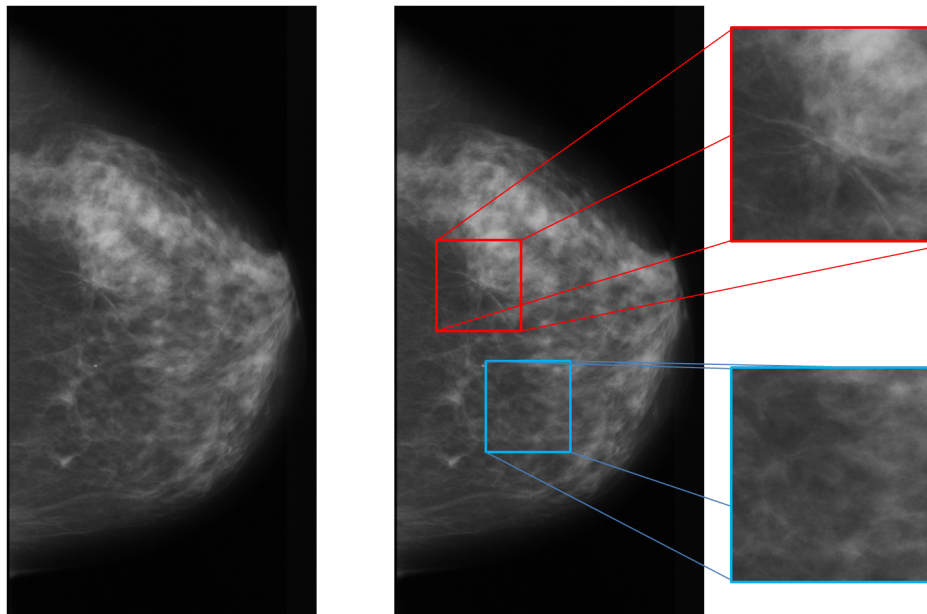
Figure 1-7: Architectural distortion: Left mammogram with AD. Right zoom of a AD lesion in the upper square, lower square show a normal region.

The importance of detecting AD lies in its standard BIRADS categorization. BIRADS assigning AD as class 4 with a positive predictive value of 50 % due to its subtle appearance and variability in presentation, suggesting taking a biopsy and MR image to confirm the presence of breast cancer. However, if is a false positive case it generates unnecessary costs and patient discomfort. AD Has been related with posterior malignancy in the breast [2, 59, 9, 38].

It is more difficult for physicians to detect architectural distortion than micro calcifications and masses, is estimated that this abnormality is the cause of between 12 % and 45 % of cancer cases omitted or misinterpreted [68]. Therefore is considered to be the sign associated with the development of cancer more difficult to detect.

1.5.   CAD Systems

CADS are a number of automatic computational tools to assist radiologist in detection and / or evaluation of medical images. These tools not pretend replace radiologist in their work, otherwise support the diagnosis by techniques and algorithms who allows improve visual assessment of the image and the diagnostic characteristics.

In 1955 Lee Lusted propose an automatic method to evaluate and diagnose radiography images laying the foundations to CAD systems; in 1967, Fred Winsberg describes a mammography CAD system that is able to differentiate between benign and malignant lesions [67]. Actually CAD algorithms are capable of revealing breast lesions like malign micro calcifications and masses. The commercial systems sensitivity to detect micro calcifications is reported as 99 % [66] and the reported sensitivity to masses is in a range between [75 % - 89 %][66, 50, 37]

Baker et al, focus in the study of detection of architectural distortion for commercial systems, versus radiologist diagnosis for 80 mammography images of which 45 cases were of AD, the results sensitivity for radiologist was 60 % and for commercial CAD systems were 49 % [10].The accuracy of a commercial CAD system acting as a second reader is lower than the double indepent reading method used by radiologist[8].

CAD types

There are two types of CAD systems, based on differential diagnostic and based in automatic detection. The automatic detection guides the radiologist on a suspicious region of interest marking an delimiting it. The algorithms used for this type of detection are based on textures descriptors or boundary extraction to obtain patterns for classification[40, 43, 3]. Differential systems offers an interactive feedback to radiologist [44, 45], use detection based in groups of malign or benign signs, and be conducted by visual similarity respect to a group of images previously diagnostic [22, 47], however this type of CAD system is still on test and research [22, 47, 23].

1.6.   Strategies to Characterize Architectural Distortion Lesion on Mammograms (State of the art)

From main characteristic of AD the focal radiating pattern, several methods search connective tissue radiating from a particular focus, a stellate pattern with different orientation, intensity and width of the spicules. All of this works focus on somehow enhance fibrous tissue linear structures and highlight information related to orientation, scale, or strength of lines, for the characterization process of the lesion.

Several techniques have been used to detect potential sites of spicular pattern, among others linear detectors [69], steerable filters [58], oriented field analysis by phase portraits maps [56], spicules enhancement by image transformation (Gabor and Radon spaces) [57, 6, 12] or fractal functions [28, 64]. Karssemeijer et al [34] detected stellate patterns, including spiculating masses and architectural distortion, using a multiscale method that classified the output of three-directional second order Gaussian derivative operators and reported a sensitivity of about 90 % with one False-Positive per image (FP/Image). Zwiggelaar et al. [70] evaluated the performance of several method for detecting and classifying linear structures in the image spatial domain, where the type of linear structure depends on the criterion parameters. Sampat et al. [58, 57] proposed a linear filter of the image Radon transform which is filtered by radial spiculation filters that detect spiculated masses and architectural distortion, reporting a sensitivity of 80 % and 14 FP/image for AD and a sensitivity of 91 % and 12 FP/image for spiculated masses. Guo et al [28] applied five several methods that estimate the RoI Fractal Dimension. Afterward, a SVM differentiated the masses and AD from the normal parenchymal tissue, obtaining an area under curve ROC (AUC) of 0.875. Ichikawa et al. [31] evaluates the concentration index of linear structures obtained by the mean curvature of the image, reporting a sensitivity of 68 % and 3.4 FP/image. Nemoto et al. [46] proposed the likelihood of spiculation and a modified point convergence index weighted by the likelihood to enhance AD characteristics, establishing a sensitivity of 80.0 % and 0.80 FP/image. Biswas et al. [14] introduced a generative model with bank of filters that learns "texton"patterns and used 19 images with AD and 21 normal tissue images (a subset of images from MIAS dataset). Evaluation was performed under a leave-one-out scheme, obtaining 3.6 false positives per image and 81.3 % sensitivity with a ROI radius of 5 mm (the minimum radius of an AD region in MIAS).

On the other hand, Ayres and Rangayyan [4] studied the phase portrait maps for capturing information from linear structure orientations for AD identification, strategy that is based on Gabor filters. Recent works have reported promising results with this strategy [52], for instance, Banik et al. [12] detected the AD using the Gabor filters and phase portrait analysis. For each detected ROI, the fractal dimension, the entropy of the angular power spread and 14 Haralick's features were computed. Results showed an area under ROC curves (AUC) of 0.76 with the Bayesian classifier, 0.75 with Fisher linear discriminant analysis, and 0.78 with a single-layer feed-forward neural network. These aforementioned approaches face the

detection problem of A as a classification task of RoIs of the entire mammogram. The major drawback is the choice of parameters to construct linear detectors, Gabor filters, steerable filters, linear filters or features in both spatial and transform image domain.

In general methods based in gabor filters and phase portrait Analysis suffers a main limitation in their parameter choice to measuring spicules and the intended detection of a node-pattern, representing radiating-out linear structure, often fails because the spicules do not fully describe a well-defined converging pattern. In practice, the spicules mostly make an incomplete "star"shape with many missing parts, making the texture orientation field extremely noisy to deal with.

2 Using Node Saliency

Characterization of Architectural Distortion on Mammograms Using a Linear Energy Detector

Presented on the "SIPAIM 2013"

Architectural distortion is a breast cancer sign, characterized by spiculated patterns that define the disease malignancy level. In this paper, the radial spiculae of a typical architectural distortion were characterized by a new strategy. Firstly, previously selected Regions of Interest are divided into a set of parallel and disjoint bands (4 pixels × the ROI length), from which intensity integrals (coefficients) are calculated. This partition is rotated every two degrees, searching in the phase plane the characteristic radial spiculation. Then, these coefficients are used to construct a fully connected graph whose edges correspond to the integral values or coefficients and the nodes to x and y image positions. A centrality measure like the first eigenvector is used to extract a set of discriminant coefficients that represent the locations with higher linear energy. Finally, the approach is trained using a set of 24 Regions of Interest obtained from the MIAS database, namely, 12 Architectural Distortions and 12 controls. The first eigenvector is then used as input to a conventional Support Vector Machine classifier whose optimal parameters were obtained by a leave-one-out cross validation. The whole method was assessed in a set of 12 RoIs with different distribution of breast tissues (normal and abnormal), and the classification results were compared against a ground truth, already provided by the data base, showing a precision rate of 0.583 %, a sensitivity rate of 0.833 % and a specificity rate of 0.333 %.

2.1. Introduction

Breast cancer is the most frequently diagnosed cancer in women, representing a 16 % of all female cancers [48]. However, breast cancer is fully treatable when detected in early stages and mammography is yet the more effective method for detecting early abnormalities, even though some studies have reported that between 10 % and 25 % of breast cancers are not detected [32]. Such variability is present because mammographic examination is particularly difficult, especially when radiological signs are hidden by the complex superposition of soft tissues. Ultimately, development of Computer Assisted Diagnosis Systems (CAD) in mammography has proven to be effective in finding calcifications and masses [9]. Nevertheless, they have been found to fail in detecting architectural distortion with adequate levels of accuracy

[9], typically of 49 %. An architectural distortion is any change of the spatial distribution of the breast connective tissue, in general with no visible mass. The typical retraction of tissues is variable and gradual, sometimes showing spiculations radiating from a point, sometimes accompanied of a sort of focal retraction and at some point, including also distortion at the parenchyma boundary. The set of radiological signs of an architectural distortion has been collected and formalized within the Breast Imaging Reporting and Database Systems (BI-RADS) standard [1]. Among the different lesions directly related with the presence of breast cancer, the architectural distortion (AD) has been reported as the most commonly missed abnormality in false-negative cases [36].

