AUTOMATIC ANALYSIS OF MAMMOGRAPHY IMAGES: CLASSIFICATION OF BREAST DENSITY

Rita Filipa dos Santos Teixeira

MSc in Biomedical Engineering

Faculdade de Engenharia da Universidade do Porto

July 2013

AUTOMATIC ANALYSIS OF MAMMOGRAPHY IMAGES: CLASSIFICATION OF BREAST DENSITY

M.Sc.Thesis

Rita Filipa dos Santos Teixeira

Supervised by:

Prof. João Manuel R. S. Tavares

Associate Professor at the Mechanical Engineering Department

Faculdade de Engenharia da Universidade do Porto

Porto, July 2013

Acknowledgements

I am grateful to Professor João Manuel R.S. Tavares for availability, orientation and support granted to this work.

I would also like to thank to Inês Domingues, PhD Student at INESC Porto, for the support and assistance provided in the explanation of the documentation of the INBreast database and the continuous availability.

I would also like to acknowledge to Professor Sandra Ventura for the availability to evaluate a lot of images, it was very important for this work.

Lastly, I would like to show my immense gratitude to my family and friends for supporting me all the time along my studies and for being proud on my academic path.

Abstract

Breast cancer is the most common malignancy of women and is the second most common and leading cause of cancer deaths among them. At present, there are no effective ways to prevent breast cancer, because its cause is not yet fully known. Early detection is an effective way to diagnose and manage breast cancer can give a better chance of full recovery. Therefore, early detection of breast cancer can play an important role in reducing the associated morbidity and mortality rates.

Mammography has proven to be the most effective tool for detecting breast cancer in its earliest and most treatable stage, so it continues to be the primary imaging modality for breast cancer screening and diagnosis. Furthermore, this exam allows the detection of other pathologies and may suggest the nature such as normal, benign or malignant. The introduction of digital mammography is considered the most important improvement in breast imaging.

Computer-aided detection/diagnosis (CAD) has been shown to be a helpful tool in the early detection of breast cancer by marking suspicious regions on a screening mammogram, allowing thus to reduce the death rate among women with this disease. These systems use computer technologies to detect abnormalities in mammograms and the use of these results by radiologists for diagnosis play an important role, once characterize lesions through automatic image analysis. The CAD performance can vary because some lesions are more difficult to detect than others, this is because they have similar characteristics to normal mammary tissue. However, it is important to continue working in order to decrease the number of failures.

The pectoral muscle represents a predominant dense region in medio-lateral oblique views of mammograms. Its segmentation has been considered an important factor for an adequate performance of automatic cancer detection methods. Mammograms images are hard to interpret because of the complex tissue morphology of the breast and the number of imaging parameters that affect mammogram acquisition. Classification of breast density is important for epidemiological studies investigate the relationship between mammogram density and the occurrence of cancer. For these reason, there is an increasing interest in using measurements of mammographic density patterns in computer aided-detection.

Keywords

- Breast
- Cancer
- Medical Imaging
- Image analysis
- Mammography
- Pathologies
- Computer aided detection/diagnosis
- Pectoral muscle
- Breast Density

Resumo

O cancro de mama é o mais comum nas mulheres e é a segunda principal causa de morte de cancro entre as mesmas. Até ao presente, não existem formas de prevenir o cancro da mama, pois a sua causa não é ainda totalmente conhecida. A deteção precoce é uma forma eficaz de diagnosticar e tratar o cancro da mama dando uma maior possibilidade de total recuperação. Assim, a deteção precoce do cancro da mama tem um papel importante na redução dos índices de mortalidade.

A mamografia tem mostrado ser a ferramenta mais eficaz na deteção do cancro da mama numa fase precoce e mais facilmente tratável, continuando assim a ser a principal modalidade da imagiologia mais usada para diagnosticar o cancro da mama. Além disso, este exame permite detetar outras patologias e sugere a sua natureza, isto é, normal, benigna ou maligna. A introdução da mamografia digital é considerada o melhoramento mais importante na imagiologia mamária.

A deteção/diagnóstico assistida por computador (DAC) tem sido vista como sendo uma ferramenta de ajuda na deteção precoce pela marcação de regiões suspeitas num mamograma permitindo, assim, reduzir as taxas de morte nas mulheres com esta doença. Estes sistemas usam tecnologia computacional para detetar anormalidades num mamograma, e o uso destes resultados por radiologistas para diagnóstico possui um papel importante, uma vez que caracteriza lesões através da análise automática da imagem. O desempenho da DAC pode variar porque algumas lesões são mais difíceis de detetar do que outras, isto acontece pois têm características semelhantes ao tecido mamário normal. Contudo, é importante continuar a trabalhar para diminuir o número de falhas.

O músculo peitoral apresenta-se como uma região densa predominante nas incidências médio-laterais oblíquas dos mamograma, sendo a sua segmentação considerada importante para o desempenho adequado dos métodos de deteção automática de cancro. As imagens dos mamogramas são difíceis de interpretar porque possuem uma morfologia complexa do tecido da mama e o número de parâmetros de imagem que afetam a aquisição do mamograma. A classificação da densidade mamária é importante para em estudos epidemiológicos ser investigada a relação entre a densidade mamária e a ocorrência de cancro. Por esta razão, tem aumentado o interesse em usar medidas de padrões de densidade mamográfica no DAC.

Palavras – Chave

- Mama
- Cancro
- Imagem Médica
- Análise de imagem
- Mamografia
- Patologias
- Deteção assistida por computador
- Músculo Peitoral
- Densidade mamária

Index

Figures Index

Tables Index

Acronyms List

ACR – American College of Radiology

AEC – Automatic Exposure Control

AUC – Are Under the Curve

BI – RADS – Breast Imaging Reporting and Data System

BPNN – Back Propagation Neural Network

CAD – Computer Aided Diagnosis

CC – Cranio Caudal

DCIS – Ductal Carcinoma *in situ*

DWT – Discrete Wavelet Transform

FCC – Fibrocystic Change

FFDM – Full Field Digital Mammography

FN – False Negative

FP – False Positive

FPF – False Positive Fraction

FROC – Free Response ROC curve

FSM – Film Screen Mammography

IBC – Inflammatory Breast Carcinoma

ISSAC – Interactive Selective and Adaptive Clustering

LCIS – Lobular Carcinoma *in situ*

MIAS – Mammography Image Analysis Society

MIC – Multiple Image Comparison

MRI – Magnetic Resonance Imaging

MLO – Medio-Lateral Oblique

PCA – Principal Component Analysis

RBF – Radial Basis Function

RBST – Rubber Band Straightening Transform

ROC – Receiver Operating Characteristic

ROI – Region of interest

SIC – Single Image Comparison

SVM – Support Vector Machine

TN – True Negative

TP – True Positive

TPF – True Positive Fraction

UWT – Undecimated Wavelet Transform

CHAPTER I – Introduction

1.1 – Motivation

1.2 – Main Goals

1.3 – Report Organization

1.4 – Contributions

1.1 Motivation

Breast cancer is the most common malignancy of women and cause many deaths. The early detection is the best solution in order to avoid mastectomy, reduce the probability to return, and decrease the rate of mortality.

The mammographic exam allows to detect and characterize lesions in the breast, so it is important that women be awareness of this disease and do auto exams frequently, and after of a specific age it is recommended doing regularly mammography exams. There are many techniques for detect breast lesions, like ultrasonography and magnetic resonance imaging, but mammography is the more common choice. Mammography is a very useful exam; however, the similarity between normal and pathologic patterns may cause difficult diagnosis.

Computer aided diagnosis systems intend to support the doctor and other technicians in detection of mammary lesions, giving, thus, a second opinion.

Breast density is an important measure which shows the possibility for the detection of abnormalities in mammograms. Higher breast density usually indicates a higher possibility for the presence of malignant tissue. A human observer can distinguish different structures very well without the information of their overall brightness. In automatic breast density classification it is important to decide which parameters give the best division between categories.

1.2 Main Goals

The aim of this thesis is to develop a completely automated method to analyze and classify the mammographic images according to the density. In the future it can be used in a hierarchically CAD system that classify the images firstly according density, after indicates if there is a lesion and finally what kind of lesion is.

The developed techniques have been implemented in MATLAB ® and tested on real case studies.

1.3 Report Organization

The present thesis is organized and divided into six chapters. In the second chapter is explained the anatomy of the breast, as well as some pathologies that occur in the breast, and the main disease, the cancer in order to understand its imaging features. Also in this chapter some technical concepts and physics of mammography exam and equipment are explained.

The third chapter explains the history of computer aided detection, evaluation, it application in other organs and a little state of art of developed algorithms to detect breast lesions.

Afterwards, there is a full and detailed explanation of the steps developed since the distinction between directions and projections until the classification, including also the segmentation of pectoral muscle. The fifth chapter includes the experimental results and their evaluation. The last chapter presents the conclusions of this thesis where all work will be commented, as well as some future perspectives and tasks are proposed.

1.4 Contributions

This dissertation aimed to provide an initial study about breast and mammography. It provided also a review of computer aided detection algorithms in mammography and in other areas.

An algorithm that detects what direction and projection of breast is in cause was developed, in order to avoid wasting time and solve the problems with the lack of standard nomenclature.

An algorithm that detects segments and extracts pectoral muscle was developed thereby removing a portion of the image that can interfere with the results of the following steps. Once breast density has proved to be a very important factor in the analysis of mammography images because it can be a sign of cancer, in this dissertation a method was developed for the extraction and combination of features in order to classify breast density images of two image databases. As happened in this work does not always images are classified using the same system, then it is possible to classify an image as dense or non-dense.

This dissertation allows the creation of a completely automatic system from the simplest steps to the most complex because it also allows identify a specific type of lesions in mammography images, the masses. After the calcifications these lesions are the most frequent.

CHAPTER II – Breast and Mammography

2.1 – Breast Anatomy

2.2 – Breast Pathologies

2.3 – Breast Cancer

2.4 – Breast Imaging Reporting and Data System

2.5 – Breast Density

2.6 – Mammography equipment

2.7 – Digital Mammography

2.8 – Other technologies

2.9 – Summary

2. Breast and Mammography

The fundamental knowledge of the breast structures and some breast pathologies is essential to understand the importance of breast cancer study. Breast cancer is a malignant neoplasia produced by a cellular division dysfunction.

Mammography is a particular form of radiography, using radiation levels between specific intervals with a purpose to acquire breast images to diagnose an eventual presence of structures that indicates a disease, especially cancer. In case of mammary pathologies, their early detection is extremely important. The technological advances verified in imaging have contributed for the increase in the successful detection of breast cancer cases. In this area, mammography has an important role to allow detect lesions in initial stages and make a favorable prognosis.

2.1. Breast anatomy

During the fetal period is created, by epidermis, a depression which forms a mammary pit on the local of mammary gland. The region where appear the mammary glands is located in left and right sides of the upper ventral region of the trunk. The breasts exist in woman and man, but the mammary glands are normally most developed in female, except in some particular circumstances related with hormonal problems. The nipple is a small conical prominence surrounded by a circular area of pigmented skin, the areola, which contains large sebaceous glands that are often visible to the naked eye. The base of the female breast, roughly circular, extends from the second rib above to the sixth rib below. Medially, it borders the lateral edge of the body of the sternum, and laterally it reaches the mid axillary line, Figure 1 (Moinfar, 2007; Moore, 2004).

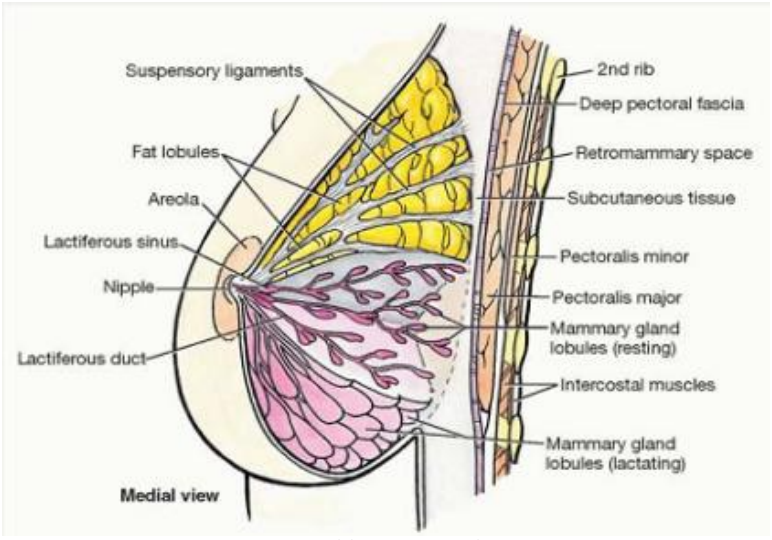


Figure 1 – Sagittal section of female breast (from (MOORE, 2004)).

At puberty, the female breasts normally grow according to the glandular development and increase of fat deposition; furthermore, also the nipples and areolas grow. The size and shape of breast depend of genetic, racial and dietary factors. During the pregnancy, the areola colour becomes dark, and after it keeps pigmentation. This colour diminishes as soon as lactation is over, but is never entirely lost throughout life (Gray, 2000; Moore, 2004).

The breast consists of gland tissue, fibrous tissue, connecting its lobes and fatty tissue in the intervals between lobes. The breast contains 15 to 20 lobes of glandular tissue, which constitute the parenchyma of the mammary gland. These lobes give a shape characteristic to the breast due to a considerable amount of fat, and these are composed of lobules, connected together by areolar tissue, blood vessels and ducts. Each lobule is drained by a lactiferous duct, which opens independently on the nipple. Just deep to the areola, each duct has a dilated portion, the lactiferous sinus, which accumulates milk during lactation. The smallest lobules include also the alveoli, which open into the smallest branches of the lactiferous ducts (Dixon, 2006; Moinfar, 2007; Moore, 2004).

Many changes happen in the breast tissue during the menstrual cycle and pregnancy, due to hormones progesterone and estrogens. In a woman who is not pregnant or suckling, the alveoli are very small and solid, but during the pregnancy enlarge, and the cells undergo rapid multiplication. The mammary glands only produce milk when the baby is born, despite be prepared for secretion since mid-pregnancy. The first milk, colostrums, eliminates the cells in the centre of the alveolus that suffered fatty degeneration. In a woman who has given birth more than twice the breast become large and pendulous, and in elderly women, they usually become small because of the decrease in fat and glandular tissue atrophy. But, normally in young women the breasts are supported and keep in their position by the Cooper's ligaments. These ligaments, particularly well developed in the upper part of the gland, help to maintain the lobes of the gland (Moore, 2004; Seeley, Stephens, & Tate, 2004).

As referred previously, hormones influence alterations in breast. Estrogens stimulate growth of the mammary glands, deposition of fat and development of lobules and alveoli. Progesterone and prolactin promote the final growth and are responsible for the function of these structures (Guyton & Hall, 2000).

2.2. Breast pathologies

2.2.1 Fibroadenoma

Fibroadenomas are the most common breast tumors in pubertal females, and there are three types of fibroadenoma classified as: common, giant and juvenile. These tumors are

characterized by a proliferation of both glandular and stromal elements, have well-demarcated borders, are firm, rubbery, freely mobile, solid, usually solitary breast masses. There is not pain or tenderness due to fibroadenomas and their size do not change with the menstrual cycle. Women aged in their 20s and adolescents are the most common people affected with this disease. A rapid growth sometimes occurs, but usually that growth is extremely slow. A giant fibroadenoma should measure over 5 cm in diameter, but the average is 2.5 cm. These tumors may be return, approximately 20% recur, and women should be aware of this risk and make periodic examinations (Dixon, 2006; Moinfar, 2007).

2.2.2 Mammary Dysplasia

Mammary dysplasia also can be called as fibrocystic changes (FCC), fibrocystic disease, fibrous mastopathy or fibroadenosis cystic. In reality, these alterations not indicate a disease.

This pathology is defined as being a benign alteration of the breast consisting of cystic dilatation of intralobular glands with or without stromal fibrosis. The age distribution of this lesion is between 20 and 50 years. Normally, fibrocystic changes are associated to the cyclic levels of ovarian hormones, because during ovulation and before menstruation, the hormone level changes often lead the breast cells to retain fluid and develop into nodules or cysts, which feel like a lump when touched. The texture of the breast is, in these cases, similar to the breast in pre-menstrual phase. The signs of fibrocystic changes include increased engorgement and density of the breasts, excessive nodularity, rapid change and fluctuation in the size of cystic areas, increased tenderness, and occasionally spontaneous nipple discharge. It can be unilateral, bilateral or just affect a part of the breast (Malik, 2010; Moinfar, 2007).

2.2.3 Mastitis and breast abscess

Inflammatory conditions of the breast, particularly acute mastitis and breast abscess are rare pathologies. Often these infections can happen in postpartum situations or after a lesion. There are two types of mastitis: acute and chronic. In acute mastitis, it is predominantly composed of neutrophilic granulocytes, seen mostly in lactating women. Chronic mastitis may be due to reinfection or a relapsed infection; the first case occurs sporadically and commonly is transmitted from the baby, and the second case means that eradication of the pathogen was failed (Jatoi, 2010; Moinfar, 2007).

Breast abscess arises when mastitis was treated inadequately and exist milk retention. The most common diagnosis techniques used to treat include ultrasonography of the breast and needle aspiration under local anesthesia with a purpose of identify collections of fluid or pus (Jatoi, 2010; Moinfar, 2007).