In despite of CAD systems may act as second reader in diagnostic tasks, they have shown relatively low performance for AD detection[9]. Parenchymal distortions have been approached using geometrical information of breast tissues, namely, orientation fields, spiculation level detectors, steerable filters and multiresolution image analysis. Several studies reported characteristic patterns that allow to find changes related to distortion. Ayres and Rangayan focus their work on capturing radiating out-line patterns [5, 53, 7], using gabor filters and phase portrait and detecting potential sites of AD with a 80 % - 90 % of sensitivity. Rangayan et al. used coherence with respect to the breast tissue [55] and a measure of angular dispersion for detection of potential sites of AD [54], reporting a sensitivity of 90 % and a rate of 6.3 false positives per patient. Zwiggelaar et al. used factor analysis to discover dominant directions of intensity, reporting a sensitivity of 80 %, with a false positive rate of 0.014 [71]. Karssemeijer and Brake [35] proposed a method that is based on statistical analysis of a pixel orientation map and studied three different features for finding stellate lesions, the Laplacian of a Gaussian, correlation with a mass-based model, and gradient orientation [16], reporting a sensitivity of 90 % with 1 false positive per image.

This paper presents a new approach, devised to characterize the radial spiculae of a typical architectural distortion. For doing so, the image is split into thin parallel and disjoint rectangular areas, typically of four pixels × the length of the ROI, from which intensity integrals are calculated. This partition is rotated every two degrees, searching in the phase plane the characteristic radial spiculation. The obtained coefficients are herein named Linear Energy Detectors (LED) and are used to construct a graph that represent such linear relationship. A node in this graph is associated to the image coordinates, from which a thick line (rectangular area) is either leaving or arriving. The edges stand for the connection strength between the image coordinates, i.e., the LED. A centrality measure of that graph is the first eigenvector and used to feed a binary classifier, a conventional SVM that uses a radial basis function as kernel. The particular eigenvector is then mapped back to the image and zones with changes are highlighted and then interpreted.

2.2. Methods

An overview of the proposed framework is shown in Fig. **2-1**. Firstly, a set of regions of interest (RoIs) with both distorted and normal tissues are manually selected and preprocessed to enhance the visual details. Afterward, a set of coefficients are computed by simply capturing the particular energy within a band in the image, These coefficients become the edge weights of a fully connected graph and used to find the most probable path in the graph or the orientations with larger levels of energy. Afterward, these coefficients and their relationship are used as feature vector. Finally, regions are classified with a binary SVM strategy using a radial basis function.
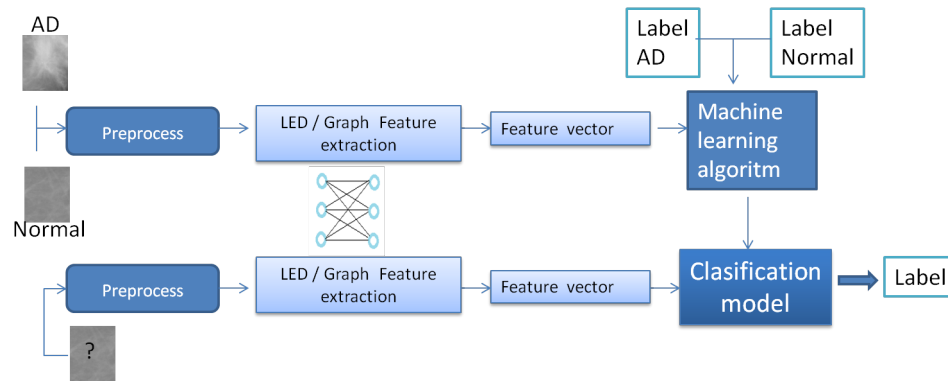
Figure 2-1: Framework) First we obtain the LED coefficients, with which the graph is completed by using them as the weights of the relative image positions. The first eigenvector, containing those regions with a higher level of linear information, is used to train a classifier.

2.2.1. ROI Pre-procesing

Visual details were enhanced by mapping the signal dynamical range to the maximum and minimum gray level values of the interval [0; 255], using an histogram adaptive equalization. These images were smoothed out by a median filter to remove remaining noise see fig. **2-2**.

2.2.2. Feature Extraction

According to BI-RADS, the typical retraction of tissues is variable and gradual, showing spiculations radiating from a point, sometimes including also distortion at the parenchyma boundary. In terms of image processing, a main goal is to find features related to textural patterns. Such features can be extracted as described in the following sections.
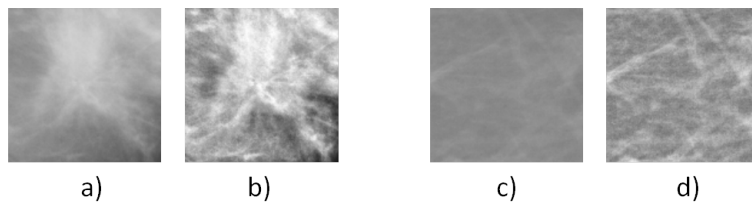
Figure 2-2: Preprocessing stage. a) and c) RoI of Architectural Distortion and Normal tissue respectively. b)and d) Adaptive equalization and smoothing.

2.2.3. Linear Energy Detection

A main aim of the radiological examination is to determine those areas with some degree of retraction, for which a bright area is the center of a lesion characterized by certain number of spiculae associated to that center. The main purpose of a Linear Detection of such a lesion is to systematically scan the phase plane and to find those directions with more energy. A drastic dimensionality reduction is achieved by associating such analysis structure with a graph that stores the intensity linearly distributed.

A LED is simply a direction in which the energy is calculated as the integral within a particular domain. The domain was herein selected as a thin rectangle or a thick line since the signal through a conventional line integral is noisy. The image coordinates are then joined by lines with a thickness of 4 pixels, aiming to detect those spiculae. the image coordinates used to project the lines are computed from a small unitary square $[0,1]^2$.

Starting in the upper left corner and marking in counter clockwise, a set of the image coordinates serve to generate the particular geometrical domains in which the integrals are calculated see fig. **2-3**. The collection of integral surfaces are the LED coefficients.
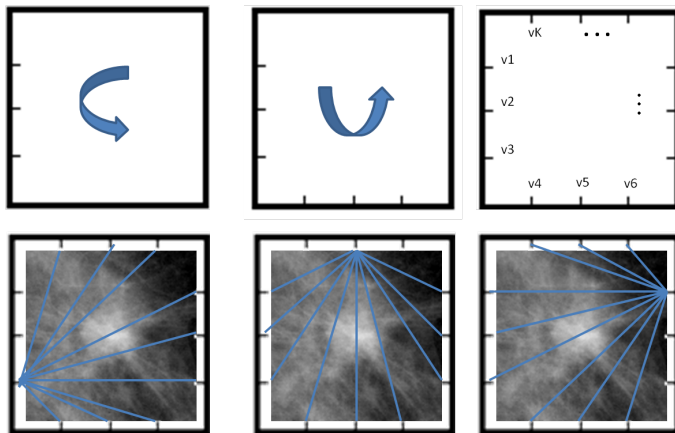
Figure 2-3: Vertice indexing and line connection in unitary square $[0,1]^2$, the blue arrow indicates the way in which the vertices are indexed and the blue lines shows some possible connections between vertices.

2.2.4.   The Graph-based Representation

Graphs and induced Markov chains are used to model saliency and attentional fixations in natural images. Typically, to construct a graph onto an image, the image pixels are represented as vertices, and the relations between them are encoded within the graph edges, in some cases with a specific weight associated to each connection. Global relations are represented by interconnecting all vertices, forming a fully-connected graph, while local relations can be analysed by connecting each vertex only with its immediate neighbors. Graph relations are summarized as an adjacency matrix A, where each row and column corresponds to a vertex in the graph, and the value stored at each position i, j corresponds to the edge connection weight between the pair of vertices i and j. Algebraic operations of the adjacency matrix A allows to pick out important information about the graph structure. For example, measures such as centrality, proposed by [25], allow to estimate the global importance of each vertex, if importance is defined in terms of how çentral"the vertex is within the graph structure.

Once the image graph is constructed, an induced Markov chain allows to analyze the graph flow by associating the graph vertices to states, and the edge weights to transition probabilities. The adjacency matrix A of the graph corresponds now to the stochastic matrix M of the Markov chain, and under some appropriate assumptions (irreducibility of M due to a strongly connected graph) it has been proved that the Markov chain tends to a unique stationary probability distribution. This stationary (equilibrium) distribution represents the amount of time (or frequency of visits) that a random walker would spend at each state if he were allowed to walk for an infinitely long time. This can be interpreted as a proportional measure of the vertex importance, with respect to all other vertices in the chain. This measure has been shown to be equivalent to the eigenvector centrality described by [25], where the vertex importance is calculated in a self-referential way: a vertex is important because it is adjacent to other important vertices. [27] proposed a robust approach where global and local image properties are computed from Markov random walks on a complete graph and a sparse k-regular graph, respectively. From this information, a small number of the most salient (object) and background nodes are set and labeled as object or background.

In practice, the fully connected graph in fig **2-4** stores in its adjacency matrix A the strength of the edge connections. The first eigenvector of the adjacency matrix represents the steady state of the associated induced Markov chain and includes the nodes with larger importance in terms of the graph connections.

2.3.   Experimental Results

The strategy was evaluated on a total of 36RoIs, that included distortion and normal cases, previously annotated by a group of radiologists from the *Mammographic Image Analysis Society (Mini-MIAS)* [62]. This dataset was split into training (24) and test (12), (6 normal tissues - 6 architectural distortion) subsets.

Accuracy	0.5833
Sensitivity	0.833
Specificity	0.333
Error rate	0.4167

Table 2-1: Results of svm performance for AD and Normal classes

Classification performance was evaluated by a SVM strategy using the radial basis function as the kernel. The whole strategy, set to 128 dimensions by the graph method, reported an accuracy of 60 %,. The obtained confusion matrix reads as: $accuracy = \frac{TP+TN}{TP+FP+FN+TN}$, where TP is the number of True Positives, TN true negatives, FP false positives, FN false negatives, respectively. Results are shown in **2-1**.

2.4. Conclusions and Future Works

In this paper we have proposed a new approach to characterize this AD using a simple linear descriptor that aims to capture the spiculated pattern. This method is simple and allows to reconstruct a large variety of spiculated patterns. The method demonstrates a sensitivity of 83 % better than reported for commercial CAD systems.
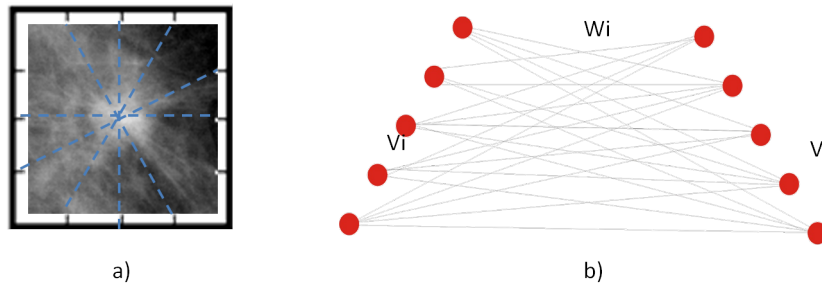
Figure 2-4: a) Linear Energy Detectors cross linearly the image connecting the different opposite sides of the image. b) Fully connected graph of the LED drawn on the image at the left panel.