2.2.4 Gynecomastia

The male breast in some particular circumstances grows, and that disease is called gynecomastia. This pathology is benign and is defined as a potential reversible enlargement of the male breast with a proliferation of ductal epithelial and mesenchymal components. In general, gynecomastia occurs during the adolescence because of excessive estrogens related to testosterone or androgen deficiency. It is normal that this problem resolves spontaneously; however, it is a significant problem in men undergoing hormonal therapy for prostate cancer (Jatoi, 2010; Manni, 1999).

2.3. Breast Cancer

Cancer is a condition that affects people all over the world. Research in this area beginning since 1900 and cancer was a disease without cure. As other cancers, breast cancer arises when cells growth and multiply uncontrollably, which produces a tumor or a neoplasm. The tumors can be benign when the cancerous cells do not invade other body tissues or malignant if cells attack nearby tissues and travel through the bloodstream or lymphatic system to other parts of the body, spreading a cancer by a process known as metastasis (Alvin, 2006; Panno, 2005).

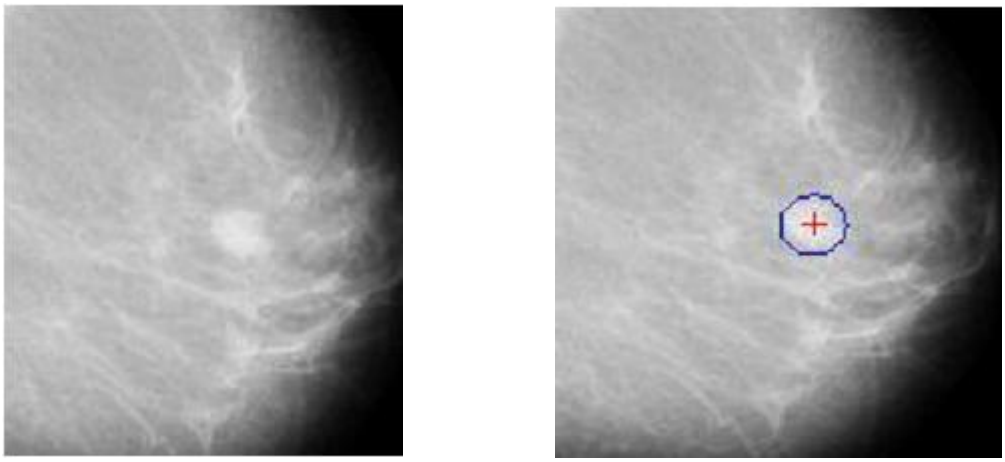
The damage that cancerous tumors cause to various important organs in the body can lead to serious illness, so an early detection is important for a better treatment and recovery. The most common range of age attacked by breast cancer is between 40-50 years old and until the menopause, the breast cancer rate incidence increases dramatically. There are other risk factors that lead to the development of a breast cancer as like age at menarche and menopause, age at first pregnancy, family history, previous benign breast disease and radiation (Dixon, 2006).

2.3.1 Breast Cancer Lesions

Breast cancer has some characteristic lesions such as microcalcifications, masses, architectural distortions and bilateral asymmetry.

a) Microcalcifications

Microcalcifications are small deposits of calcium of size from 0.33 to 0.7 mm and are slightly brighter than surrounding tissues. These lesions are difficult to detect in mammography because appear with low contrast due to their small size, although have high inherent attenuation properties. Associated with extra cell activity in breast tissue microcalcifications may show up in clusters or in patterns. A typical mammogram from MIAS database with microcalcifications is shown in Figure 2a) and its ground truth in Figure 2b) (Chang, 2005; Kavitha, 2007).



(a) (b)

Figure 2 – (a) Original mammography image; (b) Correspondent ground truth (from MIAS Database).

A microcalcification cluster normally is more detectable than an isolated microcalcification, and contributes for the diagnosis of early stages of breast cancer. These clusters may have three or more microcalcifications present in a mammogram region with an area around 1 cm^2. Once microcalcification may be a sign to malignancy it is important to be able to distinguish benign and malignant microcalcifications. Table 1 presents the grade, degree of suspicion and mammographic appearance (Kavitha, 2007; Rovere, 2006).

Table 1 – Grading of imaging reports of microcalcifications according to risk of malignancy (adapted from (Rovere, 2006)).

Grade	Degree of suspicion	Mammographic appearance
1	Normal	No abnormality seen
2	Consistent with a benign lesion	Popcorn, ring, micro cystic or diffuse bilateral calcification
3	Atypical or indeterminate but probably benign	Localized cluster of round, fine or punctuate calcification
4	Suspicious of malignancy	Localized cluster of granular calcification
5	Consistent with malignancy	Comedo calcification

b) Masses

Masses are lesions more difficult to detect in mammograms than microcalcifications because the features of a mass bear semblance to those of the normal breast parenchyma. In general, mass shape can be round, oval, lobular or irregular, and margins can be from circumscribed to spiculated, Figure 3.

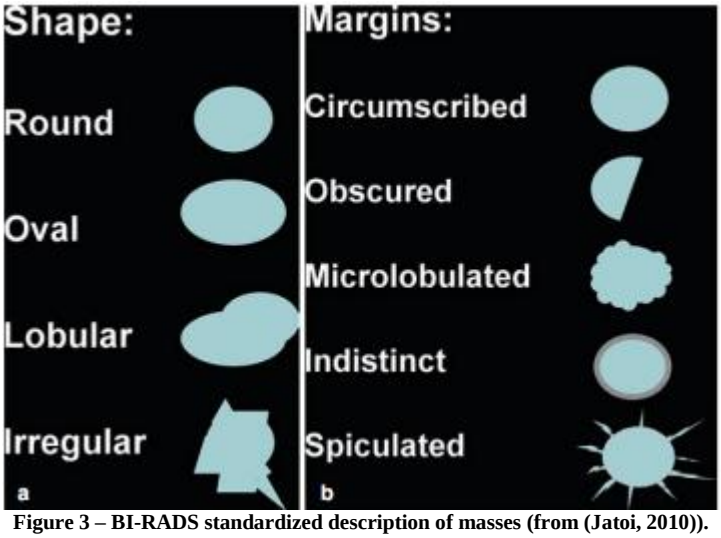


Figure 3 – BI-RADS standardized description of masses (from (Jatoi, 2010)).

When a mass is detected it is difficult to distinguish if is benignant or malignant but there are differences in the features of shape and texture between them. Benign masses are typically smooth and distinct, and their shapes are similar to the round. On the other hand malignant masses are irregular and their boundaries are usually blurry, Figure 4. A mass with regular shape has a higher probability of being benign whereas a mass with an irregular shape has a high probability of being malignant (Saki, 2010; Sun, 2010; Zhao, 2011).

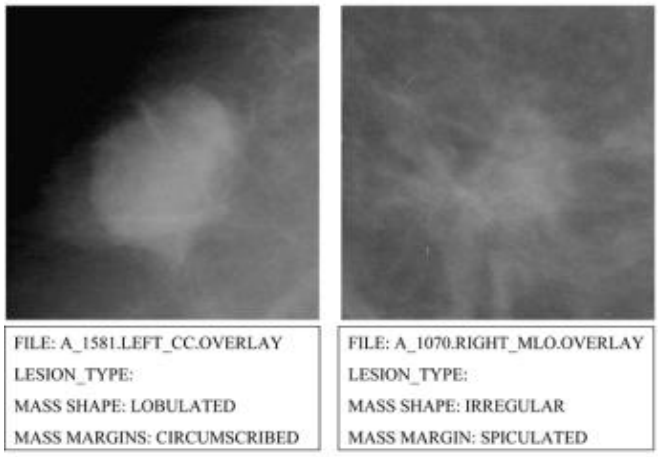


Figure 4 – Two malignant mass ROIs cropped from full-sized mammograms with their pathological characteristics specified (from (Digital Database for Sreening Mammography)).

c) Architectural distortions

The anatomy of the breast included several linear structures that cause directionally oriented texture in mammograms, so a way to detect architectural distortions is the change of normal texture of the breast. Architectural distortion could be categorized as malignant or benign; the

former, includes cancer, and the latter, includes scar and soft-tissue damage due to trauma. Due to its subtle appearance and variability in presentation, architectural distortion is the most commonly missed abnormality in false negative cases (Ayres, 2003; Banik, 2011).

d) Bilateral Asymmetry

Bilateral asymmetry of breast means a difference between corresponding regions in left and right breast and can be classified into global asymmetry and focal asymmetry. The first, happens when a greater volume of fibroglandular tissue is present in one breast compared to another in the same region. The latter, corresponds to a circumscribed area of asymmetry seen on two views, and usually is an island of healthy fibroglandular tissue that is superimposed with surrounding fatty tissue (Bozek, 2009; Li, 2009).

2.3.2 Types of Breast Cancer

Breast cancer can be classified into invasive/infiltrating or non-invasive/in situ, and both have characteristic patterns by which they can be classified. The noninvasive breast cancer is divided into ductal and lobular types and as the names induce ductal carcinomas arise from ducts and lobular carcinomas arise from lobules and do not destroy other tissues (Jatoi, 2010).

Ductal carcinoma in situ (DCIS) is the most common form of noninvasive carcinoma. The most common mammographic manifestation of DCIS is microcalcifications, and usually it is more easily to diagnose microcalcifications than masses, and thus is more effortlessly to detect DCIS than infiltrating carcinoma. This carcinoma is characterized by the proliferation of malignant mammary ductal epithelial cells that line the breast milk ducts, without evidence of invasion beyond the basement membrane. Once is important the size and extent of a DCIS to determine optimal surgical management and mammography had difficulties, some studies demonstrate most interest in other modalities, such as magnetic resonance imaging (MRI) that is more sensitive than mammography (Dixon, 2006;Jatoi, 2010; Moinfar, 2007).

Lobular carcinoma in situ (LCIS) arises in milk glands, and generally occurs before the menopause. This type of carcinoma it is hard to detect in mammography and when the woman does palpation. May be due to the difficulty to diagnosis LCIS, between 25 and 20% of women presenting this kind of tumor develop an invasive breast cancer (Dixon, 2006; Jatoi, 2010).

There exist other types of breast carcinomas, but with lower incidence, such as inflammatory breast carcinoma (IBC) and Paget's disease. Inflammatory breast carcinoma, Figure 5, are uncommon and characterized by brawny, redness, warmth, swelling, and erythematous skin changes and have the worst prognosis of all locally advanced breast cancers.

Figure 5 – Inflammatory breast carcinoma (from (Dixon, 2006)).

The IBC is produced by dermal lymphatic obstruction, and the diagnosis of this kind of cancer is done in advanced stages. The survival to this carcinoma it is reduced to 25-50 % after five years of evolution, due to an existence of the early lymphatic invasion (Dixon, 2006; Jatoi, 2010).

Paget's disease consists in a chronic eczematoid lesion of the nipple associated with an underlying breast malignancy. This disease can be due to a carcinoma in subareolar ducts (Dixon, 2006; Moinfar, 2007).

2.4 Breast Imaging Reporting and Data System

Several classifications have been used for classify the breast lesions and, although all of them are similar, the more accepted is the classification proposed by American College of Radiology (Breast Imaging Reporting and Data System – BI-RADS), Table 2.

Table 2 – BI-RADS report final assessment categories (adapted from (Jatoi, 2010)).

Category	Definition	Recommendation
0	Incomplete assessment	Additional imaging workup
1	Negative	Routine screening
2	Benign finding(s)	Routine screening
3	Probably benign	Short term follow-up (6 months)
4	Suspicious abnormality	Biopsy
5	Highly suggestive of malignancy	Appropriate action should be taken
6	Known (biopsy – proven) cancer	Appropriate action

2.5 Breast Density

Breast density is a strong indicator for breast cancer, which shows the possibility for the detection of abnormalities in mammograms. There is a great difficulty in detect tumors in the initial stage in density breasts, because it tends to mask the abnormal tissue. To minimize this problem, the main recommendation is the execution of periodic exams. The breast density decreases with the woman's age, reflecting the gradual substitution of fibroglandular tissue by fat tissue.

The relationship between mammographic density and the risk of development breast cancer was studied by Byng et al. They presented a tool that could be helpful during the preventive monitoring. The first step of this work was detect breast edges dividing the mammary structure and the background of the image, and then identify density region (Byng, 1998).

Other researchers also proposed the classification of mammograms by density through computation techniques. The classification presented considers local statistics and texture measures to divide the mammograms in fat or dens (Hajnal, 1993; Taylor, 1994).

The mammographic characteristics and the risk of breast cancer also was been studied by Byrne et al. The researchers affirmed that in nearly 2000 cases analyzed the risk almost duplicate when the woman presents some density in the breast. For women with a density above 75% the risk increases to more than four times.

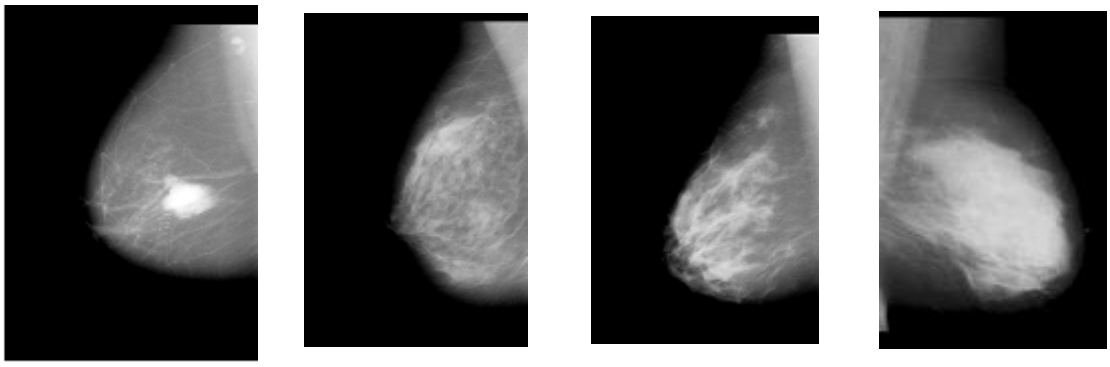
Many studies have been done to prove that the density should be include as a criteria of selection in women for screening programs and to determine how often is needed to take exams.

2.5.1 Classification of the breast according density

The mammary density is not perceptible by palpation but it is related with the fact of the x-rays penetrate the different types of breast tissues, so the relative amounts of fat, connective and epithelial tissue determines the radiographic appearance of the breast on a mammogram. Light (non-radiolucent) areas on the mammogram represent the fibrous and glandular tissues in the breast, whereas, the dark (radiolucent) areas are mostly fat (Ursin, 2009; Yaghjyan, 2011).

It was necessary create a method of classification of density patterns, so Wolfe's classification has been traditionally used in clinical practice and research. That classification had four breast parenchyma patterns or six category classification system based on the visual estimation of the percent density. The American College of radiology introduced a little modification in Wolfe's classification known as Breast Imaging- Reporting and Data System (BI-RADS) classification. The four categories of this density system are: BIRADS-1 indicates a predominantly fatty breast; BIRADS-2 scattered fibroglandular densities; BIRADS-3 a breast

that is heterogeneously dense and BIRADS-4, the highest level, an extremely dense breast that could obscure a lesion (Yaffe,2008; Yaghjyan,2011). Examples of these breast types are shown in Figure 6.



(a) (b) (c) (d)

Figure 6 – Mammograms with differing mammographic breast densities (a) predominantly fat; (b) fat with some fibroglandular tissue; (c) heterogeneously dense; (d) extremely dense (from (Moreira, 2012)).

Other measure that is related with density is sensitivity, because with high mammographic density we have a reduced ability of detecting an existing cancer. It is very difficult locate ill-defined cancers within an opaque uniform background. Thus, as mammographic breast density is a strong predictor of breast cancer the first step of image and processing analysis of mammogram is the density classification.

2.6 Mammography Equipment

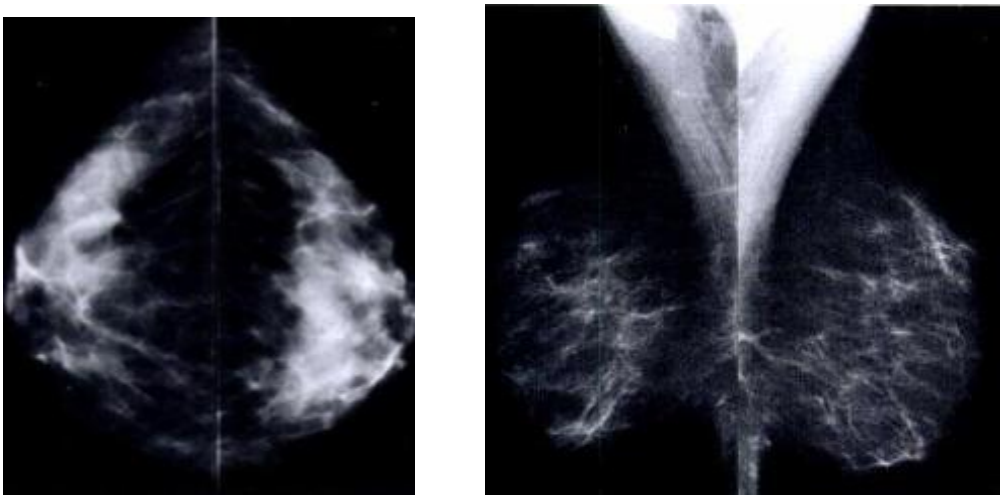
Mammography is an imaging procedure for examination of the breast that gives information about breast morphology, anatomy and pathologies. It is used for detection and diagnosis of breast cancer, as well as evaluates mass lesions in breast. The early detection of breast cancer is an important factor to treat this disease with success.

This procedure is similar to the other X-Rays, however, are used low doses, presenting a high quality that leads to high contrast and resolution and low noise (Sivaramakrishna, Gordon, 1997). The breast is sensitive to ionizing radiation, so it is desirable to use the lowest radiation dose compatible with excellent image quality.

Mammography is more sensitive and specific in assessing fatty breasts than dense breast. Dense breast tissue is particularly difficult to assess in young women. Mammography is also used in assisting needle core biopsies and for localization of non-palpable lesions (Chinyama, 2000). In screening mammography the uniform compression of the breast is important to ensure image contrast, thus these tools have to be highly sensitive, identifying as correctly as possible

those tumors that could be malignant. On the other hand, diagnostic mammography is usually more time consuming, expensive and provides higher radiation dose to the patient that screening mammography, so diagnostic tools must have a great specificity in order to really detect those tumors that are malignant (Perez, 2002).

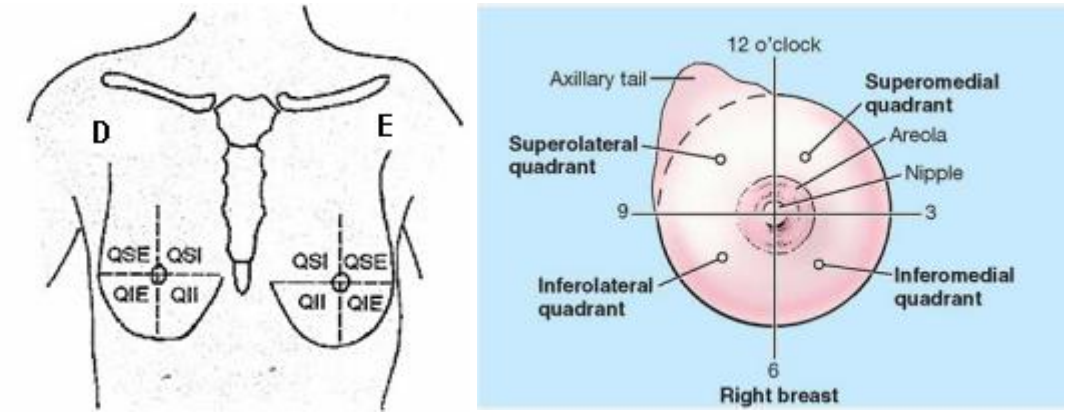
In order to assess differences in density between the breast tissue image acquisition is done using two views: a cranio caudal (CC) and mediolateral oblique (MLO). Generally, on MLO view, more breast tissue can be projected than on the CC view because of the slope and curve of the chest wall. On CC projection, Figure 7a, must be included all breast tissue with the exception of the axillaries portion. During the positioning is important to make sure that is included the upper and posterior breast tissue, through the elevation of breast within the limit of its natural mobility. The projection MLO, Figure 7b, should include as much breast tissue as possible. The image should include the free margin of the pectoral muscle to ensure that the patient is as far over the detector as possible and that the tail of the breast is imaged (Kopans, 2007; Engeland, 2003).



(a) (b)

Figure 7 – Two distinct mammographic projections: (a) Cranio Caudal view; (b) Medio Lateral Oblique view (from (Kopans, 2007)).

Once localized lesions are needed methods for divided the breast in small areas. The method of the quadrants divides the breast using the nipple as reference. Figure 8 shows two forms of represent this method.



(a) (b)

Figure 8 – Breast localization: (a) quadrants method; (b) nomenclature of each quadrant (from (Kopans, 2007; Moore&Agur, 2003)).

The essential components of the mammographic imaging system are a mammographic X-ray tube, a device for compressing the breast, an anti-scatter grid, a mammographic image receptor, and an automatic exposure control (AEC), as shown in Figure 9.

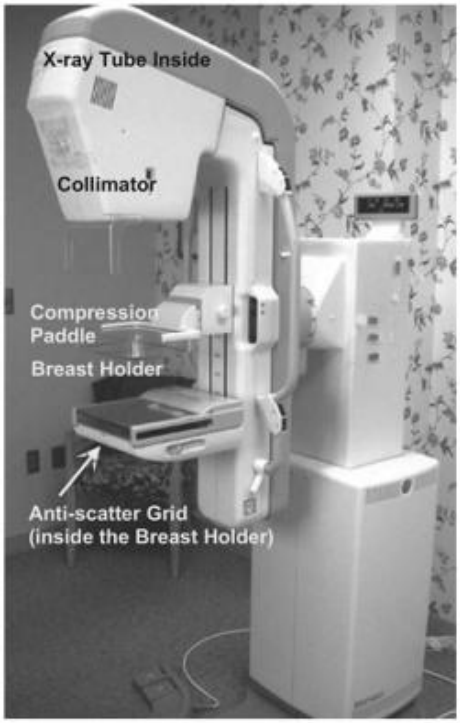


Figure 9 – A typical mammography unit (from (Akay, 2006)).

X-ray tube

An important difference in mammographic tube operation compared with conventional radiographic tube operation is the low operating voltage, typically below 35 kVp. The X-ray tubes are formed by a glass ampoule under vacuum, where there is an internal filament, a cathode and a target, the anode, having a window through which passes the useful beam. The electric current passing through the cathode produces electrons that are accelerated by the potential difference between cathode and anode (Akay, 2006; Webster, 2006). The voltage applied to cause the potential difference between cathode and anode, determines the maximum photon energy which will be transmitted for the breast. So, it is necessary the existence of a collimator and filters to limit and direct the output of radiation. The low voltages applied in the tube, there will cause a reduction in the energy beam produced, also reducing the dose received by the patient. The focus used to establish the image is provided by two filaments contained in the x-ray tubes, the large focus with 0.3 mm and the thin focus with 0.1 mm. These filaments must be made of a material having a high melting, usually made of tungsten, molybdenum or rhodium (Webster, 2006; Bronzino, 2000).

Compression Unit

During the image acquisition process, it should be avoided the loss of resolution caused by the patient movement, for that the patient must be hold breath, using the lowest exposure time possible and immobilizing the breast by compression. As seen previously, the uniform compression decreases the thickness of breast, allowing a better penetration of the X-ray beams. This procedure minimizes superimposition and geometric unsharpness, while resolution is improved as the distance to the object of interest is decreased. Furthermore, it is important to ensure that the area of concern is included in the field of view, because lesions can "roll" or "squeeze" out. It also improves the image contrast and minimizes superposition from different plans (Bronzino, 2000; Cardenosa, 2004).

Anti-Scatter Grid

These grids are composed of linear lead (Pb) separated by a rigid interspace material and placed between the breast holder and the cassette slot. The absorption grid is projected with adequate width and spacing with the purpose of absorbing mainly the scattered radiation without interfering on primary radiation. Anti-scatter grids are used to avoid an image contrast decrease produced by scattered radiation when reaches the image receptor. The static grids have thin slices with a low resolution compared with the system in order to not to be displayed in the image (Akay, 2006; Bronzino, 2000).

Image Receptor

The fluorescents screens in conjunction with single coated radiographic film allow that X-rays pass through the cover of a light-tight cassette and the film to impinge upon the screen. The cassettes guarantee a good contact between the screen and film. The crystals absorb the phosphor energy and produce light with an isotropic distribution. Higher sensitization of the film velocities is related to reducing the radiation dose to the patient. However, faster films tend to be noisier. The sensitization velocity is related to the film thickness; thinner films tend to be slower, but have better resolution. More recent films in mammography have higher qualities resolution, velocity and low noise. However, the scattering of light from the fluorescents screens reduces the resolution of the system. The radiographic films needs of chemical processes that absorb most of the light and display the image (Bronzino, 2000).

Automatic Exposure Control (AEC)

These devices of AEC compensate automatically the variations of absorption generated by the breast anatomy, reducing the final dose of radiation and avoiding repetitions. The sensors allow conjugate time exposure with the density and thickness of the tissues (Akay, 2006).

Noise

The doses in mammography are kept as low as possible because of health effects. However, X-ray quantum noise becomes more apparent in the image as the dose is lowered. This noise is due to the fact that there is an unavoidable random variation in the number of x-rays reaching a point on an image detector. The quantum noise depends on the average number of X-rays striking the image detector and is a fundamental limit to radiographic image quality. The granularity of the film is another source of noise and increases the higher the speed of film used. Hence, there is a necessity to adjust the speed to maintain a high image quality. The grain of the film is finite in size, and if the film image is enlarged, the inhomogeneity of the film grain becomes apparent (Bronzino, 2000; Kopans, 2007). In mammography, high image quality is essential because most of the relevant information of the mammogram corresponds to small details, such as microcalcifications, which can only be identified with a high spatial resolution image (Akay, 2006).

2.7 Digital Mammography

Digital mammography is the most recent significant development in mammography, which uses essentially the same conventional mammography system, but the screen film system is replaced by a detector, which produces an electronic signal that is digitized and stored, Figure 10. This technique has many practical advantages over film screen mammography (FSM), but only when its performance reaches the standards of FSM this new technology can be fully justified. Screening mammography has been shown to be effective in reducing breast cancer mortality. Full field digital mammography (FFDM) is increasingly used in the clinical setting, with a number of advantages resulting in better detection of breast cancer. With FFDM, the processes of image acquisition, processing, and display are decoupled or separated. This allows an optimization of each process (Bluekens, 2010; Pisano, 2004).

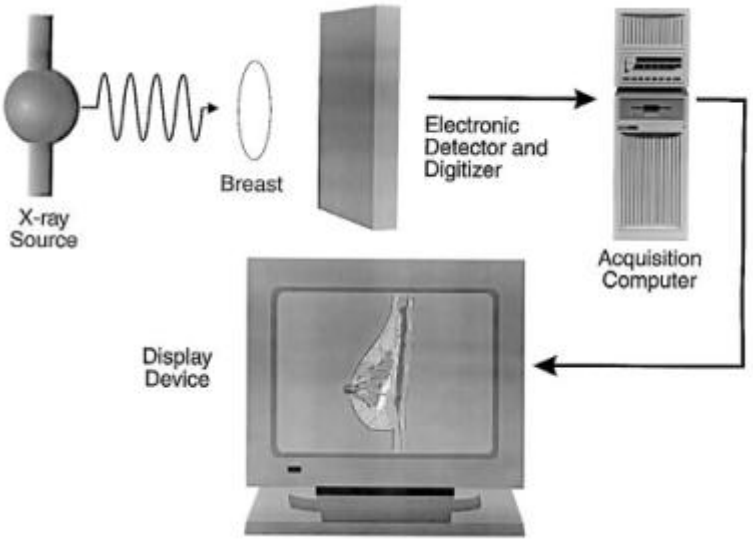


Figure 10 – Schematic representation of a digital mammography system (from (Bronzino, 2000))

.

In digital mammography, some of the limitations of FSM are overcome. The detected X-ray photons are converting directly to numeric values. The digital images can be processed by a computer, displayed in multiple formats, and fed directly to computer aided detection (CAD) software. Moreover, this technique allows the adjustment of the magnification, orientation, brightness and contrast of the image, after the exam (Bronzino, 2000; Tartar, 2008).

Digital images are sampled images, generated instantly after the exposure and have lower noise than SFM because of reduction in quantum mottle and elimination of granular artifacts from film emulsion. Digital mammography provides a broader dynamic range of densities and greater contrast resolution in dense breasts (Nees, 2008; Pisano, 2004).

In image processing, the contrast and brightness may be changed and enlarge part or entire breast. Post-acquisition processing may compensate for problems of underexposure or over

exposure. Image display systems for diagnosis include one or more high resolution monitors, a computer and image display software. For interpretation of image annotation and measurement tools are available, including linear, area and angle measurements. Sometimes, it is important to invert the gray-scale used for a better view of microcalcifications. Digital storage is able to store images in DICOM, it allows a rapid mode to access a large amount of data, but that requires a lot of computer memory. It is important that these systems be carefully designed to ensure that data are not lost, and personnel has to be trained to maintain the storage systems and retrieve new data for interpretation and prior exams for comparison (Hashimoto, 2008; Nees, 2008).

Although in some aspects, the digital mammography having a best performance than SFM, it is more expensive. It is necessary still working and does some improvements in order to reduce the disadvantages that this technique has.

2.8 Other technologies

The ultrasound is another promising technique in breast imaging, and it is especially benefic to distinguish between cysts and carcinogenic masses without calcification. Dynamic, contrast-enhanced Magnetic Resonance Imaging (MRI) of the breast has been shown to be extremely sensitive in the detection of invasive breast cancer and is not limited by the density of the breast tissue. However, this technique did not present enough resolution for identify pathological masses and calcifications as the conventional mammography. In proper situations, use MRI with other methods could have benefits. In terms of molecular imaging of the breast, a method is scintimammography which consists on use of gamma camera to obtain a breast image of a patient injected with radioisotopes. It is useful for patients who have dense breasts and suffered mammary surgery. The other method, galactography/ductography, is a specific exam using contrast radiologic agents for imaging the breast ducts. This technique can help diagnosing the cause of a discharge of abnormal nipple and diagnosing intra ductal papillomas.

2.9 Summary

The breast is studied by clinicians to understand better and make more easily the detection of breast pathologies, but cancer has a special attention due to their fatality in women. Different types of lesions can indicate breast cancer, such as microcalcifications, masses and architectural distortions. Furthermore, breast cancer can be classified according to the extent of the spread and the breast tissue where was originated. Some other diseases have patterns similar to the breast cancer, which difficult the diagnosis. A breast imaging reporting data system is used to classify the suspicion of breast cancer. Ever more the density of the breast is taken into account in the analysis of mammography images has been considered a risk factor for the development of breast cancer.

Mammography is very important for detect breast lesions when they are asymptomatic and when they are invasive and are in advanced stages.

This technique is an X-ray imaging procedure for examination of the breast that gives information about breast morphology, anatomy and pathologies. The main structures present in a mammogram is an X-ray tube, an anti scatter grid, a breast compressor, an image receptor and finally, an automatic exposure control to adjust the amount of radiation.

The digital mammography improved some aspects that the conventional mammography cannot. In order to try getting best results than mammography other methods such as Magnetic Resonance Imaging and ultrasound are also used in this area.

CHAPTER III – Computer Aided Diagnosis (CAD)

3.1 – CAD History

3.2 – CAD Systems

3.3 – CAD Evaluation

3.4 – CAD in brain and retinal pathologies

3.5 – CAD algorithms in breast lesions

3.6 – Summary

3. Computer Aided Diagnosis (CAD)

Computer aided detection or diagnosis (CAD) is an important application of pattern recognition, computer science and image processing and analysis techniques, aiming at assisting doctors in making diagnostic decisions. Once the data are often not easily interpretable the CAD systems help doctor detecting subtle lesions and reducing the probability of failure. These computational systems are promising in detection of suspect cases and for helping in the diagnostic decision as a "second opinion". Thus, in the past several years, CAD systems and related techniques have attracted the attention of both research scientists and radiologists (Marroco, 2010; Tang, 2009).

3.1 CAD History

The idea of using a computer to help in diagnosis started in 1960. In 1963, (Lodwick et al, 1963) investigated the use of a computer in diagnosing bone tumors. One year later, (Meyers et al, 1964) propose a system for distinguish, automatically, normal and abnormal chest radiographs. Winsberg et al (Winsberg, 1967) developed an automated method to diagnosis breast cancer in mammograms by means of optical scanning. However, this method was largely unsuccessful, like other pioneer studies. A computer system for breast cancer detection was designed to classify breast lesions on X-radiographies by Ackerman and Gose (Ackerman, 1972). Several years later, 1990, a new study conducted by Chan et al consisted in analyzes mammogram without and with the computer aid. This study contributed considerable for the growing of this field. Chan et al (Chan, 2008) found a statistically significant improvement in the performance of radiologists when they used the computer aid (Webster, 2006).

The 1998 year was very important because Food and Drug Admnistration (FDA) approved a commercial CAD system for screen mammography, ImageChecker^{TM}. In 2002 and 2004, respectively, FDA also approved iCAD and Eastman Kodak's Health Group for mammography. In breast ultrasonography, the software was designed to analyze breast ultrasound images by automatically segmenting and analyzing shape and orientation characteristics of suspicious lesions in user selected regions of interest (ROI). Today, it is estimated that more than 5000 mammography CAD systems are in current use in hospitals, clinics and screening center in the U.S. (Fujita, 2008).

The number of papers related to CAD research presented at the RSNA meetings from 2000 to 2005 is listed in Table 3. The majority of these presentations were concerned with three organs: chest, breast, and colon, but other organs such as brain, liver, and skeletal and vascular systems were also subjected to CAD research (Doi, 2007).

Table 3 – Number of CAD papers related to seven different organs, presented at the Annual Meetings of the Radiological Society of North of America (RSNA) in Chicago from 2000 to 2005 (from (Doi, 2007))

Year	2000	2001	2002	2003	2004	2005
Chest	22	37	53	94	70	48
Breast	23	28	32	37	48	49
Colon	4	10	21	17	15	30
Brain	-	4	2	10	9	15
Liver	3	-	5	9	9	9
Skeletal	2	7	7	9	8	5
Vascular etc	5	-	12	15	2	7
Total	59	86	134	191	161	163

3.2 CAD Systems

CAD systems have been developed for the detection of a variety of cancers, including breast and lung cancer as well as colonic polyps. Less usual, but when expert human observers are not available, these systems can be used in emergency situations. CAD systems generally perform automatic assessments of patient images and present to radiologist areas that they have determined as having the appearance of an abnormality. It is important have awareness that in different contexts CAD can have different performance, so it needs to be adjusted to produce the most accurate result (Chan, 2008; Paquerault, 2010).