3 Linear Saliency

Characterizing Architectural Distortion in mammograms by Linear Saliency

Architectural distortion (AD) is a common cause of false-negatives in mammograms. This lesion usually consists of a central retraction of the connective tissue and a spiculated pattern radiating from it. This pattern is difficult to detect due the complex superposition of breast tissue. This paper presents a novel AD characterization by representing the linear saliency in mammography Regions of Interest (RoI) as a graph composed of nodes corresponding to locations along the RoI boundary and edges with a weight proportional to the line intensity integrals along the path connecting any pair of nodes. A set of eigenvectors from the adjacency matrix is then used to extract discriminant coefficients that represent those nodes with higher salient lines. A dimensionality reduction is further accomplished by selecting the pair of nodes with major contribution for each of the computed eigen-vectors. The set of main salient lines is then assembled as a feature vector that inputs a conventional Support Vector Machine (SVM). Experimental results with two benchmark databases, the MIAS and DDSM databases, demonstrate that the proposed linear saliency domain method (LSD) performs well in terms of accuracy. The approach was evaluated with a set of 246 RoI extracted from the DDSM (123 control and 123 AD) and a set of 38 RoI from the MIAS collections (19 control and 19 AD) respectively. The classification results showed respectively for both databases an accuracy rate of 89 % and 87 %, a sensitivity rate of 85 % and 95 %, and a specificity rate of 93 % and 84 %. Likewise, the A_z was 0.93 for both databases.

3.1. Introduction

Breast cancer is the most frequently diagnosed form of cancer and the second largest cause of death for women worldwide [60]. When detected in its early stages, there are therapeutical alternatives to manage and fully cure this form of neoplasia. Currently mammograpy, *i.e.* the visual inspection by an expert of X-ray images of the "compressed" breast, is considered the most cost-effective method for breast cancer detection in the early stages [65, 18]. From a mammography an expert radiologist is able to detect abnormalities such as calcifications,

bilateral asymmetry, masses and Architectural Distortion (AD) [18].

Despite the widespread use and proven effectiveness of mammographies, screening mammography programs showed a high intra and inter-observer variability in the early 90s of last century with undetected cancer rates of between 10 % and 30 % [13]. More recent studies have reported rates of between 10 % and 25 % [32]. This high rate is partly due to the fact that identifying radiologic signs for mammographic interpretation remains a complex visual task because of the blurriness originating from the overlaying of a complex superposition of breast tissues.

3.1.1.   Architectural distortion

An AD lesion is any change of the spatial distribution of the breast connective tissue, in general with no visible mass. In such lesion, the typical retraction of tissues is variable and gradual, commonly showing radial spiculations. Focal retraction or distortion at the boundary of the parenchyma without a visible or palpable mass may also be present. The set of radiological signs of an AD has been collected and formalized within the Breast Imaging Reporting and Database Systems (BIRADS) standard [59]. Figure **3-1** illustrates some differences in both AD and normal tissue, manually selected by a radiologist. AD has been reported as the most commonly missed abnormality in false-negative cases [36].

3.1.2.   Assisted AD detection

The development of Computer Assisted Diagnosis Systems (CAD) in mammography has proved to be effective in finding calcifications and masses [**?**]. Nevertheless, CAD systems have not been as effective in the detection of AD with adequate levels of accuracy (about 49 % of precision) [**?**, 42]. Several methods have been introduced to detect AD connective tissue radiating from a center which have the appearance of radiolucent lines or stellate patterns. In these methods linear structures are enhanced to highlight information related to orientation, scale, or strength of lines, facilitating the characterization process. Several techniques have detected areas with potential stellate patterns, among others linear detectors [69], steerable filters [58], oriented field analysis by phase portraits maps [56], spicules enhancement by image transformation (Gabor and Radon spaces) [57, 6, 12] or fractal functions [28, 64].

Karssemeijer *et al.* [34] detected stellate patterns, including spiculating masses and AD using a multi scale method that classified the output of three-directional second order Gaussian operators and reported a sensitivity of about 90 % with one False-Positive (FP) per image (FP/image).

Zwiggelaar *et al.* [70] evaluated four distinct methods for detecting and classifying linear structures in synthetic images. The reported Area Under the Curve (AUC) using line operators, Orientated Bins, Gaussian Derivatives and Ridge Detectors were of 0,94, 0,91, 0,9 and 0,82, respectively. The type of linear structure detector and the parameter selection depends
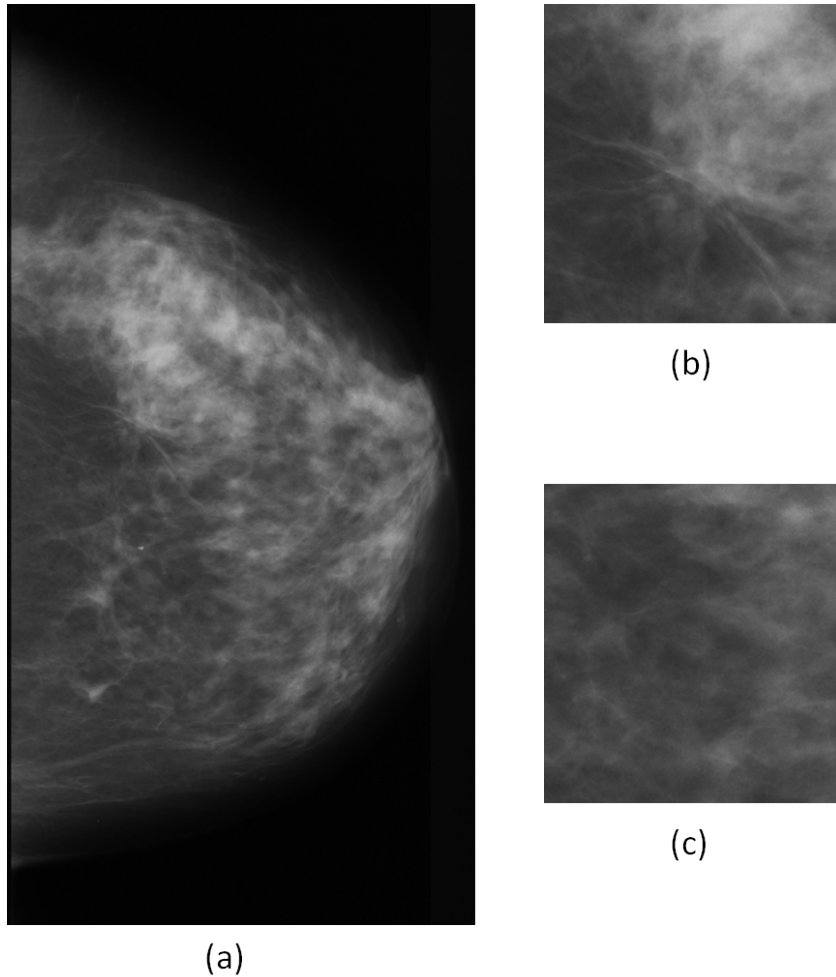
Figure 3-1: a). Mammogram showing AD. b) A 128×128 AD RoI showing spiculations radiating from a central point. c) . Normal 128×128 RoI tissue portion. RoIs (b) and (c) were both exracted from the mammogram (a).

on prior knowledge of the lesion being imaged.

Sampat *et al.* [58, 57] proposed a transformed image and radial spiculation filters in the Radon space to detect spiculated masses and AD. The reported sensitivity was of 80 % and 14 FP/image for AD and the reported sensitivity was of 91 % and 12 FP/image for spiculated masses.

Guo *et al.* [28] applied five methods to estimate a Region of interest (RoI) of a Fractal Dimension that was used to train a support vector machine classifier (SVM). The authors reported an AUC of 0.875 when differentiating masses and AD from the normal parenchyma.

Ichikawa *et al.* [31] used a concentration index of linear structures obtained by the mean curvature of the image, resulting in a sensitivity of 68 % and 3.4 FP/image.

Nemoto *et al.* [46] proposed the likelihood of spiculation and a modified point convergence index weighted by the likelihood to enhance AD, establishing a sensitivity of 80.0 % and 0.80

FP/image.

Biswas *et al.* [14] introduced a generative model with a bank of filters that learned "texton" patterns using a subset of images from the Mammographic Image Analysis Society (MIAS) dataset [62] (19 AD and 21 control images). Evaluation was performed under a leave-one-out scheme, obtaining 3.6 FP/image and a sensitivity of 81.3 % with a RoI radius of 5 mm (the minimum radius of an AD region in MIAS).

Kamra *et al.* [33] use a combination of Spatial Gray Level Co-occurrence Matrix (SGLCM), fractal and Fourier power spectrum based features to characterize the AD lesion, quantified in four directions $\theta = 0°, 45°, 90°, 135°$. This strategy was evaluated on a subset of RoIs from MIAS and the Digital Database for Screening Mammography (DDSM) [41] datasets with different size of RoI and a subset of fixed ones, the reported accuracy, sensitivity, specificity of the fixed RoI versions of DDSM was 92.9 %, 93.3 %, 92 %, for MIAS was 97.2 %, 85.7 %, 95.3 % respectively.

Matsubra *et al.* [39] applied directional and background filters to analyze the linear structure in eight directions. The detection sensitivity and the FP/image reported was 81 % and 2:5 - 4:2 respectively. The dataset used in this study contains 174 AD and 580 normal diagnosed control cases.

Several studies have used the phase portrait maps to capture the outward radiating linear structure, a strategy based on Gabor filters that highlight linear patterns at certain orientations [4, 52, 37, 11] with significant results.

Banik *et al.* [11] detected the AD using the Gabor filters and phase portrait analysis. For each detected RoI, the fractal dimension, the entropy of the angular power spread and 14 Haralick's features were computed. Results showed an AUC of 0.76 with the Bayesian classifier, 0.75 with Fisher linear discriminant analysis, and 0.78 with a single-layer feed-forward neural network.

These approaches face the AD detection problem as a classification task of RoIs of the entire mammogram. The major drawback is then the need of choosing a large number of parameters to construct linear detectors, Gabor filters, steerable filters, linear filters or features in both spatial and transform image domain to measure spicules.

Additionally, the intended detection of a node-pattern, representing radiating-out linear structures often fails because the spicules, in general, do not fully describe a well-defined converging pattern but instead spicules mostly make an incomplete "star" shape with missing parts, making the texture orientation highly noise sensitive.

3.1.3.   Contribution

Unlike previous approaches, in this paper we explore the search of the linear saliency of a stellate pattern in the spatial domain. The proposed Linear Saliency Domain method (LSD) represents a RoI as an (initially) fully connected graph with nodes corresponding to the RoI boundary and edge weights calculated using line integrals of the intensities along

the path connecting any pair of nodes. By doing this, the edges corresponding to the most salient lines are assigned a larger weight in the graph. A centrality measure based on the graph's adjacency matrix's eigenvectors is used to identify the most relevant nodes. A further dimensionality reduction is accomplished by selecting only the most salient edge originating from relevant nodes. The set of salient lines is then assembled as a feature vector used to train an SVM classifier. Experimental results in two benchmark databases, the DDSM and MIAS databases, demonstrate that LSD outperforms the baseline techniques [4, 52] in accuracy and speed with a precision rate of 89 % and 87 %, a sensitivity rate of 85 % and 95 %, a specificity rate of 93 % and 84 % respectively.

The rest of this article is organized as follows: Section 3.2 explains in detail the methodology and steps of the LSD. In Section 3.3 the datasets used to test the method are presented. In section 3.4, results of the validation experimentation, are shown. Section 3.5 presents the discussion, conclusions and possible future work directions are explored.

3.2.   Method

The LSD method consists of five consecutive steps: Image Enhancement, Graph Construction, Central Node Detection, Detection of Salient Edges and Classification. The pipeline of the whole process is presented in Figure **3-2** and each step is explained in detail in the following sub-sections.

3.2.1.   Image enhancement

The RoIs of the images selected as input for LSD are smoothed using a non-linear median filter [15] aiming to remove the image noise. Spicular details are enhanced by mapping each image's dynamical range to the maximum and minimum gray level values of the interval ([0, 255]), followed by a contrast limited adaptive histogram equalization approach (CLAHE) [51].

3.2.2.   Graph Construction

The perimeter of a squared RoI is divided into 32 segments per border for a total of $N = 4 \times 32 = 128$ segments. The segments' width, 4 pixels in the MIAS database and 32 pixels in the DDSM database, were chosen to approach the typical width of a single spicula, $i.e.$ approximately 200 μm [6]. A graph $G(V, E)$ is built with the N segments corresponding to nodes V. The weights of the edges E are computed as the integral of the RoI intensities along the lines connecting every pair of nodes, $i.e.$ the value of the integral between node V_i and node V_j corresponds to edge $E_{ij} = E_{ji}$. Notice that when nodes lie on the same border, $e.g.$ two nodes on the top border, the thickness of the line and the integral result are both zero, making G is a partially connected graph.
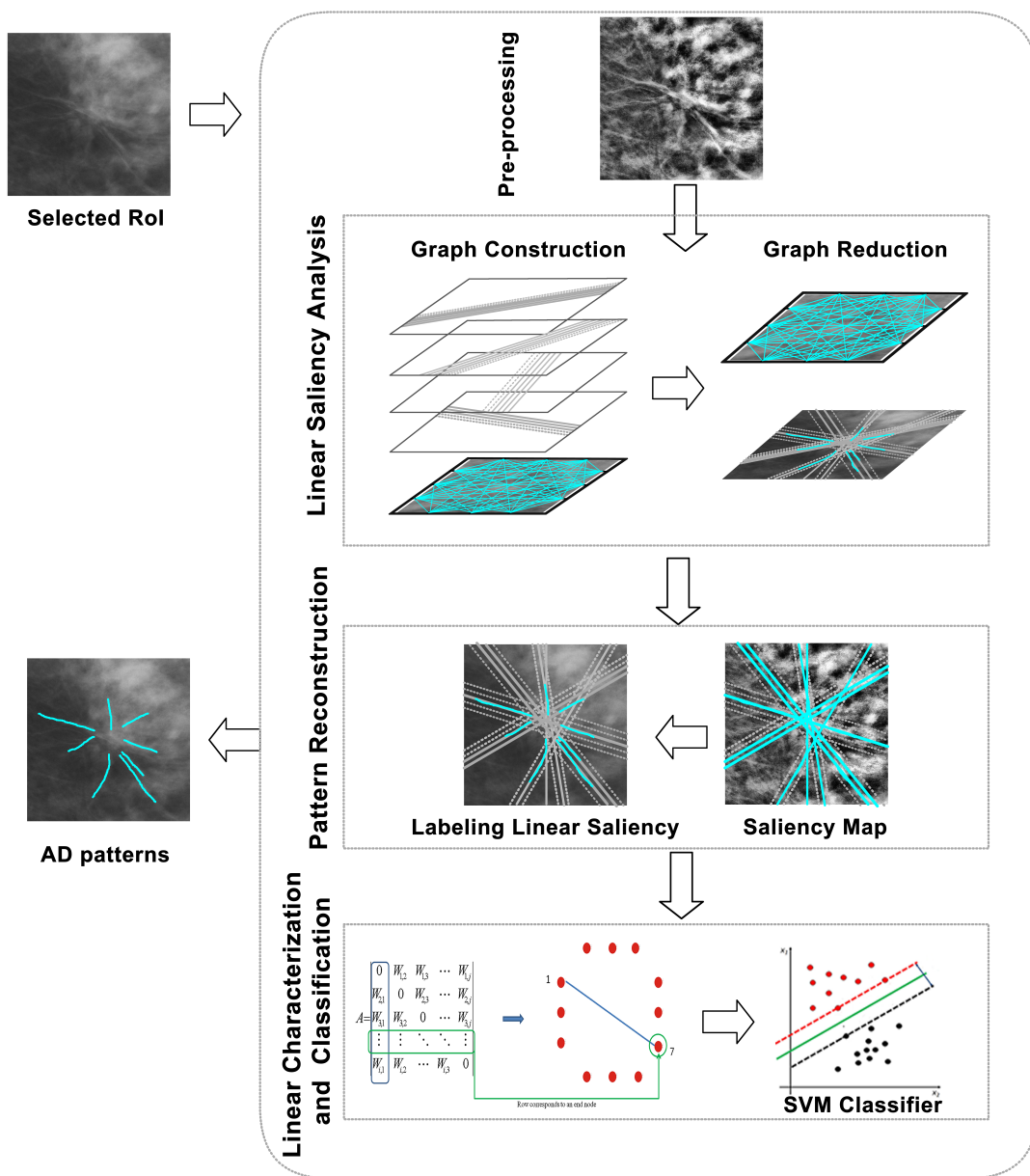
Figure 3-2: Pipeline of LSD.

Graph relations are summarized as an adjacency matrix A, where each position $[i, j]$ corresponds to the weight $w_{ij} = w_{ji}$ of the edge between two vertices (V_i, V_j). An illustrated example of the construction of the graph may be seen in Figure **3-3** and its adjacency matrix is shown in Equation 3-1.
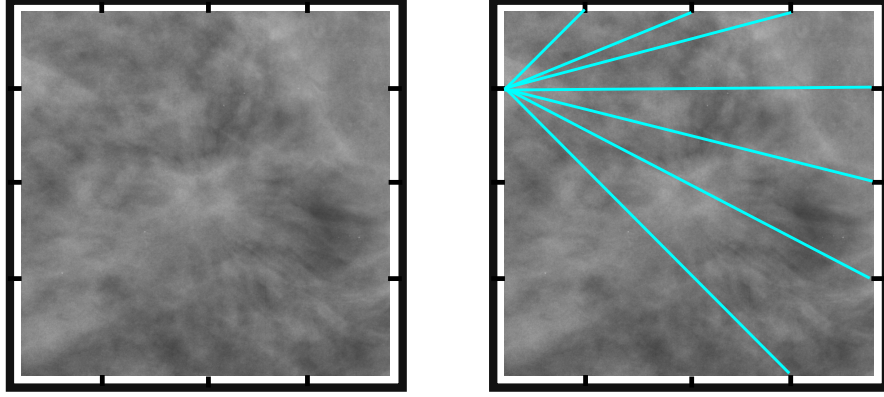
Figure 3-3: The RoI boundary is divided into N segments as illustrated in the left panel. The center of each segment corresponds to a node V_i. The weight w of the edges connecting any pair of nodes corresponds to the value of the line integrals.

$$
A = \begin{pmatrix}
0 & w_{1,2} & w_{1,3} & \cdots & w_{1,j} \\
w_{2,1} & 0 & w_{2,3} & \cdots & w_{2,j} \\
w_{3,1} & w_{3,2} & 0 & \cdots & w_{3,j} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
w_{i,1} & w_{i,2} & w_{i,3} & \vdots & 0
\end{pmatrix} \tag{3-1}
$$

3.2.3. Central Node Detection

A Markov chain may be associated to the graph $G(V, E)$ by normalizing the weights of the edges. Nodes can then be interpreted as the states of the Markovian process, while the weights are equivalent to transition probabilities. Under the appropriate assumptions, such as irreducibility due to strong connectivity, this chain tends to a stationary probability distribution that represents the frequency of node visits that a random walker would make if allowed to walk forever. In the present problem, this probability distribution would naturally be concentrated at nodes that have higher linear intensities since transitions into subgraphs containing them is more likely.

This type of representation has been shown to be equivalent to the eigenvector centrality analysis described in [26], where the node importance is calculated in a self-referential way, *i.e.* a node's importance is larger if it is adjacent to other important nodes. Different artificial vision strategies have represented the image relationships by a graph structure. Gopalakrishnan *et al.* [27], for example, have proposed that Markov random walks may represent the global or local image properties by performing the centrality analysis on either the complete or sparse k-regular graphs. From this information, the most salient foreground and background objects may be set and labeled.

The most important nodes may be extracted from the eigenvectors sorted using the eigenvalues of A. The first eigenvector represents the steady state of the associated induced Markov chain and includes the nodes with the most important connections. Subsequently, the nth eigenvector corresponds to the most important nodes except those in the mth eigenvector, for all $m < n$.

In each eigenvector the largest element $\overrightarrow{V}_k$ corresponds to the position of the most relevant node v_m :

$$v_m = max(\overrightarrow{V}_k) \quad \forall k \in [1, \ldots, N]. \tag{3-2}$$

.