It is clearly that CAD systems are very useful for health professional. These professionals have repetitive tasks, they observe many images in one day, and these systems allow for automate processes. There is another obvious advantage, related with the ability to observe and work with 3D and even 4D images. Two main types of CAD approaches are possible, the first is called CADe and identify suspicious regions in the image, and the second is CADx that classifies the suspicious regions. Sometimes, CADe is associated with computer aided detection and CADx with computer aided diagnosis. It is believed that CAD could be an effective and efficient solution for implementing double reading, which provides double perception and interpretation. As the name of these systems suggests, just an aid is given, and radiologists take the final decision (Doi, 2007; Webster, 2006).

The mainly goal of CAD systems is to improve the accuracy and efficiency of radiologists. These technicians can be attracted by some features and miss a lesion; this can occur when they are focused on identifying some suspect disease. In mammography, the variability between radiologists can be very large, and sometimes when a radiologist analyzes a same case with a spaced time he gives a different interpretation. The CAD systems read images faster without reducing accuracy, what not happens with radiologists, but medical community needs to be confidant in the results (Webster, 2006).

3.3 CAD Evaluation

The performance of CAD systems is variable and depends on the organ, disease, type of image finding, and so on. In general, CAD systems tend to err on the side of caution, presenting images with a large number of false positive (FP) marks. The CAD systems can be classified as true positive (TP), false positive (FP), true negative (TN) and false negative (FN) in the context of detecting the presence or absence of abnormality. In this classification, positive/negative refers to the decision made by the algorithm and true/false refers to how the decision agrees with the actual clinical state. The result false positive may put the patient in delicate and fragile position but, with the help of complementary exams, this result can be excluded. However, when in the results are a false negative, is a more worrying situation once the person has the lesion but the algorithm does not detect (Boyer, 2009; Sonka, 2000).

The performance of a CAD system can be limited to the detection of obvious cancers with a moderate sensibility and a relative good specificity. These metrics are based on true/false, positives/negatives metrics. The sensitivity refers to how often the algorithm used reports that an abnormality exists in the instances where it actually exists, and is also called true positive fraction (TPF) (Boyer, 2009; Sonka, 2000):

$$sensitivity = \frac{TP}{(TP + FN)} \quad \textbf{(1)}$$

On the other hand, specificity refers to how frequently the algorithm correctly reports normal when no abnormal exists, and False Positive Fraction (FPF) is the same as (1- specificity):

$$specificity = \frac{TN}{(TN + FP)} \quad \textbf{(2)}$$

Conversely, CAD systems can be oriented toward high sensibility without regard for decreasing specificity. The developers of CAD systems can choose whatever the system will have high specificity or high sensibility. At present, CAD systems have a sensitivity of detection around 88 to 92% in mammography. However, despite the software improvements, masses have the highest rate of false positives (Boyer, 2009).

The space of possible tradeoffs between sensitivity and specificity are summarized in a receiver operating characteristic (ROC) curve, Figure 11. The ROC curve is typically plotted with the TPF (sensitivity) on Y axis and the FPF (1- specificity) on the X axis. Thus, the ROC curve produced allows the detection of massive lesions with predictable performance. In the ROC, the ideal operating point is the upper left corner, with TPF=1 and FPF=0 (Sampat, 2005; Sonka, 2000).

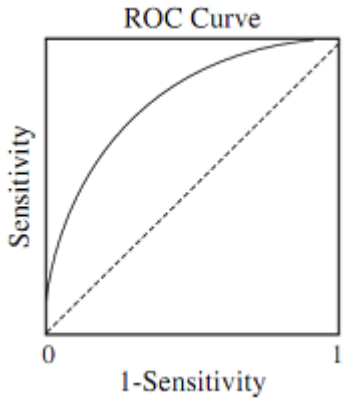


Figure 11 – Example of a ROC curve (from Sampat, 2005)).

It is common to use the area under the ROC curve (AUC) as a performance measure for evaluating the goodness of the curve. An algorithm that decides by random guessing will generate a curve that is a diagonal line, and will have an AUC of 0.5. The greater the area under the curve much higher the probability of making a correct decision. One ROC curve is globally better than another ROC curve if it has a greater AUC. However, since the AUC is an overall measure, it may hide important details (Sampat, 2005; Sonka, 2000).

To evaluate true-positive detection, sometimes is also required the localization of the tumor. The Free Response ROC (FROC) curve, Figure 12, is appropriate when the goal is to identify all instances of abnormalities in an image. In this scenario, as the algorithm can make an arbitrary number of false detections in an image, the number of FP is in principle unbounded. Since the maximum possible number of FP responses is not fixed, the x dimension of the FROC represents the average number of FP per image. The FROC represents the space of tradeoffs between the sensitivity and the generalized specificity of an algorithm. Two FROC curves may still be compared using the areas under the curves over the same range of FP (or TP) (Sampat, 2005; Sonka, 2000).

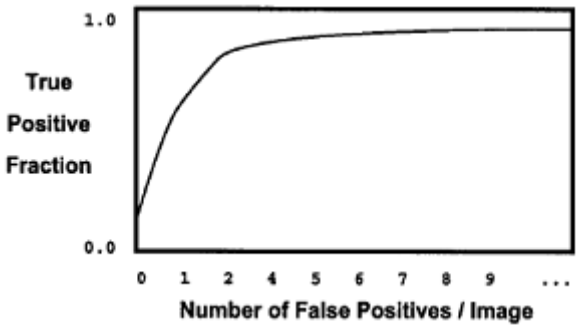


Figure 12 – Example of a FROC curve (from (Sonka, 2000)).

3.4 CAD in brain and retinal pathologies

The prevention of brain diseases is of paramount importance. MRI is very useful for the early detection of cerebral and cerebrovascular diseases. Lacunar infarcts, unruptured aneurysms, and arterial occlusions can be detected using MRI, and it is very important to detect them. However, visualization of these structures is not always easy for radiologists and neurosurgeons. CAD systems can assist neuroradiologists and general radiologists in detecting intracranial aneurysms, asymptomatic lacunar infarcts, and arterial occlusions and in assessing the risk of cerebral and cerebrovascular diseases (Doi, 2007).

Retinal fundus images are useful for the early detection of a number of ocular diseases, if left untreated, can lead to blindness. Examinations using retinal fundus images are cost effective and are suitable for mass screening. In view of this, retinal fundus images are obtained in many health care centers and medical facilities during medical checkups for ophthalmic examinations. The increase in the number of ophthalmic examinations improves ocular health care in the population, but it also increases the workload of ophthalmologists. Therefore, CAD systems developed for analyzing retinal fundus images can assist in reducing the workload of ophthalmologists and improving the screening accuracy (Doi, 2007).

3.5 CAD Algorithms in breast lesions

The most computer-aided detection and diagnosis computational algorithms in mammographic image analysis consist of the following typical steps: segmentation, feature extraction, feature selection and classification. But, before that, image enhancement is usually needed. The aim of the segmentation step in mammographic image analysis is to extract regions of interest (ROIs) containing breast abnormalities from the normal breast tissue. Another aim of the segmentation is to locate the suspicious lesion candidates from the region of interest. Following the segmentation, feature extraction consists in determine the characteristics of the region of interest. The input of this stage is a set of lesion candidate regions that are characterized by a number of features extracted from those regions. Some of the features extracted from the regions of interest in the mammographic image are not significant when observed alone, but in combination with other features, they can be significant for classification. The best set of features for eliminating false positives and for classifying lesion types as benign or malignant are selected in the feature selection step. The features extracted and selected in the previous steps are used to classify them into lesion non-lesion state aiming at substantially reduction of the false positives. Further classification can be considered, identifying different types of lesions, particularly if they benign or malignant.

3.5.1 Microcalcifications

The detection of microcalcifications in breast imaging has always been the role of mammography. The computer system should process the mammograms as digitized images for the detection of suspicious points as being microcalcifications. Then, microcalcifications have to be separated through a classification process and grouped into clusters. Finally, the obtained microcalcification clusters have to be classified as indicative of benign breast tissue or malignant, a sign of breast cancer.

A high contrast is always required to differentiate very fine structures with slight differences in density, such as microcalcifications, so enhancement is the first step in the process of microcalcification detection and classification. The methods of computer assisted enhancement emerged as consequence of the difficulty of radiologists in detection of small microcalcification clusters. However, the computer assisted programs often generate many false positives, thus it was necessary to reduce that. In 1995, an algorithm called Issac (Interactive Selective and Adaptive Clustering) had been developed that reduced the number of false positives and improved the diagnosis results. As referred previously, this algorithm comprised two parts, the first part identified sensitive sample domains in the feature space reducing the number of false positives. The second part was an interactive adaptation in which the radiologist or other user can improve the results by interactively identifying additional false positive or true negative samples. This tool was very useful in the sense that it helped radiologists bring their attention to possible microcalcification malignancy regions for a closer examination (Estevez, 1995).

Wavelet transform has been very used in computer analysis of mammographic images. Yu et al (Yu, 1996) enhance the microcalcifications using an undecimated wavelet transform (UWT). Once exists a similarity between high frequency noise and microcalcifications, UWT with their multiresolution capacity allows enhance microcalcification and suppress noise. Therefore, the input image are decomposed into different resolution levels that are sensitive to different frequency bands, and then is chosen an appropriate wavelet which will allow separate microcalcifications and high frequency noise into different resolution levels. By choosing the right resolution level can effectively enhance microcalcifications. This technique had the potential to become a useful method in detecting microcalcification (Yu, 1996).

Strickland et al. proposed a discrete wavelet transform (DWT) with bi-orthogonal spline filters. Four dyadic and two additional interpolating scales are extracted by introducing binary thresholds in every scale (Strickland, 1996). Laine et al. applied a wavelet-based enhancement methodology utilizing redundant transformation and linear/nonlinear mapping functions with Laplacian or gradient wavelet coefficients (Laine, 1995).

Fractal and fuzzy logic methodologies are also approaches resulting in sophisticated enhancement techniques. Li et al. proposed a fractal modeling scheme. The breast background

tissue results in high local self-similarity, but the presence of microcalcifications reduces it. Microcalcifications were enhanced based on the difference between the original and the modeled image (Li, 1997). Cheng et al. proposed a fuzzy logic contrast improvement methodology resulting in the enhancement of con tours and fine details of the mammographic structures. Local contrast values are improved by the utilization of the maximum fuzzy entropy principle and global information that are extracted by histogram analysis (Cheng, 1998). Jiang et al. applied a fuzzy enhancement operator restructuring the mammograms in fuzzy domain aiming on the detection of microcalcifications. The utilization of maximum fuzzy entropy principle results in the identification of regions with high local fuzzy contrast values which corresponds to microcalcifications (Jiang, 2005).

Wirth et al. enhance the contrast of microcalcifications based on morphological analysis. The suppression of background artifacts and the isolation of the breast region are the first steps. Then the perspicacity of microcalcifications is improved using morphological enhancement. This method allows the preservation of small details that were eliminated when they used traditional enhancement techniques. Moreover, it is able to enhance microcalcifications without noise emphasis or the over accentuation of background tissue (Wirth, 2004).

Segmentation is also an important step for detection and classification of microcalcifications. Chan et al. describe an algorithm for the segmentation of microcalcifications. The gray value resolution of their images is 10-bit. They work with disk-shaped simulated calcifications superimposed on a normal mammographic background. The authors state that the shape is not well reproduced. In one group, there are always eight to ten calcifications with a constant contrast of thirty gray levels relative to the background, and they are all smaller than 0.5 mm in diameter. The group is localized arbitrarily. These assumptions are questionable in the light of real mammograms, where neither the number of calcifications in a cluster, nor their contrast to the background is constant. For the segmentation, the difference between the signal-enhanced and the signal-suppressed image is calculated. Unless their method is tested with real images, its relevance remains doubtful (Chan, 1987).

Wróblewska et al. developed a method that has more than 90% of efficiency in microcalcification detection. Initially, they segment the individual objects, i.e., potential microcalcifications by applying opening by reconstruction top-hat technique and image thresholding based on approximation of an image local histogram with probability density function of Gauss distribution (Wróblewska, 2003).

Chan et al. and Wu et al. implement segmentation techniques based on local thresholding for individual pixels using the mean pixel value and rms noise fluctuation in a selected square region around the thresholded pixel. The threshold for a pixel is set as the mean value plus the rms noise value multiplied by a selected coefficient. A structure is segmented by connecting pixels that exceed the threshold. This method allows detect single or clustered

microcalcification because if a microcalcification is close to another or bright structure, the window used to compute the rms noise value around the first microcalcification will include the other bright structures, the noise rms may be overestimated, and the threshold may be set too high (Chan, 1987; Wu, 1992).

An extensive process of segmentation was done by Dengler et al. They use several steps that include high pass filtering, difference of Gaussian filtering, four computations of the standard deviation of the image, a smoothing, an opening, as well as an iterative thickening process with two erosions, two intersections and a union operation in each iteration (Dengler, 1993).

Yu and Guan have proposed a CAD system automatic detection of clustered microcalcifications in digitized mammograms. Using mixed features the potential microcalcification clusters are segmented, and then the individual microcalcification cluster is segmented using 31 features. The discriminatory power of these features is analyzed using general regression neural networks via sequential forward and sequential backward selection methods (Yu, 2000).

Netsch and Heinz-Otto, uses the Laplacian scale-space representation of the mammogram. First, possible locations of microcalcifications are identified as local maxima in the filtered image on a range of scales. For each finding, the size and local contrast are estimated, based on the Laplacian response denoted as the scale-space signature. A finding is marked as a microcalcification if the estimated contrast is larger than a predefined threshold which depends on the size of the finding (Netsch, 1999).

Kim and Park proposed a three layer Back Propagation Neural Network (BPNN) for automatic detection of microcalcification clusters. Texture features are extracted to classify the ROI containing clustered MC and ROI containing normal tissues (Kim, 1999).

3.5.2 Masses

Very different approaches have been applied for mass detection, and researchers have applied a number of distinct processing techniques to prompt masses on mammograms. Li et al. extracted suspicious areas by adaptive thresholding followed by a modified Markov random field model-based method (Li, 1995). Petrick et al. identified potential mass regions with a density weighted contrast enhancement segmentation combined with an object-based region growing technique. They merged morphological and texture features in a linear discriminant analysis classifier to discriminate suspicious regions as either mass or normal breast tissue (Petrick, 1999). Zheng et al. implemented a Gaussian band pass filtering technique for segmenting potential masses. A rule-based multilayer topographic feature analysis was then utilized to classify suspect regions. With the same purpose, but in a different way some investigators have developed mass detection algorithms based on the analysis of gradient

orientation maps (Zheng, 1995). Karssemeijer et al. proposed a statistical analysis of a map of pixel orientations to detect stellate patterns (Karssemeijer, 1995). Brake et al. followed the same approach to detect all types of masses at different scales (Brake, 1999). Kobatake et al. developed an iris filter to detect malignant masses on mammograms. The filter was intended to enhance malignant tumors with very weak contrast to their background; features extracted from the border of the enhanced areas were used to eliminate false- positive detections (Kobatake, 1999).

Usually, segmentation methods used for classification of mammographic masses can be manual, semi-automated or fully automated. Mudigonda et al. used hand-segmented mass boundaries to extract gradient-based features and texture measures from a ribbon surrounding the mass, and used the features in a stepwise discriminant analysis classifier (Mudigonda, 2000). Bruce et al. used multiresolution features to quantify mass shapes that were segmented manually by a radiologist (Bruce, 1999). Pohlman et al. segmented masses using an adaptive region growing algorithm. If a mass could not be segmented after repeated manual adjustments of the region growing algorithm, it was excluded from the data set. They found that in the task of differentiating invasive cancer and benign lesions, their tumor boundary roughness feature achieved classification accuracy comparable to those of two experienced radiologists who specialized in mammography (Pohlman, 1996). Kilday et al. extracted mass shapes using interactive gray-level thresholding. Masses that were not successfully segmented were excluded from analysis. Morphological features and patient age were used to classify the masses into cancer, cyst, and fibroadenoma categories (Kilday, 1993). Huo et al. used a fully automated region growing algorithm with a multiple transition point technique to segment the masses. Features extracted from the margin, and the density of the masses was used in a two-stage hybrid classifier consisting of a rule-based stage and an artificial neural network stage (Huo, 1998).

In order to extract and calculate features of the region of interest, Huo et al. developed a technique that analyzes patterns and quantifies the degree of speculation present. This study was divided into two steps, where the first consists in use region growing for automatic lesion extraction, and the second allows feature extraction using radial edge gradient analysis. Sahiner et al., introduced a new concept for characterize masses as malignant or benign, rubber band straightening transform (RBST). The RBST implies the transformation of a band of pixels localized around a segmented mass onto the Cartesian plane, and is obtained the RBST image. The border of a mammographic mass appears approximately as a horizontal line, and possible speculations resemble vertical lines in the RBST image (Sahiner, 1998). Similar to what happens in microcalcifications, wavelets are also used in mass detection. Qian et al. used multiresolution and multi-orientation wavelet transforms analyzing the influence of WTs on image feature extraction for mass detection. Moreover, is compared the discriminant ability of

features extracted with and without the wavelet-based image preprocessing using computed ROC (Qian, 1999). Based on temporal changes is possible extract features that allow the correct classification of masses. Hadjiiski et al. take current and prior mammograms automatically segmented by active contour method and from each mass extract several characteristics. The feature space was obtained using four main combinations of features. Stepwise feature selection and linear discriminant analysis classification were used to select and merge the most useful features (Hadjiiski, 2001).

Feature selection is important because of this step depends the choose of the best set of features for eliminating false positives and for a correct classification. In this stage, genetic algorithms are very useful, so Sahiner et al. implemented a method for feature selection that allows the differentiation mass / normal tissue (Sahiner, 1996).

Nandi et al. developed a technique called genetic programming, where several features related with edge sharpness, shape and texture were computed and, for refine the set of features available to the genetic programming classifier, they used features selection methods. This approach had a good evaluation because the accuracy and the success rate were incredibly high (Nandi, 2006).

3.5.3 Architectural Distortions

The breast contains several piecewise linear structures, such as ligaments, ducts, and blood vessels, that cause directionally oriented texture in mammograms. The presence of architectural distortion is expected to change the normal oriented texture of the breast. Characterization of such subtle changes from a pattern recognition perspective is the goal of many approaches.

Mudigonda and Rangayyan proposed the use of texture flow-field to detect architectural distortion, based on the local coherence of texture orientation. Only preliminary results were given, indicating the potential of the technique in the detection of architectural distortion (Mudigonga & Rangayyan, 2001). Matsubara et al. used mathematical morphology to detect architectural distortion around the skin line and a concentration index to detect architectural distortion within the mammary gland. The authors reported a sensitivity of 94% with 2.3 false-positives per image, and 84% with 2.4 false-positives per image, respectively (Matsubara, 2003). Evans et al. investigated the ability of a commercial CAD system to mark invasive lobular carcinoma of the breast: the system correctly identified 17 of 20 cases of architectural distortion (Evans, 2002). Birdwell et al. evaluated the performance of a commercial CAD system in marking cancers that were overlooked by radiologists; the software detected five out of six cases of architectural distortion (Birdwell, 2001). However, Baker et al. found the sensitivity of two commercial CAD systems to be poor in detecting architectural distortion: fewer than 50% of the cases of architectural distortion were detected (Baker, 2003).

Banik et al. presented methods based in Gabor filters, phase portrait modeling, fractal analysis, and Haralick's texture features for architectural distortions detection. There exist a phase for detect candidates and for each ROI detected fractal dimension, and Haralick's texture features were computed (Banik, 2009). These authors introduced some improvements in this approach and also introduced a novel method for the analysis of the angular spread of power and Law's texture energy measures derived from geometrically transformed regions of interest. Thus, they reduce the number of false positive per image and increase the sensitivity (Banik, 2011).

3.5.4 Bilateral Asymmetry

Asymmetrical breasts could be reliable indicators of future breast disease in women, and this factor should be considered in a woman's risk profile. Asymmetries of concern are those that are changing or enlarging or are new, those that are palpable and those that are associated with other findings, such as microcalcifications or architectural distortion. Nishikawa et al. analyzed asymmetry densities to detect masses by performing a non-linear subtraction operation between the right and the left breast (Nishikawa, 1994). Katariya et al. assumed a conic shape from the using the CC view (Kataryia, 1974). The study by Kalbhen et al. has used both view and information recorded from breast compression using volumes of 32 breast specimens. They determined the most accurate method of calculating the volume of the breast. Assuming a half elliptic cylinder shape for the compressed breast for the CC view, the result reported better-measured volume than those assumed on the MLO view (Kalbhen, 1999).

Lau and Bischof used detection of structural asymmetries between corresponding regions in the left and right breast to detect breast tumors. They used a set of measures related to the mammographic appearance of tumors but without characteristics of specific tumors. They developed measures of brightness, roughness, brightness-to-roughness and directionality. The method was evaluated using a set of 10 pairs of mammographic images where structural asymmetry was a significant factor in the radiologist's diagnosis. These mammographic images contained a total of 13 suspicious areas. Final test results revealed sensitivity of 92% with 4.9 false positives per pair of mammographic images (Lau, 1991). Méndez et alsubtracted left and right mammographic images and applied a threshold to obtain a binary image with the information of suspicious areas. The suspicious regions or asymmetries were delimited by a region growing algorithm. The method was tested on a set of 70 pairs of digital mammographic images and obtained a true positive rate of 71% with 0.67 false positives per image (Méndez, 1998).

Stamatakis et al. proposed two methods for the comparison of left and right breast. The first method, Single Image Comparison (SIC), involved finding the corresponding areas whose intensities differed more than a preset threshold. The second method, Multiple Image Comparison (MIC), involved generating eight pairs of images from each original pair of left and right images. These eight pairs of images were then bilaterally compared, and the resulting images were recombined into one pair of images. A set of 10 features was calculated from the suspicious areas detected by the comparison process. Stepwise discriminant analysis was used to select 5 features for final feature set. The results were tested on a data set of 50 pairs of mammographic images. The SIC method obtained 86.8% accuracy with 4.9 false positives per image and MIC method obtained 89.2% accuracy with 4.3 false positives per image (Stamatakis, 1996).

Georgsson proposed two methods for differential analysis of bilateral mammographic images. The first method was based on the absolute difference between the registered images, and the second method was based on the statistical differences between properties of corresponding neighborhoods. He tested both methods on 100 mammographic images obtained from the MIAS database. The absolute difference method achieved a detection rate of 18 with 10 false positives, and statistical method achieved detection rate of 21 with 10 false positives (Georgsson, 2003).

Good et al. developed a method which performs feature-based analysis of local and global differences between mammographic images. The extracted features are assumed to be relatively invariant with respect to breast compression. After the registration of the images with a template, they applied a series of global and local feature filters to characterize the image properties of interest. Features are placed into feature vector and appropriate comparison of feature vectors reveals differences between image properties (Good, 2003). Miller and Astley used anatomical features for detection of breast asymmetry. Using texture analysis, they segmented mammographic images into regions of fat and nonfat tissue that presents glandular disc. Glandular discs in left and right breast were compared using properties of their shape and grey-level distribution. They tested the proposed method on a data set of 30 pairs of mammographic images and achieved classification accuracy of 86.7% with false positive rate of 16.7% (Miller, 1993). In another work, Miller and Astley compared glandular discs in left and right breast using shape, brightness and topology features. Measurements of shape were used to detect architectural distortion of the glandular disc, and brightness distribution was used to detect masses and focal densities and topological features specified other suspicious asymmetries. The method was tested on the data set of 104 pairs of mammographic images and obtained classification accuracy of 74% (Miller, 1994).

Tahmoush and Samet proposed the use of image similarity to determine asymmetry. They filtered the mammographic images to find the contextually similar suspicious points that could be malignant. Their similarity or asymmetry analysis used comparison of the values of the sets

of suspicious points using modeling and supervised learning of model parameters. They used Bayes' Theorem to incorporate asymmetry into CAD system. Their method properly classified 84% of images in their test set. The method correctly classified 97% of cancerous cases and 42% of non-cancerous cases (Tahmoush, 2006). Ferrari et al. developed a method for the asymmetry analysis based on the detection of linear directional components using a multiresolution representation based on Gabor wavelets. They segmented the fibroglandular disc and decomposed the resulting image using a bank of Gabor filters. The Gabor filter responses for different scales and orientation were analyzed using Karhunen-Loève (KL) transform and Otsu's thresholding method. For quantitative and qualitative analysis of the oriented patterns, the rose diagrams were used. Their method obtained classification accuracy of up to 74.4% (Ferrari, 2001). Rangayyan et al. extended the method of Ferrari et al. by including morphological measures and geometric moments of the fibroglandular disc related to density distributions. Directional features were obtained from the difference in rose diagrams and additional set of features, including Hu's moments, eccentricity, stretch, area and average density were extracted from the segmented fibroglandular discs. The differences between the pairs of the features for the left and right mammographic image were used as measures for the analysis of asymmetry. They obtained the overall classification accuracy of 84.4% (Rangayyan, 2007).

3.6 Summary

Computer-aided diagnosis has become a part of clinical work because it is a useful tool for diagnostic examinations but is still in the infancy of its full potential for applications to many different types of lesions obtained with various modalities. CAD is a concept based on the equal roles of physician and computer, and thus is distinctly different from automated computer diagnosis.

The CAD evaluation tools are based on their values of false positives and negatives and true positive and negatives, and thus on the sensitivity and specificity. CAD can help to decrease errors of detection of cancers, notably those that manifested, as microcalcifications in mammography.

CAD systems have a typical design that includes pre-processing, detection, classification and evaluation. Processing stage aiming at normalizes image size and intensity and reduces image noise in order to allow more robust systems to acquisition changes. It is in this stage that it is defined the region of interest to be processed in subsequent stages. The detection stage, introduced for enhancement the lesions in order to facilitate the detection. The output of this stage is first indication of lesion location, most likely with many false positives. Finally, classification stage aims at reducing the number of false positives by measuring local features at

the detected locations and uses those features for classification in lesion and non-lesion locations. An evaluation about the performance of CAD system allows the comparison of results with ground-truth.

During years many researchers dedicated themselves to study intensively breast cancer lesions. Some of them invested in work in only one or two stages, and others created a complete CAD system giving one more advance in this area. Microcalcifications and masses are the most common lesions of the interest for researchers may be because the challenge exists in this problem. However, there exists some literature related with architectural distortions and bilateral asymmetries. Hence, the development of new breast cancer computer-aided detection is an active research field, particularly regarding the detection of subtle abnormalities in mammograms.

CHAPTER IV – Methods and Implementations

4.1 – Distinction between left and right breast

4.2 – Distinction between CC and MLO projection

4.3 – Segmentation and extraction of pectoral muscle

4.4 – Features extraction

4.5 – Classification of breast density

4.6 – Classification of masses

4.7 – Summary

4. Methods and Implementations

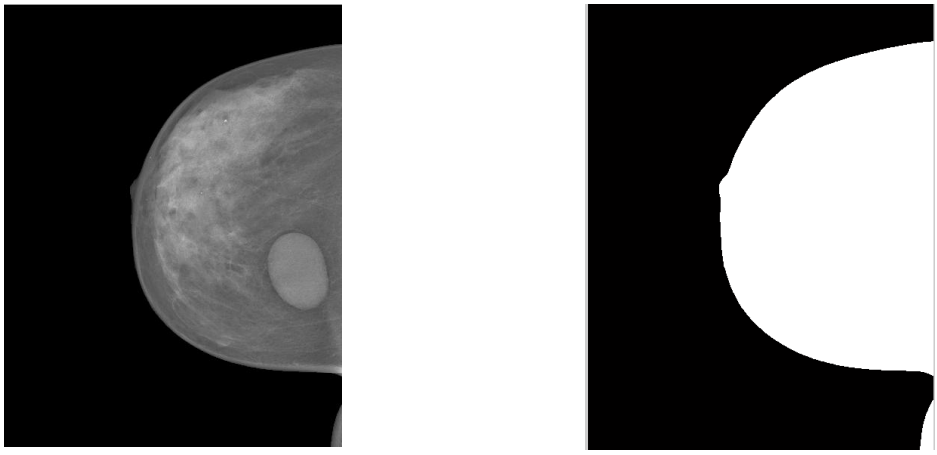
This chapter presents a description about the different methodos implemented, as well as exemplifications of their results. The techniques have been implemented in Matlab® (R2012b) in a computer with CPU T2390 with 1.86 GHz and 2GB RAM memory.

All the mammographic images correspond to real cases obtained from mini MIAS Database (Suckling, 1994), which contains MLO views of both left and right breasts. The images in the database were digitized at a resolution of 50 μm *per* pixel with 1024x1024 pixel size at 256 gray levels. The images are in the grayscale file format (.pgm). The other image database is INBreast which contains MLO and CC views of left and right breasts. The image matrix was 3328x4084 or 2560x3328 pixels, depending on the compression plate used in the acquisition, and the images were saved in the DICOM format, with pixel size of 70 μm and 14-bit contrast resolution. Once in this work pretend do the classification of breast density it is important to say that MIAS database just have three categories to classify density, on the other hand the INBreast database use the most common classification that have four categories, as seen previously in chapter II.

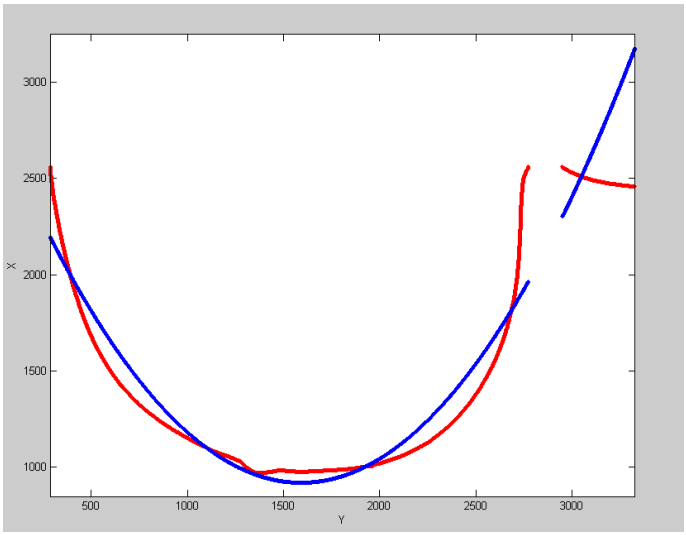
4.1 Distinction between left and right breast

The first step of this work consists in doing the distinction between left and right breasts, because it is important do a diagnosis fast that allow an early detection of any malignancy in order to avoid mastectomy, reduce the probability to return, and decrease the rate of mortally. So, the initial work developed was to create a mechanism that was automatic from the simplest procedures to the most complex. The goal of this tool is to avoid errors in the image observation during the diagnosis, preventing the use of unnecessary therapies.

Therefore, it is proposed an algorithm that by isolation of the border of the breast allows obtaining an approximation to a parabola, Figure 13, to determine if is a left or a right breast.



(a) (b)



(c)

Figure 13 – (a) Original Image; (b) Isolation of the border of the breast; (c) Approximation of breast edge to a parabola.

Firstly the image was read and submitted to a median filter, smoothing it and preventing the loss of details that could be important for the following steps. Another way to smooth the image and also tested consisted of using a Gaussian filter in order to verify if could influence the final results.

After that the image is binarized, so the next step can easily extract the contour through the variations in intensity. To obtain the contour of the image is applied the *edge* function of Matlab using the *Canny* operator.

In order to distinguish between the left and right breast it was used a linear iterator with the direction of the x-axis, which ran all the lines of the output image and save the coordinates(x,y) of the pixels belonging to the border of the breast in a vector. Through *polifit* function it was performed a quadratic regression which returned the parabola corresponding to the entered

values. From the results obtained it was possible to distinguish the left and right breast through of quadratic index of the function$y = Ax^2 + Bx + C$. In the right cases this index must be positive otherwise it is a left breast.

4.2 Distinction between CC and MLO projection

To determine the projection that is associated with the image analysis were developed and tested several methods among which stood out for its high accuracy rate that will be presented in this section.

The method chosen to distinguish between projections uses part of the algorithm developed to distinguish the left and right breasts presented in the previous section. The starting point is the image with the contour obtained by applying the *Canny* operator. The image is reduced approximately to half with the goal of draw a line between the extremes of the contour, Figure 14. According to the research done it is known that the slope of a line belonging to the MLO projection is more pronounced than in the CC projection. Thus, it is possible to distinguish the projections taking into account on the absolute value of the slope of the line. A straight line was traced between the start and end points of the contour of the breast for a left breast this line must have negative slope, and for right breast a positive slope.

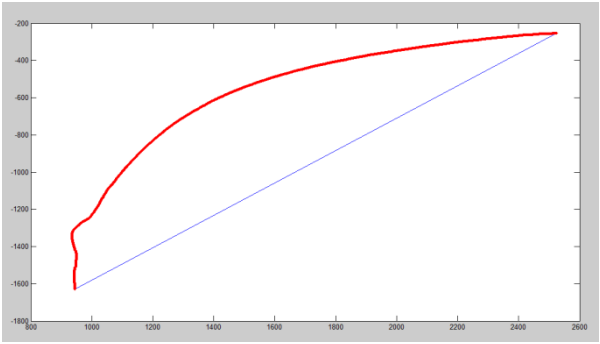


Figure 14 – Contour of the image and a straight line between the extremes of the contour.

4.3 Segmentation and extraction of pectoral muscle

In a mammographic image there are structures which are not part of the breast as *background*, identification labels, and pectoral muscle. The segmentation of pectoral muscle is very useful in many areas of mammographic analysis, once this muscle presents characteristics

of textures similar to mammary parenchyma, that means the results of the imaging processing methods cannot be the corrects in automatic detection of cancer.

There are several processes which depend of the location and segmentation of pectoral muscle:

- Evaluation of the position of the breast during the exam (Naylor, 1999);
- Location of clusters of microcalcifications (Pimentel, 2004);
- Combination of information's from pairs of mammograms (Yam, 2001)
- Determination of density patterns of the mammary tissue (Karssemeijer, 1998);
- Suppression of pectoral muscle for not interfere in methods of mammograms processing(Karssemeijer, 1998;Hatanaka, 2001;Raba, 2005)

There are some works that approach the edge of the pectoral muscle by a straight line as a first step in the processing (Karssemeijer, 1998; Yam, 2001; Kwok, 2004; Ferrari, 2004). The region growing (Raba, 2005), the Hough transform (Karssemeijer, 1998; Yam, 2001; Ferrari, 2004) and the local adaptive thresholding are some different techniques that can determine this approach. Masek et al. applied an algorithm that using the calculation of the minimum cross-entropy threshold in areas around the pectoral muscle to determine several thresholds as a function of area size (Masek, 2001). In 2005, Bajger et al developed a method to identify the pectoral muscle based in minimum spanning trees and the active contour. This technique was applied in 84 images of MIAS data, and as result were obtained 1,64% of false-positives and 12,03% of false- negatives (Bajger, 2005).