3.2.4.   Detection of salient edges

Once the central nodes have been detected, it is necessary to select the most salient edges that radiate from them. This is done by locating the highest weights in the original A adjacency matrix. As stated before, higher values correspond to the more radiolucent lines. In particular, each relevant node stands for a column of the adjacency matrix A, as illustrated in Figure **3-4**.
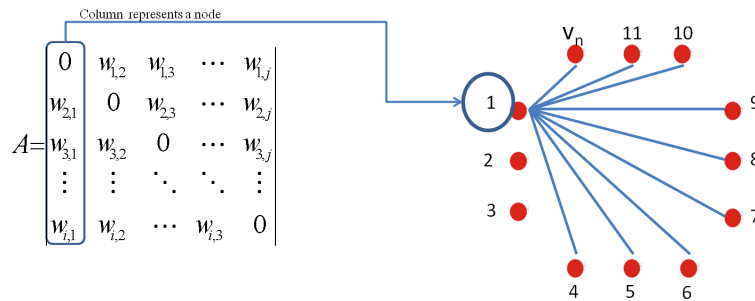
Figure 3-4: In the matrix, a column represents a node in the graph and its connections

The highest value of the column corresponds to the most important connection of that node, as shown in the Figure **3-5**. This connection is then the most salient pattern for the selected central node.

3.2.5.   Classification

The binary AD classification problem is approached by an SVM classifier [21]. This strategy is applied to datasets with complex separation boundaries between classes and is based on maximizing the separating margin between the two classes in an alternative space in which data relationships are linear. In this case two classes are defined: normal breast tissue and ADs [28].
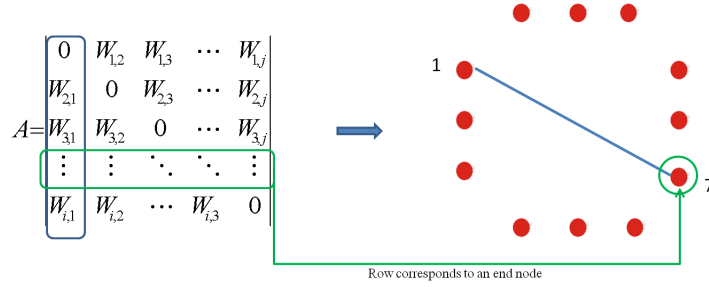
Figure 3-5: each column correspond to a node and its connections

The proposed method trains the classifier model using radial basis functions (RBF) as a kernel. Positive values correspond to AD and negative values to control cases. The optimal hyperplane is found by solving the discriminating function:

$$f(x) = sign\left(\sum_{i=1}^{n} y_i a_j \phi^T(x_i).\phi(x_j)\right) \tag{3-3}$$

where sets of labels y_i module the Lagrange multipliers a_j and the kernel function $K(x_i, x_j)$ is a similarity measure defined by computing the inner-product between the feature vectors as

$$\phi^T(x_i).\phi(x_j) = e^{-(\frac{1}{2\sigma^2}\|x-x_i\|^2)}. \tag{3-4}$$

The optimal kernel parameters are the default parameters of the libSVM [20] ($\gamma = \frac{1}{5128} = 0,00019$ and $\sigma = 1$) .

3.3.   Datasets

Two public mammography databases were used to test the method: DDSM and MIAS. Both collections contain digitized versions of mammography screen-films and an independent set of healthy patients, 695 control cases in DDSM and 207 in MIAS.

Two sets of RoIs consisting of 123 controls and 123 AD were selected from the DDSM [41] database and 19 control and 19 AD from the MIAS [62] database. The control cases were randomly selected from the set of controls provided by each database. Once a control case was chosen, an expert radiologist selected the normal fibroglandular tissue RoI coordinates. In contrast, the AD RoI was selected using the size and location annotations stored in the database. A RoI corresponds to the smallest square region containing the complete lesion, 128×128 for the MIAS database and 1024×1024 for the DDSM collection. Additionally, synthetic images were developed to test the proposed LSD method.

3.3.1. Synthetic Data

A set of synthetic images was constructed to assess and fine tune the method's parameters. These digital phantoms aim to emulate actual AD and therefore must keep the radiated pattern of linear structures around a focus. The method implemented was presented by Parr *et al.* [49] and consists of finding a representative element of a series of artificially constructed radiating patterns. Briefly, the method splits a square image of 81×81 pixels into tiles of 9×9 that serve to focus radiated patterns starting from the center of each tile. For a particular centered pattern, the same 9×9 representation grid is used but this time each tile stores the average angle of every line crossing the tile. From these images, the obtained 81 radiating patterns storing the mean angle for each tile, are then vectorized and used to feed a Principal Component Analysis strategy (PCA). The main twelve eigenvectors of the covariance matrix, that explain a 90 % of the variability, are set as the statistical representative elements. Afterwards, the weights of the six vectors that approximate 80 % of the variance are retrieved by simply subtracting the mean to the observation and projecting the result onto the matrix of twelve eigenvectors.

The weights of the principal directions b_i are found then by centering the data x_i and projecting onto the principal direction matrix P, as stated in equation 3-5 and illustrated in Figure **3-6**

$$b_i = P^T(x_i - \mu) \tag{3-5}$$

(a)   (b)   (c)
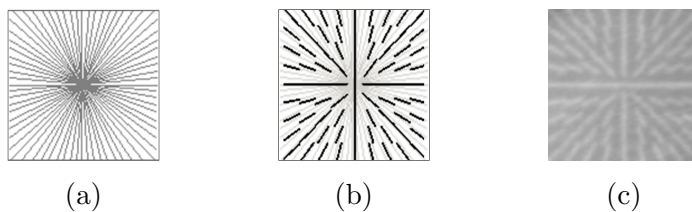
Figure 3-6: **3-6**a. Radiating pattern. **3-6**b. Angular representation of a radiating pattern. **3-6**c. Focal radiating phantom obtained.

Different linear patterns can be obtained from this representation by simply de-centering the data by one or two standard deviations, as formulated in Equation 3-6 and illustrated in Figure **3-7**. The final phantom consists of a pattern superimposed to an actual mammogram region diagnosed as normal. The dataset is composed of 162 phantoms, namely 81 focal that are de-centered as previously explained by one standard deviation to yield the 81 non-focal patterns.

$$b_i = P^T(x_i - \mu + s) \tag{3-6}$$
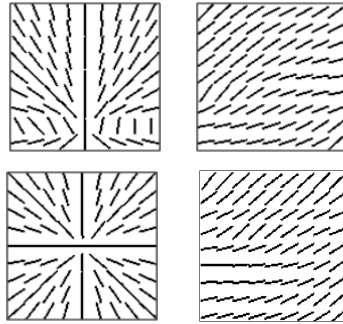
Figure 3-7: Effects of de-centering the focal line pattern

3.3.2. The DDSM database

The DDSM database[1] has been widely used as a mammography evaluation benchmark [41]. This open access database consists of digitized images of mammographic films with the corresponding technical and clinical information. The database is sorted by malignancy level and contains a total of 2620 cases, each including four images obtained from Cranio-Caudal (CC) and Medio Lateral Oblique (MLO) projections as well as the specific BIRADS descriptions that follow the agreement in $4th$ edition. The DDSM collection provided a total of 123 AD cases that were selected along with the expert annotations. The centroid of the region selected by the expert was used as the center of a 1024×1024 pixels RoI surrounding the lesion. Likewise, 123 control RoIs with the same dimensions were selected as previously explained for a total set of 246 mammograms. The selected mammograms had been digitized either with a Lumisys laser film scanner at 50 μm or a Howtek scanner at 43.5 μm pixel size and at a resolution of $2^{12} = 4096$ gray level tones.

3.3.3. The MIAS database

This collection of cases is a public dataset widely used as a benchmark, published by MIAS[2]. The database provides mammograms from a screened population at a spatial resolution of 50 μm and a grey-level depth of 8 bits. The RoIs were also set to a size of 128×128 pixels as previously explained, always excluding the nipple and pectoral muscle. Images are sorted and annotated by the parenchymal tissue type (fatty, fibroglandular or dense) and the BIRADS lesion description. This dataset contains a total of 322 cases from Medio Lateral Oblique (MLO) view, but only 19 cases are documented for AD. A set of 19 normal cases were chosen following the protocol aforementioned.

---

[1]http://marathon.csee.usf.edu/Mammography/Database.html
[2]http://peipa.essex.ac.uk/info/mias.html

3.3.4. Experimental Setup

The performance of the LSD method was tested using series of normal and AD RoI either synthetically constructed or extracted from the two public datasets described above. Classification followed an SVM strategy with an RBF kernel using the LIBSVM library implementation [20]. The strategy herein implemented consists of 128 nodes that result in 32 salient lines from the graph introduced in previous section. The different metrics used to evaluate the results were: $accuracy = \frac{TP+TN}{TP+FP+FN+TN}$, $sensitivity = \frac{TP}{TP+FP}$ and $specificity = \frac{TN}{TN+FN}$, being TP the number of True Positives, TN the true negatives, FP the false positives and FN the false negatives, respectively. Performance is also evaluated using the Region Operating Curve (ROC) generated by setting thresholds as the membership.

3.4. Results

Comparison with baseline methods was performed by implementing two different strategies: a classical approach by Ayres and Rangayyan [4] mapping to the Gabor space and then phase portraying, and the strategy proposed by Banik *et al.* [11], who analyze both the node maps of the phase portraits and Haralick texture descriptors from the Gabor magnitude. These two representations are basically a projection to the Gabor space, but while Ayres *et al.* use the phase space orientation map, from which different features are extracted, Banik *et al.* include also some characteristics of the phase space magnitude. It is worth mentioning that we implemented these methods using exclusively frequency and texture features which were reported by the authors to give good results.

3.4.1. Experiments with Synthetic data

As introduced in section 3.3.1, synthetic data emulates different AD focal and non-focal oriented patterns, corresponding to the pathological and control cases, superimposed to a regular distribution of normal tissue, as illustrated in Figure **3-8**. This background was chosen by an expert from a case which was randomly selected from the MIAS set of control cases. The obtained is composed of 162 RoIs, 81 focal and 81 no focal oriented patterns.
0
This evaluation tested not only the classification using pure phantoms, but also the quality of the result obtained when the phantom was contaminated with different levels of Gaussian noise. The added noise emulates the type of noise resulting from the contribution of many independent sources and associated to tomography and radiographic images [**?**]. The Gaussian noise was set to $\mu = 0$ and $\sigma = 0{,}05$, a level for which the peak signal to noise ratio (PSNR) between the original image and the degraded one was of 19.6 dB. A total of 162 phantoms (81 focal and 81 regular patterns) were synthetically generated, and each of these subsets was randomly sorted out. These two sets were then split to a, roughly, 70 to
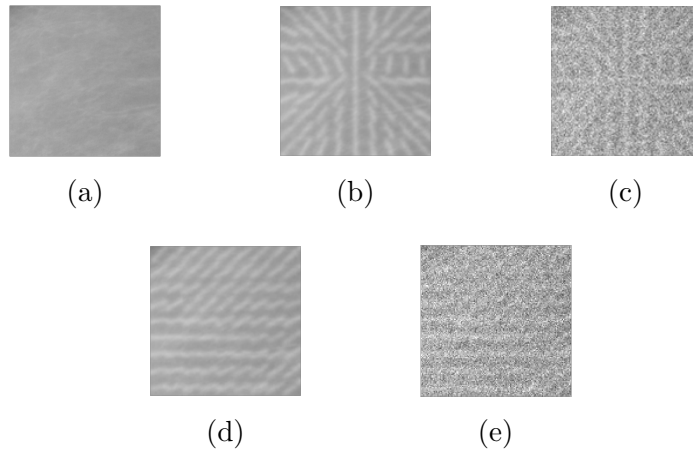
Figure 3-8: Different phantom patterns. (a) background from a control MIAS Mammogram. (b) focal AD phantom. (c) focal AD phantom with Gaussian noise. (d) Non-focal regular AD phantom. (e) Non-focal regular AD phantom with Gaussian noise. Gaussian noise was set to $\mu = 0$ and $\sigma = 0{,}05$.