As previously mentioned the technique of region growing is very used in this works. Raba et al. segmented the pectoral muscle applying this methodology and tested in 320 of these images were correctly segmented. The evaluation of the results was done by a radiologist (Raba, 2005).

The detection of this muscle is an important task to aid the diagnosis, mainly when is being developed an automatic system. The morphologic operators are not yet very tested in this field and never were used specifically to detect the pectoral muscle.

As seen previously the extraction of the pectoral muscle it is an important step to classify breast density because sometimes can induce wrongly the level of density. An algorithm to solve this problem was developed and the different steps to segment the pectoral muscle are presented in the following block diagram, Figure 15.

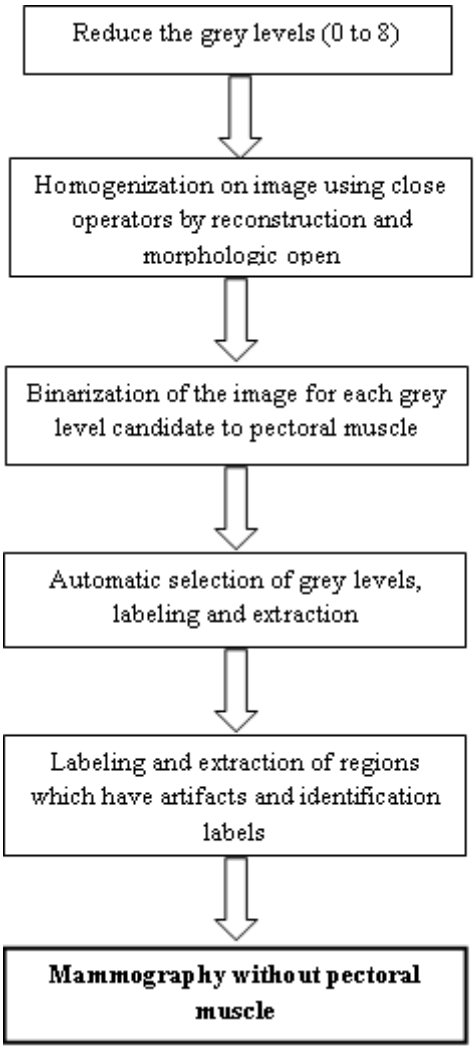
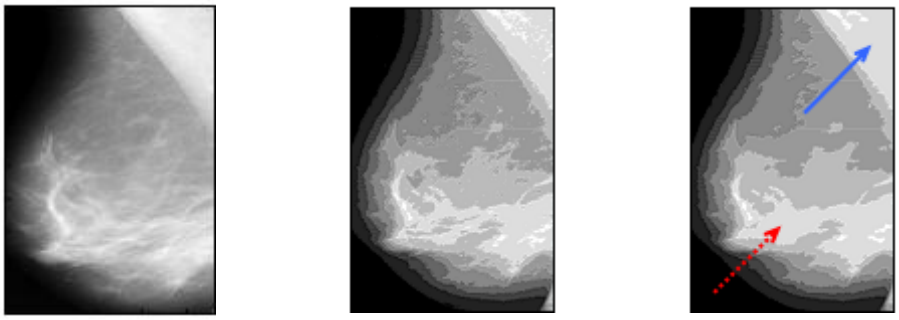


Figure 15 – Block diagram of the procedure to segment the pectoral muscle.

A preliminary study was done to investigate empirically the grey levels which are related to the edge of the pectoral muscle. To study the grey levels it was decided reduce the image to nine levels (0 to 8) in original image. So, in this way the breast present eight levels (1 to 8) and the zero was reserved to background. In general, the pectoral muscle can be related to levels 5 to 8, and the other grey levels are related to the breast edge (1, 2 and 3) and part of mammary tissue (4).

After reducing the grey levels, each level was binarized and for 5, 6, 7 and 8 it was applied the closing by reconstruction, to join regions of the same grey level and smooth the edge of the structures. Then the opening operator allowed eliminates small structures in image, as exemplified in Figure 16. For that it was used a structural element in the form of disc with a diameter of 20 pixels.



(a) (b) (c)

Figure 16 – (a)original image; (b) after reducing grey levels; (c) after closing by reconstruction(red arrow), following by morphologic opening (blue arrow).

The pre processing is important to eliminate some elements that are not connecting to the region of interest, so the next step consisted in use labels to solve this problem. In this algorithm just the entire breast interest, therefore using labels allow say that the labels which have more points is the breast and any other artifact is retired of the image.

The last step of this procedure is selecting the correct level or levels that allow exclude the pectoral muscle of the image. After a lot of experimental work it was perceptible that if were used the levels 5, 6, 7 and 8 the result of segmentation would be inadequate. Thus, the conclusion was that the most suitable levels to use for pectoral muscle extraction were 6, 7 and 8. However, after select these three levels can obtain just the pectoral muscle or other regions with the same grey levels but which are mammary tissue. Here was used the algorithm to identify if the image analyzed it was left or right because if it was a left image the algorithm chose the first label of the image, if not the algorithm chose the last label and delete all the points to the image. Note that this process is iterative in order to obtain a satisfactory contour. Every iteration will repeat the thresholding in eight levels, then the opening operator with a structural element of 20 and finally the extraction of the contour using labels.

Once are working with two data images, all images of MIAS data were the firsts tested with this algorithm, just because this data only have MLO projection. There is no ground truth elaborated by a specialist in order to compare the results. Also INbreast images were tested, but just the MLO images that correspond to 201 of 410 the images. In this case we have a ground truth done by a specialist. However, the ground truth correspond a XML file with a few points of the pectoral muscle. So, it was necessary implement a small code to join these points using splines and also the methodology to distinct between left and right breast. Thus it was created a new set of MLO images without pectoral muscle which later would serve to compare with the results obtained.

4.4 Features extraction

The definition of a set of features able to describe how density the tissue breast is, it is a challenging task for the development of a CAD system, especially if want to create a completely automated methodology. It is not very easy analyze visually the mammography and classify into one of the four types more usual of density, because sometimes it is a little questionable and therefore the automatic algorithm can help to decide what level of density is.

The texture of such region can help to distinct between different tissues and can be represented by statistical descriptors extracted from the histogram of the image intensities or the co-occurrence matrix, besides structural and spectral descriptors.

To obtain the best representation of the breast tissue density were combined various sets of features:

- Statistics texture descriptors from the histogram of the image intensities;
- Statistical texture descriptors from the co-occurrence matrix;
- Texture descriptors from the Fourier spectrum;
- Invariants Hu moments;
- PCA on the matrix image vectored;
- 2DPCA on the image array;
- 2DPCA on the co-occurrence matrix;

In respect of statistical descriptors were extracted nine. The average intensity of the image where the more dense tissue has the higher average intensity. Other measure used was the standard deviation that are related with the irregulariry of the texture. It was also important calculate the smoothness that is low to regular intensity, and the uniformity that is higher in soft textures. Also asymmetry, kurtosis, average histogram, modified standar deviation and modified asymmetry were calculated and extracted (Subashini,2010). The formulas for all of these measures are presented in Table 4 .

Table 4 – Equations of statistical descriptors from the histogram.

Average	$\mu = \sum_{i=0}^{L-1} z_i p(z_i)$
Standard deviation	$\mu_2 = \sum_{i=0}^{L-1} (z_i - m)^2 p(z_i)$
Smoothness	$R = 1 + \frac{1}{1 + (\frac{\mu^2}{L-1})^2}$
Uniformity	$U = \sum_{i=0}^{L-1} p^2(z_i)$
Asymmetry	$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i)$
Kurtosis	$\mu_4 = \sum_{i=0}^{L-1} (z_i - m)^4 p(z_i)$
Average histogram	$AH_g = \frac{1}{L} \sum_{i=0}^{N} (L-1)(i)$
Modified Standard Deviation	$\sigma_m = \sqrt{\sum_{ij} \left(X_{IJ} - \mu\right)^2 p\left(X_{ij}\right)}$
Modified Asymmetry	$\mu_{3_m} = \sqrt[3]{\sum_{ij} \left(X_{IJ} - \mu\right)^3 p\left(X_{ij}\right)}$

However, the descriptors based on the histogram do not contain information about the positioning of pixels in relation to other pixels. In order to preserve the spatial information of the pixel intensity, statistics from the co-occurrence matrix of intensity levels were extracted. The most common descriptors were used in this work, so correlation evaluates how a pixel is related to its neighbor and contrast evaluate the contrast of intensity in this relation. The other measures are maximum probability, uniformity, entropy and homogeneity (Woods, 2004). The equations of statistical descriptors from the co-occurrence matrix are shown in Table 5.

Table 5 – Equations of statistical descriptors from the co-occurrence matrix.

Correlation	$\sum_{i=1}^{K}\sum_{i=1}^{K}\frac{(m_i - r)(j - m_c)p_{ij}}{\sigma_r \sigma_c}$
Contrast	$\sum_{i=1}^{K}\sum_{i=1}^{K}(i-j)^2 p_{ij}$
Maximum probability	$\max(p_{ij})$
Uniformity	$\sum_{i=1}^{K}\sum_{i=1}^{K}{p^2}_{ij}$
Entropy	$\sum_{i=1}^{K}\sum_{i=1}^{K}p_{ij}log_2(p_{ij})$
Homogeneity	$\sum_{i=1}^{K}\sum_{i=1}^{K}\frac{p_{ij}}{1+\lvert ij\rvert}$

The spectral approach is mainly used to identify the global frequency of an image by identifying high peaks of energy spectrum and based on the properties of Fourier spectrum. The spectrum is very useful because it has three characteristics, the first relates to the fact that prominent peaks in the spectrum providing the direction of the texture patterns. Moreover the fundamental spatial period is provided by standards position of the peaks in the frequency plan. And lastly, when the periodic elements are eliminated through the process leaves the filtration elements in the periodic image, which can be described by statistical techniques.The spectrum is generally adapted to describe the direction of standards periodicals in an image (Woods, 2004).

In computer vision the moment of an image is a certain specific weighted average of the intensity of image pixels and are usually chosen because they contain some attractive properties. These can be applied to various aspects of image processing that can go from the invariant pattern recognition until the image encoding to represent an estimate.

Hu moments are invariant to translation, scale change and also rotated. Hu developed two different methods for describe the invariant moments (Hu, 1962). The first called main axis, which however failed when it was found that the images are not contained a single major axis such images are described as symmetric with respect to rotation. The second method is the absolute time invariant. Hu derived expressions of the algebraic invariants applied to the moment of the function generated from a rotation transformation. These groups consist of nonlinear expressions centered moments. The result is a set of orthogonal absolute moment's

invariants which can be used to identify patterns invariant to scale, rotation and position (Shutler, 2002). Based on normalized central moments in Table 6 are presented the seven moments invariants.

Table 6 – The seven invariant moments of Hu.

1st moment	$\emptyset_1 = \eta_{20} + \eta_{02}$
2nd moment	$\emptyset_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$
3rd moment	$\emptyset_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \mu_{03})^2$
4th moment	$\emptyset_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \mu_{03})^2$
5th moment	$\emptyset_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12)}^2 - (\eta_{21} + \eta_{03})^2]$
6th moment	$\emptyset_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$
7th moment	$\emptyset_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$

Usually tends to choose the attributes that best describe the properties of the components of the image, this is necessary because one of the major problems inherent in the process of identifying regions in images is the large dimensionality of the feature vector, because it compromises the performance and accuracy classifier. The reduction of the dimensionality consists in decrease the set of characteristics, this being possible because some features may not be relevant for the decision, and so it is possible to identify the minimum number of characteristics that is enough to describe a region of an object or identify the image (Duin, 2000).

There are two approaches to reduce the dimensionality which are the extraction and feature selection. The first creates new characteristics from combinations or changes in the original set

while the selection of characteristics is to determine a subset of characteristics that discriminate objects through previously established rules.

According to Kendall, the PCA can be used for various tasks of which highlights the simplification of the data set, which consists in finding a simplified representation of the universe of study, that may occur through the transformation being a linear combination or not a set of independent variables in another set of autonomous and inferior dimension (Kendall, 1980).

Differently, 2DPCA is a technique developed for image feature extraction based on 2D matrices, in order to solve the problem of transform the matrix of the images in one-dimensional vector which make difficult analyze the co-variance matrix. So, in this algorithm it is more easier evaluate the co-variance matrix and consequently it is necessary less time of processing to determine the eigenvectors (Yang, 2004).

4.5 Classification

4.5.1 Breast Density

As explained earlier, the diagnostic step is performed from the sets of characteristics extracted previously, which is submitted to the evaluation of a classifier, which, from a previous training, tells, in this case, what levels of density the breast has. After selected the main features the following phase it is classification, but first it was necessary to be added to the array with the main characteristics the respective value of class to which image belongs.

The support vector machines (SVMs) have been successfully applied in various areas of research. The basic idea of SVM is that it projects data points from a given two-class training set in a higher dimensional space and attempts to find a maximum-margin separating hyper plane between the data points of these two classes. The training data for SVMs should be represented as labeled vectors in a high dimensional space where each vector is a set of features that describes one case. This representation is constructed to preserve as much information as possible about features needed for the correct classification of samples (Taylor, 2000).

To ensure the accuracy of the results there are two ways of choose what set is for training and for test. In the simple validation, the training samples labeled are randomly divided into two parts, one is used as training set to adjust the parameters of the classifier and the other validation set is used to estimate the generalization error. The cross validation method used in this work was the K-fold in which the model is trained with all subsets except for one that will be used as the validation set. The process is repeated K times, each time with a different subset as the validation set (Haykin, 1999).

The SVM can be described for a binary classification and the hyper plane generated is determined by a subset from the two classes, called support vectors. However, sometimes the data set cannot be precisely separated by a hyper plane, so a function called kernel is used instead. In this case the main goal is classify the breast density that according ACR has four categories, but in case of MIAS database only exists 3 categories, so it is a multi-class problem. There are two basic approaches about this kind of problem, the first divides into types one called one against all reduces the problem of multiple classes to a set of binary problems, and the other separate classes two by two, it is called one against one. The second approach is a generalization of the binary classification to more than two classes (Crammer, 2000; Hsu, 2002).

Relatively to this work the SVM approach was based on the implementation provided by the LibSVM toolkit using a C-SVC SVM model and Radial Basis Functions (RBF) where the parameters C and gamma were individually calibrated for different combination of features tested using a grid search scheme, minimizing the classification error. Generally, the radial kernel is a reasonable choice when the relation between class labels and attributes is nonlinear. It is possible do a grid-search using a function of LibSVM where all pairs (C,γ) are tried and the one with the best cross-validation accuracy is selected. The disadvantage of that it is the high computational cost. The classifier's training/testing processes was performed using a 10-fold cross validation that allowed obtaining statistically meaningful results (Chang, 2011).

The mammograms used to classify have different labels of density, as referred previously the MIAS database is divided into three classes and the INbreast in the four BI-RADS standard classification. Table 7 shows the distribution of the instances used in this work.

Table 7 – Distribution of density classes.

Class	MIAS	INBreast
1	104	136
2	104	146
3	112	99
4		29

4.5.2 Masses

According to BI-RADS, a mass is defined as a three-dimensional structure demonstrating convex outward borders, usually evident on two orthogonal views. The presence of a mass is the most important mammographic sign of cancer. An initial study on the detection of masses was started, since this computed aided system can also include the detection of lesions and identification of same. In this case, were used the images of the base INBreast that has 65 normal images and 108 with masses, but only 27 images include exclusively the masses. In the developed algorithm the image is transformed in double format and after several experiments was found a value that allows getting only part of the image where the mass is, Figure 17.

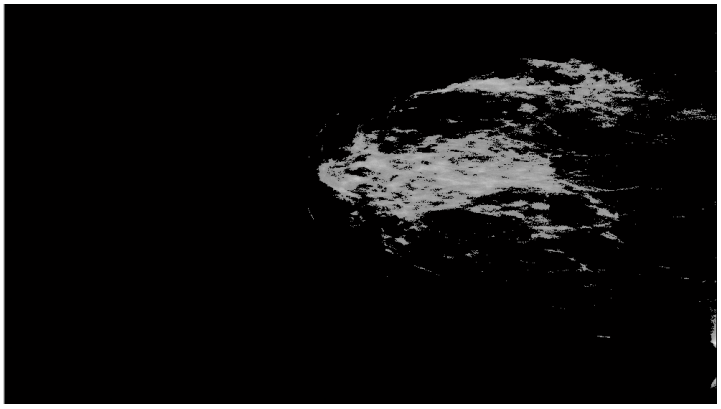
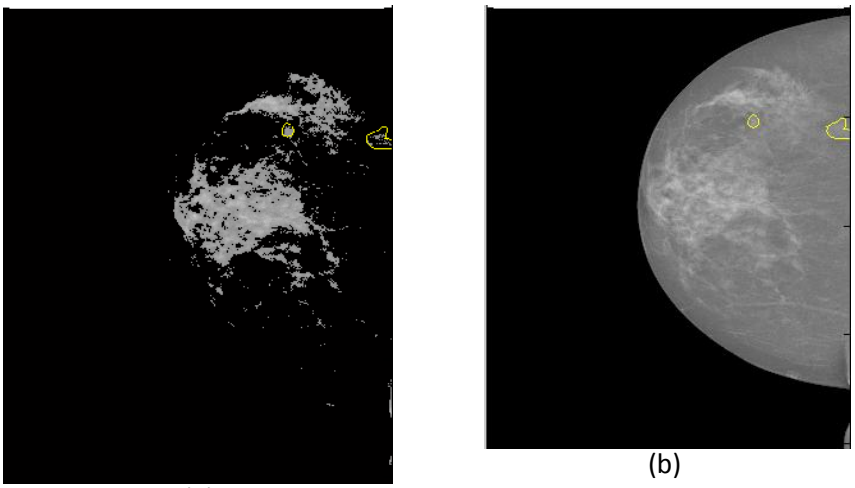


Figure 17 – Breast image after reducing for find the mass.

In INBreast documentation there is also a .XML file with the points of the lesions in every image. As can be seen in Figure 18 (a) the part of the image obtained contain the masses, sometimes not all points are in the selected area but help the technician or doctor because captures their attention to a small region of the image.



(a) (b)

Figure 18 – (a) Altered image with the masses; (b) Original image with the masses.

For now it is important to say that this method is very simplest but it is a beginning for promising results in the detection of masses and other lesions. However while this CAD system do not have a more complex technique, this can be a good help.

4.6 Summary

This chapter introduced some new techniques for the breast density classification. It explains detailed each step taken to achieve the distinction between directions and projections, the pectoral muscle segmentation, the features extraction and finally the classification.

Firstly, it shows the type of images used in this work and the classification in right or left breast and then the classification in CC or MLO projection.

The pectoral muscle segmentation and extraction is an important step of all work because the correct breast density classification can depends of that. Having done that, the following steps consist in extract and select the features to classify density.

In the end, to solve a multi class problem it was used SVM where training/testing processes was performed using a 10-fold cross validation that allowed obtaining statistically meaningful results.

CHAPTER V – Results and Discussion

5.1 – Distinction between left and right breast

5.2 – Distinction between CC and MLO projection

5.3 – Segmentation and extraction of pectoral muscle

5.4 – Classification

5.5 – Summary

5. Results and Discussion

In this chapter are presented the results of all the steps explained in the methodology. It is shown some of the best and worst results performed by the algorithm. The main objectives are classify correctly the largest possible number of images, according to density. But first it is important to distinguish between the direction and the projection of the breast and then remove the pectoral muscle that can influence the results. Finally, one can distinguish normal breast and one with injuries, specifically masses.

5.1 Distinction between left and right breast

For all the images of MIAS and INBreast data base was applied the algorithm based in quadratic regression referred in the previous chapter in order to evaluate their performance. In this work will be presented four images that represent the two projections in each breast, Figure 19.

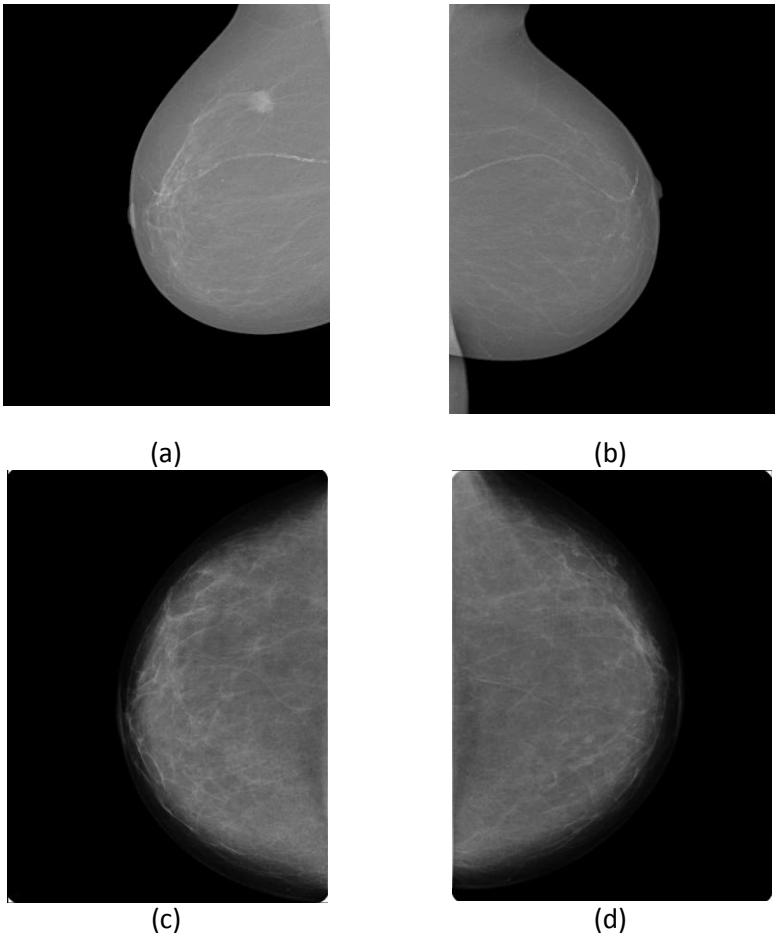


Figure 19 – Original images.

As mentioned in methodology the first step consists in smooth the original image using a median filter or a Gaussian filter. The last one present better result and the images obtained for every projection are presented in Figure 20.

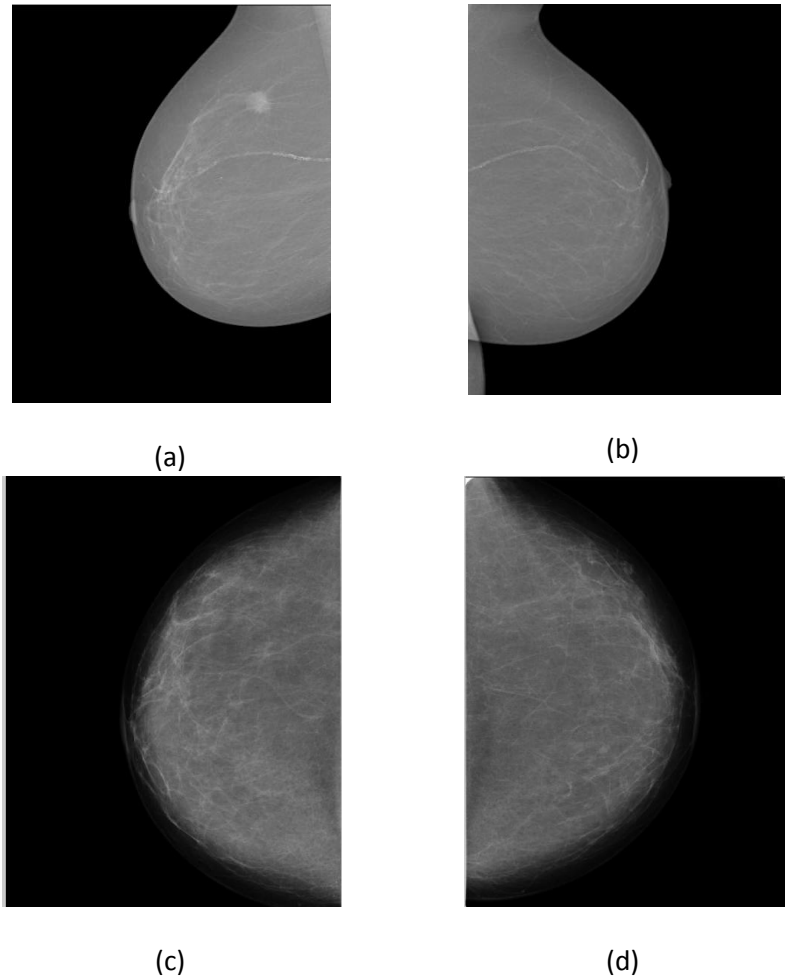


Figure 20 – Application of Gauss Filter to the originals images of Figure 19.

As can be verified the changes for these four images are not very clearly, once the images are of good quality there is practically no noise, furthermore the labels / tags that are usually present in the images of the MIAS database here do not exist. But, in certain cases the application of this linear operator can be very useful and make a difference in the final result.

The extraction of the contours it is an important step, so the images reproduced by the application of edge filter are reflected in the images of Figure 21.

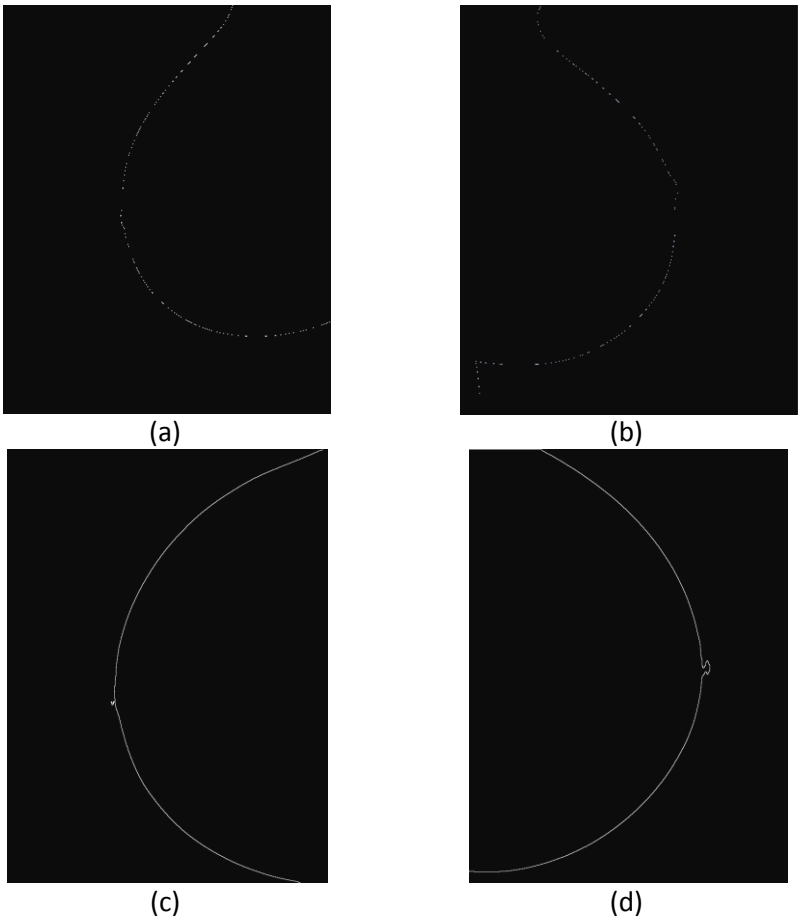


Figure 21 – Application of Canny methods to the filtered images of Figure 19.

Considering that the resultant images of this operation do not have noise and the contours were extracted correctly can be applied the next step. After all that the main decision will be made from a quadratic regression that defines if it is a left or right breast according to the concavity of the parabola and the contour of the breast obtained in the previous step. Using the contour and applying the quadratic regression are obtained the images of the Figure 22.

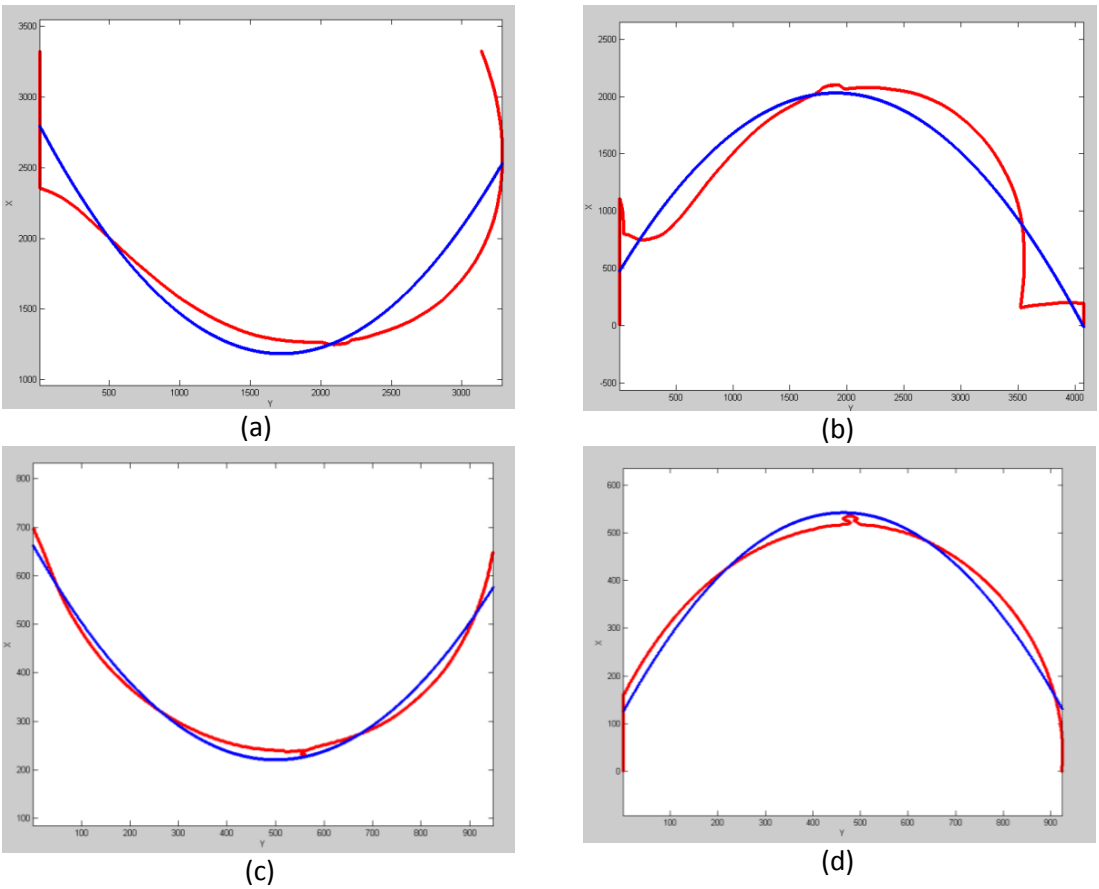


Figure 22 – Graphic representation of quadratic regression and of the breast contour.

As can be observed in Figure 18, the red line represents the parabola with the points of the vector of quadratic regression and the blue line represents just one parabola. Through a research previously done about the functioning of the quadratic regression concluded that the distinction between the two breasts is given by the value correspondent to the quadratic index of the function (x^2). Thus, when this index is superior to zero it is a right breast, otherwise it is a left breast. The values of this index to these four images are presented in Table 8.

Table 8 – Index of x^2 values of the quadratic regression.

Imagem a) da Figura 19	Imagem b) da Figura 19	Imagem c) da Figura 19	Imagem d) da Figura 19
Mama direita MLO	Mama esquerda MLO	Mama direita CC	Mama esquerda CC
5.4804e-04	-4.3209e-04	0.0018	-0.0019

Accordingly these four images where the algorithm works correctly the goal were achieved with success. However, when tested for the three hundred and twenty two images of MIAS database using a Gaussian Filter the accuracy rate is 85% corresponding to two hundred and seventy five correct decisions. For all four hundred and ten images of INBreast database the algorithm classify correctly the right and left breasts, that means an accuracy rate of 100%.

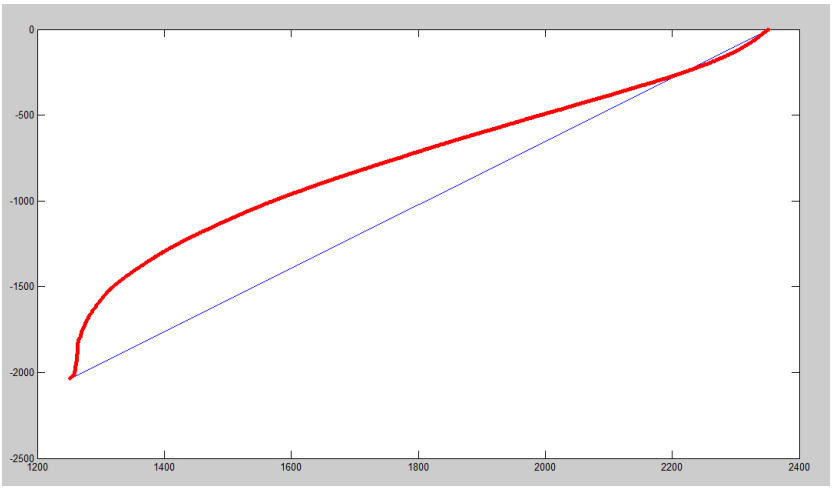
Until now the investigations are mainly related with the analysis of mammograms, especially with the segmentation of breast calcifications and not with the incidence and projection of the

image of the breast. This algorithm has low computational cost would be higher in the case of the median filter present better results, but the Gaussian filter helps reduce the runtime. Note that this process is independent of the projection, but as we will see below this is not true in distinguishing projections.

5.2 Distinction between CC and MLO projection

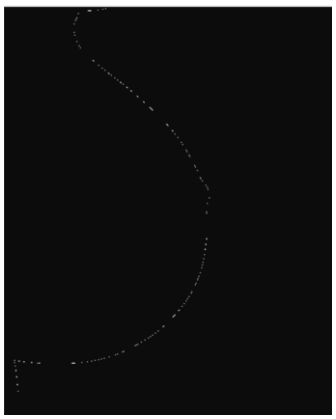
The technique developed to distinct the two most commonly used projections is dependent of the results obtained in the distinction made previously. So, will be used the same images that utilized will use the same images that are used in the algorithm to distinguish between left breast and right breast, Figures 23, 24, 25, 26.