30 training to testing cross-validation ratio with 57 of each group used for training and the remaining 24 of each group used for evaluation. A first test was made on the pure phantoms and a second one over the original patterns corrupted with the Gaussian noise. This classification task was carried out using a conventional SVM classification with an RBF kernel. Comparative results obtained for the two baseline methods and LSD, using both the original and corrupted patterns, are shown in Table **3-1**.

	Acc		**Sens**		**Spec**		**AUC**	
	P	P_G	P	P_G	P	P_G	P	P_G
LSD	0.98	0.92	1.0	0.87	0.96	0.96	1.0	0.96
Ayres *et al.*	0.54	0.48	0.66	0.33	0.42	0.62	0.52	0.52
Banik *et al.*	1.0	0.5	1.0	0.75	1.0	0.25	1.0	0.5

Table 3-1: Comparative results of the classification with synthetic images for each evaluated method. Column P shows the results for the initial synthetic image and P_G for the Gaussian corrupted phantom test sets.

The baseline method proposed by Banik *et al.*, for the original test P shows a perfect performance for the set of generated patterns, demonstrating its effectiveness for detection of focal patterns. Likewise, LSD showed an accuracy of 98 %, a sensitivity of 100 % and a specificity of 96 %, evidencing a comparable performance with these simulated lesions. These two methods outperform the implementation of the Ayres *et al.* proposal [4]. In the corrupted phantoms tests LSD outperformed the two baseline methods showing comparable figures to what was observed with the uncorrupted patterns. The baseline methods completely lose the

discriminative power shown with the uncorrupted dataset.

Figure **3-9** presents the obtained results as ROC curves that depict the classification in terms of *sensitivity* vs $1 - specificty$. Panel (a) in Figure **3-9** shows a perfect performance for the proposed method and the baseline proposed by Banik, yielding an AUC of about 1,0 in both case. In contrast the classical Ayres method shows a very low performance, with an AUC of 0,52. Interestingly, when a Gaussian noise is added to the synthetic phantoms, the proposed approach results remarkably more resistant than the baseline approaches, as illustrated in panel (b) of Figure **3-9**. The AUC of the tested methods were LSD= 0,96, Ayres= 0,52 and Banik= 0,5.

(a)   (b)
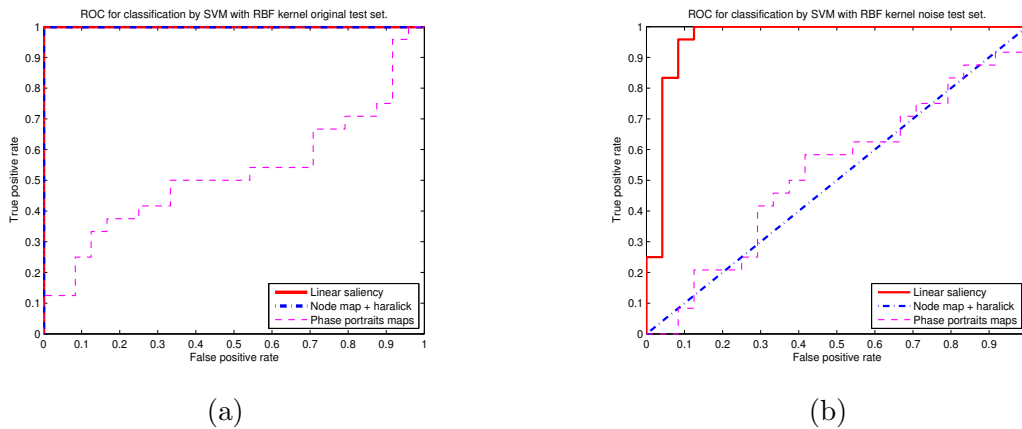
Figure 3-9: ROC curves after the evaluation of each method for both experimental groups. Results of the original synthetic images are shown in panel (a) and the results of the corrupted ones are shown in panel (b).

3.4.2.   Experiments with DDSM and MIAS Databases

Given the small number of AD cases available in the MIAS database (19), cross-validation tests were carried out using a leave-one-out scheme.

The baseline method proposed by Banik *et al.* showed an accuracy of 89 %, a sensitivity of 84 % and a specificity of 89 %, evidencing a good performance for the MIAS set. The proposed method showed for the same database an accuracy of 87 %, a sensitivity of 95 % and a specificity of 84 %, showing a performance comparable to Banik's method. These two methods strongly outperforms the Ayres *et al.* strategy. Results are shown in Table **3-2**.

The number of AD cases in the DDSM dataset allowed a k-fold cross validation scheme with k=10 folds to be used. In this test, LSD showed an accuracy of 89 %, a sensitivity of 85 % and a specificity of 93 %: a much better performance than the one observed with Banik *et al.* that yielded an accuracy of 76 %, a sensitivity of 75 % and a specificity of 76 %. The classical method proposed by Ayres *et al.* showed an even lower performance.

	Acc		Sens		Spec	
	DDSM	*MIAS*	*DDSM*	*MIAS*	*DDSM*	*MIAS*
LSD	$0,89 \pm 0,07$	$0,87 \pm 0,09$	$0,85 \pm 0,13$	$0,95 \pm 0,12$	$0,93 \pm 0,13$	$0,84 \pm 0,1$
Ayres *et al.*	$0,59 \pm 0,13$	$0,65 \pm 0,13$	$0,33 \pm 0,15$	$0,68 \pm 0,13$	$0,68 \pm 0,14$	$0,63 \pm 0,15$
Banik *et al.*	$0,76 \pm 0,11$	$0,89 \pm 0,12$	$0,75 \pm 0,13$	$0,84 \pm 0,13$	$0,76 \pm 0,15$	$0,89 \pm 0,14$

Table 3-2: Comparative classification results with the DDSM and MIAS datasets for each evaluated method. Accuracy, Sensitivity and Specificity are reported.

Figure **3-10** presents the results of SVM classification as a ROC curves. Panel (*a*) shows a high performance on the DDSM for LSD, with an AUC of 0,93 while the baseline proposed by Banik *et al.* yield 0,83 and the classical method proposed by Ayres *et al.* show an AUC of 0,55.
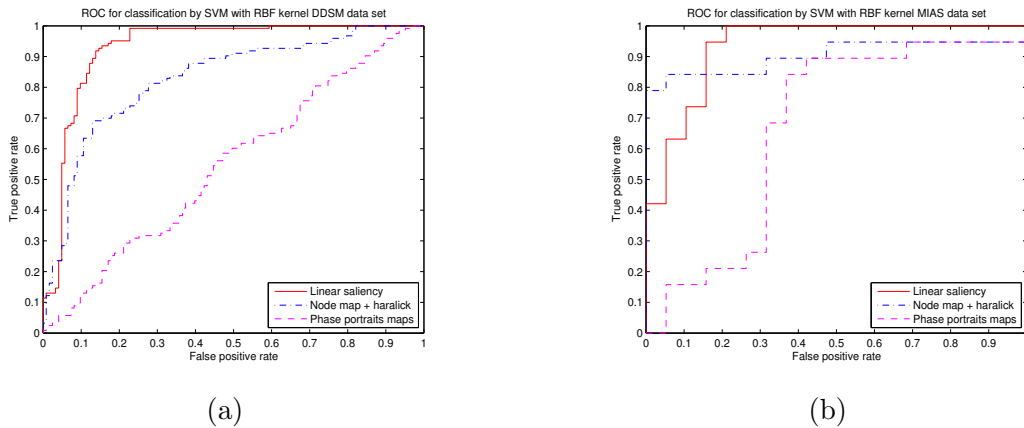
Figure 3-10: (a) ROC curves for the DDSM and MIAS databases obtained from a classical 10 fold cross validation for the set of RoIs. (b) The ROC curve corresponds to a leave one out scheme.

3.5. Discussion and Conclusions

This work has introduced a new form of characterizing AD on mammograms by representing RoI tissue distribution as a graph that captures the salient directions of linear structures within a particular RoI. The AD saliency is highlighted using a graph structure that stores the intensity linear information in the weight of the edges connecting the borders of a RoI. This external border undergoes a dyadic partition that is driven by the maximal width of an AD spicula, in the present investigation four pixels for the MIAS database and 32 for the DDSM. The spectral decomposition of the graph into its main eigenvectors reveals the RoI linear clusters, the set of ordered eigenvectors of the graph adjacency matrix. Each

eigenvector stands for a different degree of relevance or saliency and is ordered by importance. The descriptor is constructed by concatenating the lines associated to the most important nodes, i.e., the largest value of each eigenvector.This descriptor was assessed and compared against two baselines, outperforming two state-of-the-art approaches.

The classical seminal work proposed by Ayres *et al.* based on Gabor transform and phase portrait maps sets a precedent by using directional and distribution information from curvilinear structures present in the breast parenchyma. The reported evaluation for this method was done with all disponible cases of AD from MIAS dataset also used here. A subsequent and recent improvement to this method was proposed by [12] *et al.*, this makes use of node map information and combined with texture features as Fractal Dimension or Haralick results obtained are good. The presented work was compared against this method using Haralick texture features beacause is a feature being commonly and widely used.

The use of spectral decomposition techniques of the graph into its main eigenvectors allows the identification of linear clusters, represented in this case by the set of ordered eigenvectors of the graph adjacency matrix.

Results Using synthetic images, which can be viewed as an "idealized case" of an AD, showed almost perfect classification results. This indicates that, even though it does not explicitly try to extract linear or spiculated patterns, LSD does highlight the image information permitting a consistent detection under ideal conditions. The main advantage of the method, however, is its robustness to noise. This is evidenced by its continued good performance under noisy conditions, opposed to the notable drop in performance of the method proposed by Banik. The method proposed by Ayres shows a very low discrminative power which is only slightly better than chance when noise is added (AUC=0.52).

When evaluated with open databases, such as DDSM, the robustness of LSD is evidenced by showing similar results to the synthetic datasets and outperforming both the Banik and the Ayres methods. Although the features proposed by Banik are more rapidly learned, they are unable to cope with the larger biological variability introduced in the dataset.

As expected, the results are less concluding in case of the MIAS database, probably due to a much smaller population (barely nineteen cases). Still, for the MIAS basically the Banik strategy and LSD are equivalent, with very similar AUC (LSD=0,93 and Banik= 0,9).

Future work includes evaluation of the whole mammography as a screening approach. Interestingly, the graph structure might be used to compare different lesions in topological terms and to establish a similarity metric between lesions.

However, the method is dependent on parameters as the cross sectional width of a typical spicule, a source of error that might be mitigated by a previous tuning of this parameter in a particular set of mammograms, i.e., an approximation by varying this width for instance between 3 and 10 pixels. This limitation is clearly common to any method that attempts to search the lesion spiculae [69, 57, 6, 12, 34, 46, 33, 39].

4 Conclusions and Recommendations

This thesis has explored the problem of AD characterization developing a new approach. Main contribution of this work is the introduction of a linear form to compute saliency of an stellate pattern, searching in the spatial domain, sorting and characterizing it to highlight patterns related to AD or normal tissues. Method computes from ROI a fully connected graph whose nodes correspond to consecutive segments of the RoI boundary and whose edges stand for the image intensity along paths connecting two opposite nodes. The graph topology and edge weights are summarized as an adjacency matrix, from which a set of eigenvectors are computed. The adjacency matrix is then used to extract discriminant coefficients that represents locations with higher salient lines. Additionally a dimensionality reduction is further accomplished by selecting the pair of nodes with major contribution for each of the computed eigenvectors. This strategy takes advantage of saliency of tissues in mammography image and allows to approximate a large variety of spiculated patterns.

Bibliografía

[1] ACo, Radiology: *Illustrated breast imaging reporting and data system (bi-rads)*. American College of Radiology, Reston, VA, 2006

[2] A.Feig, Stephen ; J.D'Orsi, Carl ; Hendrick, R .. ; P.Jackson, Valerie ; Kopans, Daniel B. ; Monsees, Barbara ; Sickles, Edward A. ; Carol B. Stelling, Pamela Wilcox-Buchalla: American College of Radiology Guidelines for Breast Cancer Screening. En: *American Journal of Roentgenology* 171 (1998), p. 29–33

[3] Al-Shamlan, H. ; El-Zaart, A.: Feature extraction values for breast cancer mammography images. En: *Bioinformatics and Biomedical Technology (ICBBT), 2010 International Conference on*, 2010, p. 335 –340

[4] Ayres, F. J. ; Rangayyan, R. M.: Characterization of architectural distortion in mammograms. En: *IEEE Enginering in Medicine and Biology Magazine* (2005), p. 59–67

[5] Ayres, F. J. ; Rangayyan, R. M.: Characterization of architectural distortion in mammograms. En: *IEEE Eng. Med. Biol. Mag.* 24(1) (2005), p. 59–67

[6] Ayres, F. J. ; Rangayyan, R. M.: Reduction of false positives in the detection of architectural distortion in mammograms by using a geometrically constrained phase portrait model. En: *International Journal of Computer Assisted Radiology and Surgery* 1 (2007), p. 361–369

[7] Ayres, F. J. ; Rangayyan, R. M.: Reduction of false positives in the detection of architectural distortion in mammograms by using a geometrically constrained phase portrait model. En: *Int. J. Comput. Assist. Radiol. Surg.* 1(6) (2007), p. 361–369

[8] Azavedo E, Mejáre I Heibert Arnlind M.: Is single reading with computer-aided detection (CAD) as good as double reading in mammography screening? A systematic review. En: *BMC Med Imaging.* (2012)

[9] Baker, Jay A. ; Rosen, Eric L. ; Lo, Joseph Y. ; Gimenez, Edgardo I. ; RuthWalsh ; Soo, Mary S.: Computer-Aided Detection (CAD) in Screening Mammography: Sensitivity of Commercial CAD Systems for Detecting Architectural Distortion. En: *American Journal of Roentgenology* 181(4) (2003), p. 1083–1088

[10] Baker, Jay A. ; Rosen, Eric L. ; Lo, Joseph Y. ; Gimenez, Edgardo I. ; Walsh, Ruth ; Soo, Mary S.: Computer-Aided Detection (CAD) in Screening Mammography: Sensitivity of Commercial CAD Systems for Detecting Architectural Distortion. En: *American Journal of Roentgenology* 181 (2003), Nr. 4, p. 1083–1088

[11] Banik, S. ; Rangayyan, R.M. ; Desautels, J.E.L.: Detection of Architectural Distortion in Prior Mammograms. En: *Medical Imaging, IEEE Transactions on* 30 (2011), Feb, Nr. 2, p. 279–294. – ISSN 0278–0062

[12] Banik, Shantanu ; M.Rangayyan, Rangaraj ; Desautels, J. E. L.: Detection of Architectural Distortion in Prior Mammograms. En: *IEEE Transactions on Medical Imaging* 30 (2011), Nr. 2, p. 279–294

[13] Bird, R. ; Wallace, T. ; Yankaskas, B.: Analysis of cancers missed at screening mammography. En: *Radiology* 178 (1992), p. 234–247

[14] Biswas, Sujjoy K. ; Mukherjee, Dipti P.: Recognizing Architectural Distortion in Mammogram: A Multiscale Texture Modeling Approach with GMM. En: *IEEE Transactions on Biomedical Engineering* 58 (2011), Nr. 7, p. 2023–2030

[15] Bovik, Alan C. ; Huang, Thomas S. ; Munson Jr, David C.: A generalization of median filtering using linear combinations of order statistics. En: *Acoustics, Speech and Signal Processing, IEEE Transactions on* 31 (1983), Nr. 6, p. 1342–1350

[16] te Brake, G.M. ; Karssemeijer, N.: Single and multiscale detection of masses in digital mammograms. En: *Medical Imaging, IEEE Transactions on* 18 (1999), Nr. 7, p. 628–639. – ISSN 0278–0062

[17] Buseman, Mouchawar J. Calonge N. Byers. T.: Mammography screening matters for young women with breast carcinoma. En: *Cancer* 97 (2003), p. 352 – 358

[18] Buseman, S. ; Mouchawar, J. ; Calonge, N. ; Byers., T.: Mammography screening matters for young women with breast carcinoma. En: *Cancer* 97 (2003), p. 352–358

[19] contra el cancer, Liga. *Cáncer de seno, Reporte Tecnico revizado 2013*. online. December 2013

[20] Chang, Chih-Chung ; Lin, Chih-Jen: LIBSVM: A library for support vector machines. En: *ACM Transactions on Intelligent Systems and Technology* 2 (2011), p. 27:1–27:27. – Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm

[21] Cortes, Corinna ; Vapnik, Vladimir: Support-Vector Networks. En: *Mach. Learn.* 20 (1995), September, Nr. 3, p. 273–297. – ISSN 0885–6125

[22] Deserno, T.M. ; Soiron, M. ; de Oliveira, J.E.E. ; de A Araujo, A.: Towards Computer-Aided Diagnostics of Screening Mammography Using Content-Based Image Retrieval. En: *Graphics, Patterns and Images (Sibgrapi), 2011 24th SIBGRAPI Conference on*, 2011, p. 211 –219

[23] Doi, Kunio: Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential. En: *Comput Med Imaging Graph* 31(4-5) (2007), p. 198–211

[24] DT, Ramsay ; JC, Kent ; RA, Hartmann ; PE., Hartmann: Anatomy of the lactating human breast redefined with ultrasound imaging. En: *Journal of Anatomy* 206(6) (2005), p. 525 – 534

[25] Freeman, L. C.: Centrality in social networks: Conceptual clarification. En: *Social Networks* 1 (1979), p. 215–239

[26] Freeman, L. C.: Centrality in social networks: Conceptual clarification. En: *Social Networks* 1 (1979), p. 215–239

[27] Gopalakrishnan, Hu Y. Rajan D.: Random Walks on Graphs for Salient Object Detection in Images. En: *IEEE TRANSACTIONS ON IMAGE PROCESSING* 19 (2010), p. 3232 – 3242

[28] Guo, Q. ; Shao, J. ; Ruiz, V.: Investigation of support vector machine for the detection of architectural distortion in mammographic images. En: *Journal of Physics: Conference Series 15*, 2005

[29] Homer, M. J.: *Mammographic Interpretation: A Practical Approach.* Second edition. McGraw-Hill Professional, 1997

[30] Horvath, María Cecilia y schonstedt V.: EXISTEN CANCERES NO DETECTABLES EN LA MAMOGRAFIA? En: *Rev. chil. radiol. [online]* 13 n. 2 (2007), p. 84–89

[31] Ichikawa, T. ; Matsubara, T. ; Hara, T. ; Fujita, H. ; Endo, T. ; Iwase, T.: Automated detection method for architectural distorion areas on mammograms based on morphological processing and surface analysis. En: Processing, Image (Ed.): *Medical Imaging* Vol. 5370, 2004

[32] Jardines, Lori ; Goyal, Sharad ; Fisher, Paul ; Weitzel, Jeffrey ; Royce, Melanie ; Goldfarb, Shari B.: *Breast Cancer Overview: Risk Factors, Screening, Genetic Testing, and Prevention.* 2013. – http://www.cancernetwork.com/cancer-management/breast-overview/article/10165/1802560#

[33] KAMRA, Amit ; JAIN, VK ; SINGH, Sukhwinder ; MITTAL, Sunil: Characterization of Architectural Distortion in Mammograms Based on Texture Analysis Using Support Vector Machine Classifier with Clinical Evaluation. En: *Journal of Digital Imaging* (2015), p. 1–11. – ISSN 0897–1889

[34] KARSSEMEIJER, N. ; TE BRAKE, G. M.: Detection of stellate distortions in mammograms. En: *IEEE Transactions on Medical Imaging* 15 (1996), Nr. 5, p. 611–619

[35] KARSSEMEIJER, N. ; TE BRAKE, G.M.: Detection of stellate distortions in mammograms. En: *Medical Imaging, IEEE Transactions on* 15 (1996), Nr. 5, p. 611–619. – ISSN 0278–0062

[36] KNUTZEN, A. M. ; GISVOLD, J. J.: Likelihood of malignant disease for various categories of mammographically detected, nonpalpable breast lesions. En: *Mayo Clinic Proc.* 68 (1993), p. 454 – 460

[37] MALICH, Ansgar ; MARX, Christiane ; FACIUS, Mirjam ; BOEHM, Thomas ; FLECK, Marlies ; KAISER, WernerA.: Tumour detection rate of a new commercially available computer-aided detection system. En: *European Radiology* 11 (2001), Nr. 12, p. 2454–2459. – ISSN 0938–7994

[38] MATSUBARA, T. ; ICHIKAWA, T. ; HARA, T. ; FUJITA, H. ; KASAI, S. ; ENDO, T. ; IWASE, T.: Automated detection methods for architectural distortions around skinline and within mammary gland on mammograms. En: *International Congress Series* 1256 (2003-06-01T00:00:00), Nr. 1, p. 950–955

[39] MATSUBARA, T. ; ITO, A. ; TSUNOMORI, A. ; HARA, T. ; MURAMATSU, C. ; ENDO, T. ; FUJITA, H.: An automated method for detecting architectural distortions on mammograms using direction analysis of linear structures. En: *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 2015, p. 2661–2664

[40] MENCATTINI, A. ; SALMERI, M. ; CASTI, P. ; RAGUSO, G. ; L'ABBATE, S. ; CHIEPPA, L. ; ANCONA, A. ; MANGIERI, F. ; PEPE, M.L.: Automatic breast masses boundary extraction in digital mammography using spatial fuzzy c-means clustering and active contour models. En: *Medical Measurements and Applications Proceedings (MeMeA), 2011 IEEE International Workshop on*, 2011, p. 632–637

[41] MICHAEL HEAT, Daniel Kopans Richard M. ; KEGELMEYER, Philip: The digital database for screening mammography. En: *Proceedings of the 5th international workshop on digital mammography* Citeseer, 2000, p. 212–218

[42] Morton, Marilyn J. ; Whaley, Dana H. ; Brandt, Kathleen R. ; Amrami, Kimberly K.: Screening Mammograms: Interpretation with Computer aided Detection Prospective Evaluation. En: *Radiology* 239 (2006), Nr. 2, p. 375–383. – PMID: 16569779

[43] Muralidhar, Gautam S. ; Markey, Mia K. ; Bovik, Alan C.: Snakules for automatic classification of candidate spiculated mass locations on mammography. En: *Image Analysis Interpretation (SSIAI), 2010 IEEE Southwest Symposium on*, 2010, p. 197 –200

[44] Muramatsu, Chisako ; Li, Qiang ; Schmidt, Robert ; Suzuki, Kenji ; Shiraishi, Junji ; Newstead, Gillian ; Doi, Kunio: Experimental determination of subjective similarity for pairs of clustered microcalcifications on mammograms: Observer study results. En: *Medical Physics* 33 (2006), Nr. 9, p. 3460–3468

[45] Muramatsu, Chisako ; Li, Qiang ; Suzuki, Kenji ; Schmidt, Robert A. ; Shiraishi, Junji ; Newstead, Gillian M. ; Doi, Kunio: Investigation of psychophysical measure for evaluation of similar images for mammographic masses: Preliminary results. En: *Medical Physics* 32 (2005), Nr. 7, p. 2295–2304

[46] Nemoto, M. ; Honmura, S. ; Shimizu, A. ; Furukawa, D. ; Kobatake, H. ; Nawano, S.: A pilot study of architectural distortion detection in mammograms based on characteristics of line shadows. En: *International Journal of Computer Assisted Radiology and Surgery* 4 (2009), Nr. 1, p. 27–36

[47] de Oliveira, Júlia E. ; Machado, Alexei M. ; Chavez, Guillermo C. ; Lopes, Ana Paula B. ; Deserno, Thomas M. ; de A. Araújo, Arnaldo: MammoSys: A content-based image retrieval system using breast density patterns. En: *Computer Methods and Programs in Biomedicine* 99 (2010), Nr. 3, p. 289 – 297. – ISSN 0169–2607

[48] Organization., World H.: *Cáncer de mama: prevención y control*. Online. september 2012. – http://www.who.int/topics/cancer/breastcancer/es/index.html.

[49] Parr, Tim C. ; Taylor, Christopher J. ; Astley, Susan M. ; Boggis, Caroline R.: S modeling of oriented line patterns in mammograms. En: *Medical Imaging 1997* International Society for Optics and Photonics, 1997, p. 44–55

[50] Petrick, Nicholas ; Sahiner, Berkman ; Chan, Heang-Ping ; Helvie, Mark A. ; Paquerault, Sophie ; Hadjiiski, Lubomir M.: Breast Cancer Detection: Evaluation of a Mass Detection Algorithm for Computer aided Diagnosis Experience in 263 Patients. En: *Radiology* 224 (2002), Nr. 1, p. 217–224. – PMID: 12091686

[51] Pisano, Etta D. ; Zong, Shuquan ; Hemminger, Bradley M. ; DeLuca, Marla ; Johnston, R. E. ; Muller, Keith ; Braeuning, M. P. ; Pizer, Stephen M.: Contrast Limited Adaptive Histogram Equalization image processing to improve the detection

of simulated spiculations in dense mammograms. En: *Journal of Digital Imaging* 11 (1998), Nr. 4, p. 193–200

[52] RANGAYYAN, Rangaraj M. ; CHAKRABORTY, Jayasree ; BANIK, Shantanu ; MUKHO-PADHYAY, Sudipta ; DESAUTELS, J.E L.: Detection of Architectural Distortion Using Coherence in Relation to the Expected Orientation of Breast Tissue. En: *IEEE Transactions on CBMS* 2 (2012), p. 1–4

[53] RANGAYYAN, R.M. ; AYRES, F. J.: Gabor filters and phase portraits for the detection of architectural distortion in mammograms. En: *Med. Biol. Eng. Comput.* 44(10) (2006), p. 883–894

[54] RANGAYYAN, R.M. ; BANIK, S. ; DESAUTELS, J. E L.: Detection of architectural distortion in prior mammograms using measures of angular dispersion. En: *Medical Measurements and Applications Proceedings (MeMeA), 2012 IEEE International Symposium on*, 2012, p. 1–4

[55] RANGAYYAN, R.M. ; CHAKRABORTY, J. ; BANIK, S. ; MUKHOPADHYAY, S. ; DESAU-TELS, J. E L.: Detection of architectural distortion using coherence in relation to the expected orientation of breast tissue. En: *Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on*, 2012. – ISSN 1063–7125, p. 1–4

[56] RAO, A. R. ; JAIN, R. C.: Computerized flow field analysis: Oriented texture fields. En: *IEEE Trans. Pattern Anal. Mach. Intell.* 14 (1992), Nr. 7, p. 693–709

[57] SAMPAT, M. P. ; MARKEY, M. K. ; BOVIK, A. C.: Measurement and detection of spiculated lesions. En: *Image Analysis and Interpretation*, 2006, p. 105–109

[58] SAMPAT, Mehul P. ; WHITMAN, Gary J. ; MARKEY, Mia K. ; BOVIK, Alan C.: Evidence based detection of spiculated masses and architectural distortions. En: *Medical imaging* International Society for Optics and Photonics, 2005, p. 26–37

[59] SICKLES, E.A. ; DÓRSI, C.J. ; BASSETT, L.W. ; ET AL.: *ACR BI-RADS® Mammography. In: ACR BI-RADS.ᴬtlas, Breast Imaging Reporting and Data System.* Fifth Edition. American College of Radiology, Reston, VA, 2013

[60] SOCIETY, American C.: Breast Cancer Facts & Figures 2013-2014 / American Cancer Society, Atlanta, Georgia. 2013. – Informe de Investigación

[61] SOCIETY, American C. *What are the key statistics about breast cancer?* november 2013

[62] SUCKLING, John ; PARKER, J ; DANCE, D ; ASTLEY, S ; HUTT, I ; BOGGIS, C ; RIC-KETTS, I ; STAMATAKIS, E ; CERNEAZ, N ; KOK, S [u. a.]: The Mammographic Image Analysis Society Digital Mammogram Database Exerpta Medica. En: *International Congress Series* 1069 (1994), p. 375–378

[63] TANG, Jinshan ; RANGAYYAN, R.M. ; XU, Jun ; EL NAQA, I. ; YANG, Yongyi: Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances. En: *Information Technology in Biomedicine, IEEE Transactions on* 13 (2009), march, Nr. 2, p. 236 –251. – ISSN 1089–7771

[64] TOURASSI, G. D. ; DELONG, D. M. ; FLOYD, C. E.: A study on the computerized fractal analysis of architectural distortion in screening mammograms. En: *Physics in Medicine and Biology* 51 (2006), Nr. 5, p. 1299–1312

[65] VERMA, K. ; ZAKOS, J.: A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques. En: *IEEE Transactions on Information Technology in Biomedicine* 16 (2002), p. 219–223

[66] WARREN BURHENNE LJ, D'Orsi CJ Feig SA Kopans DB O'Shaughnessy KF Sickles EA Tabar L Vyborny CJ Castellino R.: Potential contribution of computer-aided detection to the sensitivity of screening mammography. En: *Radiology* 215(2) (2000), p. 554–562

[67] WINSBERG, Fred ; ELKIN, Milton ; JOSIAH MACY, Jr. ; BORDAZ, Victoria ; WEYMOUTH, William: Detection of Radiographic Abnormalities in Mammograms by Means of Optical Scanning and Computer Analysis. En: *Radiology* 89 (1967), Nr. 2, p. 211–215

[68] YANKASKAS BC, Bird RE Desrochers D.: Reassessment of breast cancers missed during routine screening mammography: a community-based study. En: *AJR Am J Roentgenol.* 3 (2001), september, Nr. 177, p. 535 – 541

[69] ZWIGGELAAR, R. ; PARR, T.C. ; SCHUMM, J.E. ; HUTT, I.W. ; TAYLOR, C.J. ; ASTLEY, S.M. ; BOGGIS, C.R.: Model-based detection of spiculated lesions in mammograms. En: *Medical Image Analysis* 3 (1999), Nr. 1, p. 39–62

[70] ZWIGGELAAR, Reyer ; ASTLEY, Susan M. ; BOGGIS, Caroline R. M. ; TAYLOR, Christopher J.: Linear Structures in Mammographic Images: Detection and Classification. En: *IEEE TRANSACTIONS ON MEDICAL IMAGING* 23 (2004), Nr. 9, p. 1077–1086

[71] ZWIGGELAAR, Reyes ; PARR, Timothy C. ; SCHUMM, James E. ; HUTT, Ian W. ; TAYLOR, Christopher J. ; ASTLEY, Susan M. ; BOGGIS, Caroline R.: Model-based detection of spiculated lesions in mammograms. En: *Medical Image Analysis* 3 (1999), Nr. 1, p. 39 – 62. – ISSN 1361–8415