(a) (b)



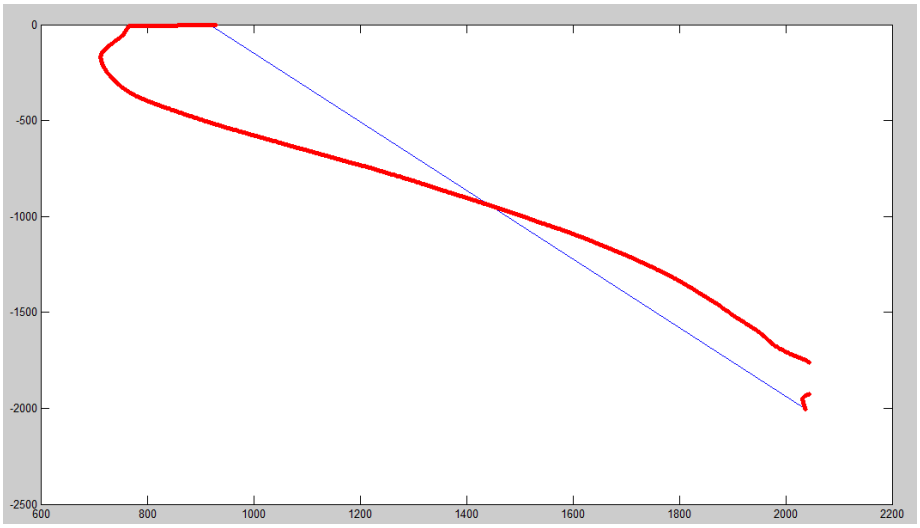
(c)

Figure 23 – (a) Image of the contour of the right breast image in MLO projection; (b) Superior half of the image; (c) Representation of superior half of the contour and straight line obtained.



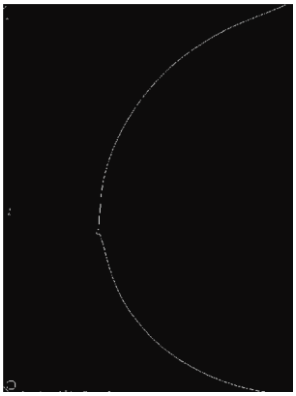
(a)

(b)

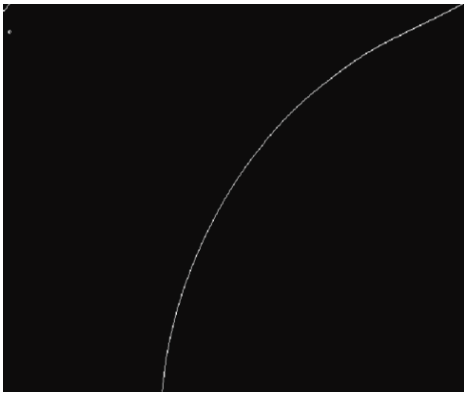


(c)

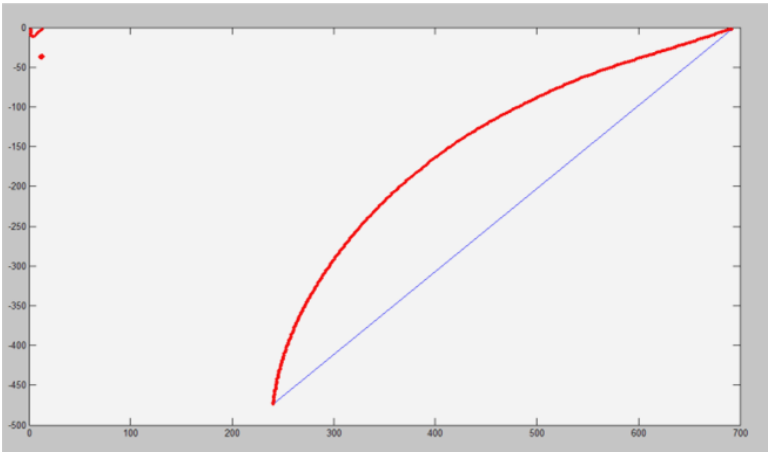
Figure 24 – (a) Image of the contour of the left breast image in MLO projection; (b) Superior half of the image; (c) Representation of superior half of the contour and straight line obtained.



(a)



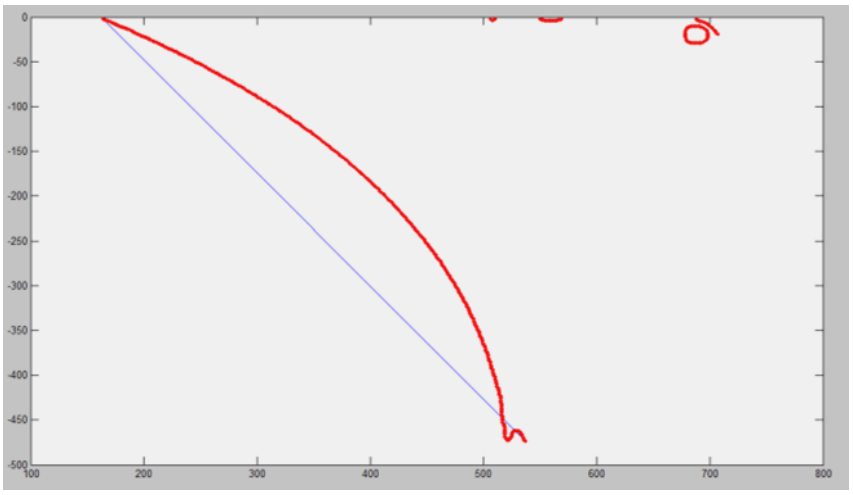
(b)



(c)

Figure 25 – (a) Image of the contour of the right breast image in CC projection; (b) Superior half of the image; (c) Representation of superior half of the contour and straight line obtained.

(a) (b)



(c)

Figure 26 – (a) Image of the contour of the left breast image in CC projection; (b) Superior half of the image; (c) Representation of superior half of the contour and straight line obtained.

In Table 9, it is clear that the technique worked correctly hitting the projection of the image in analysis. As said in methodology the slope of MLO projection is a little bit higher than in CC projection and for right breasts the slope is positive otherwise the slope is negative.

Table 9 – Slopes of the straight lines drawn over the outline and results of projections.

Images	Image a) of the Figure 15	Image b) of the Figure 15	Image c) of the Figure 15	Image d) of the Figure 15
Values of the slopes	1.8492	-1.7927	1.0442	-1.2654
Final result	Right breast-MLO projection	Left breast-MLO projection	Right breast-CC projection	Left breast-CC projection

Once this technique is dependent of the result of the distinction between left and right breast the success rate is compromised. Thus, can say that for the MIAS images have 100% of accuracy rate but only if the distinction between left and right breast were made correctly. So, as the hit rate is just 79% for the methodology of right or left breast this methodology has the same. It is important remember that MIAS database only have images of MLO projections, hence for a more reliable result the 410 images of INBreast base were tested, 203 had CC projection and obtained an accuracy rate of 95%. Although theoretically the values of CC projections are close to the expected, the values of the images MLO have fallen short of expectations it should be much higher than the CC.

5.3 Segmentation and extraction of pectoral muscle

Once the work was done take into account two databases of images, in this section firstly explain the segmentation and validation of pectoral muscle of MIAS images. Note that only MLO projections have pectoral muscle, so all the 322 images were submitted to this methodology.

In the following images of Figure 27 are presented all steps of this algorithm in a mammography of left breast, which are explained in the methodology. Also an example of segmentation and extraction of pectoral muscle in a right breast is presents in Figure 28.

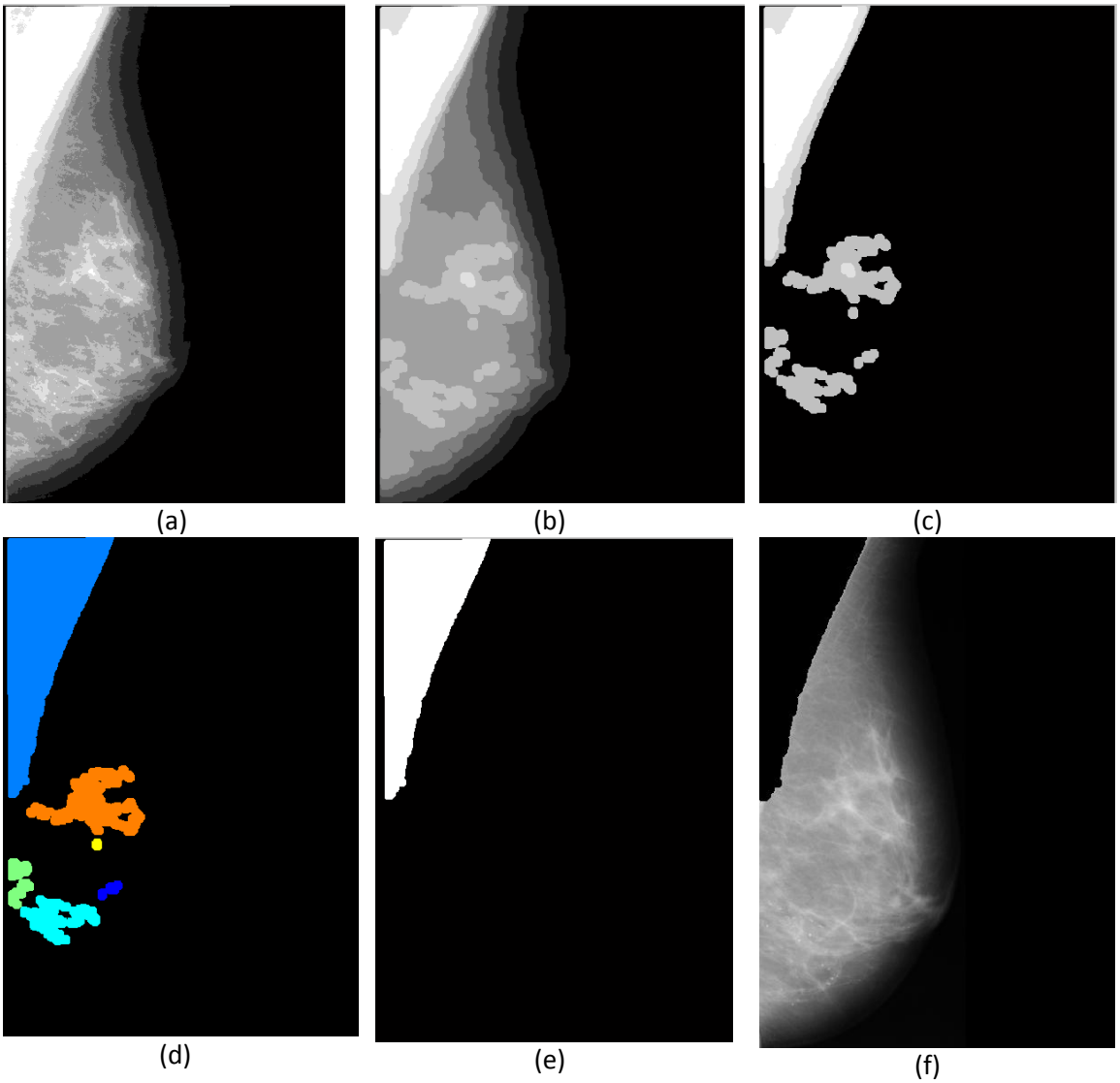


Figure 27 – (a) Original image; (b) Reduction to 9 gray levels; (c) Visualization of levels 4,5 and 6; (d) Labeling of regions; (e) Pectoral muscle; (f) Original image without pectoral muscle.

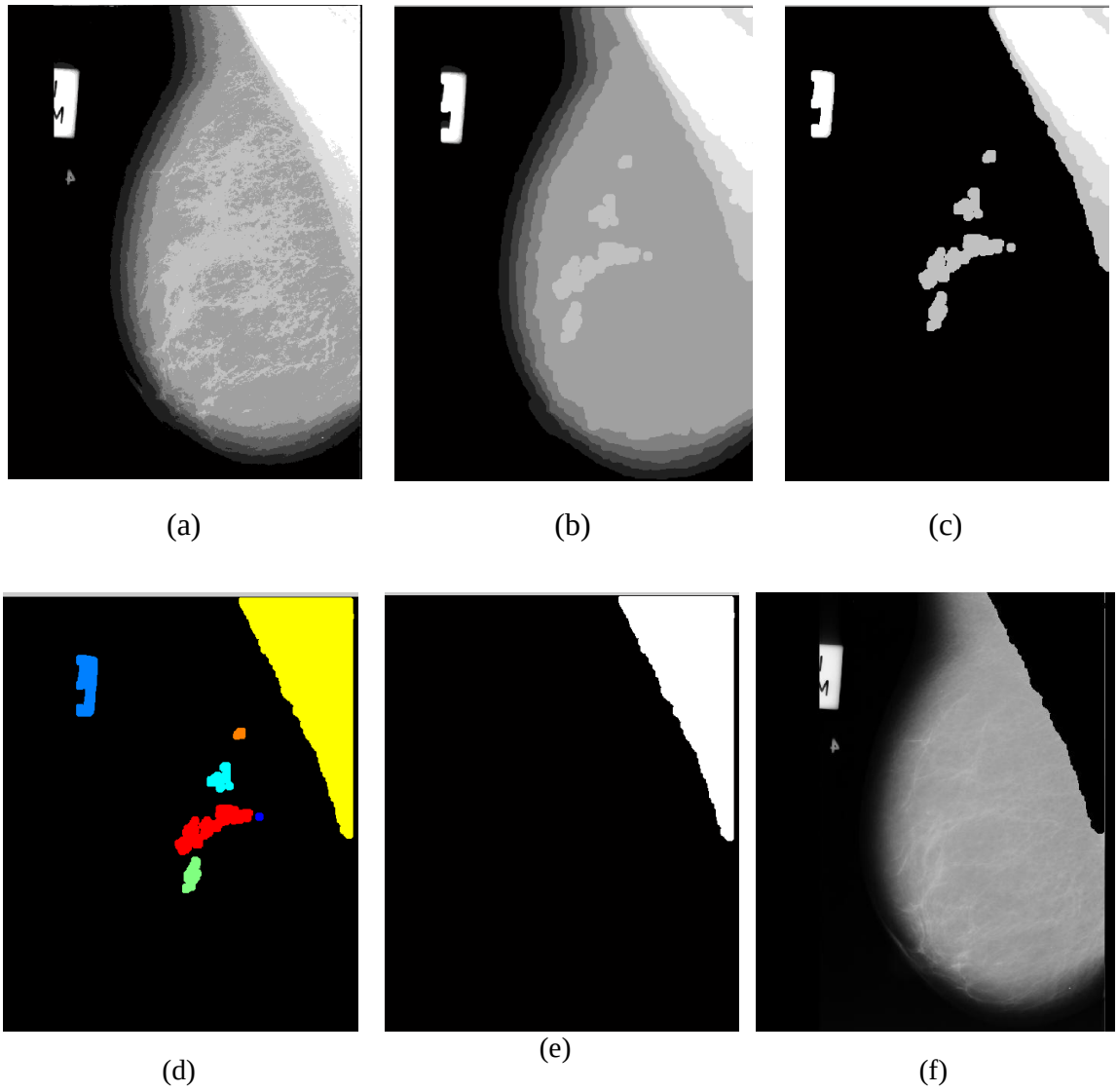
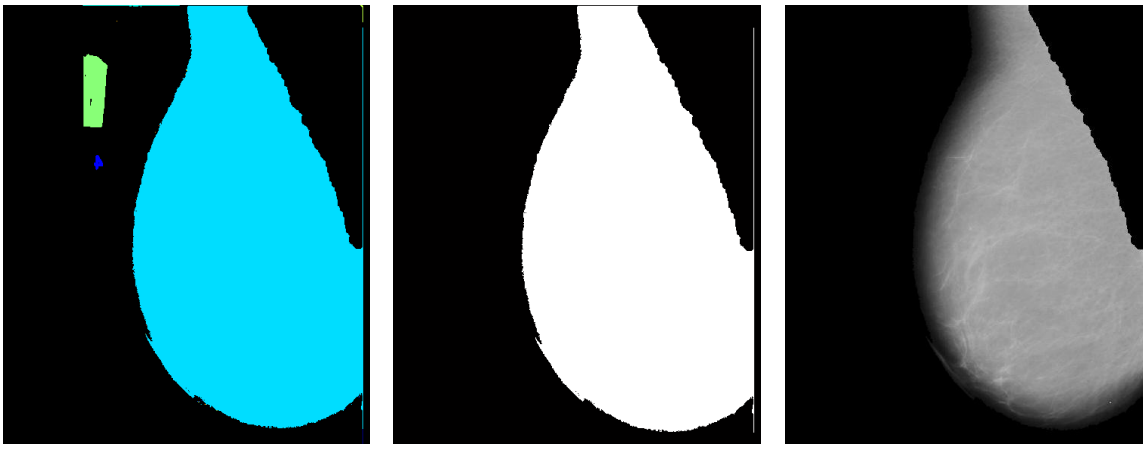


Figure 28 – (a) Original image; (b) Reduction to 9 gray levels; (c) Visualization of levels 4,5 and 6; (d) Labeling of regions; (e) Pectoral muscle; (f) Original image without pectoral muscle.

The last image shows that it is necessary eliminate some artifacts that can interfere in the breast density classification. As was used in previous steps, the use of labels allows eliminate those artifacts and Figure 29 presents the final result.



(a) (b) (c)

Figure 29 – (a) Labeling of regions; (b) Binarized image; (c) Original image without artifacts.

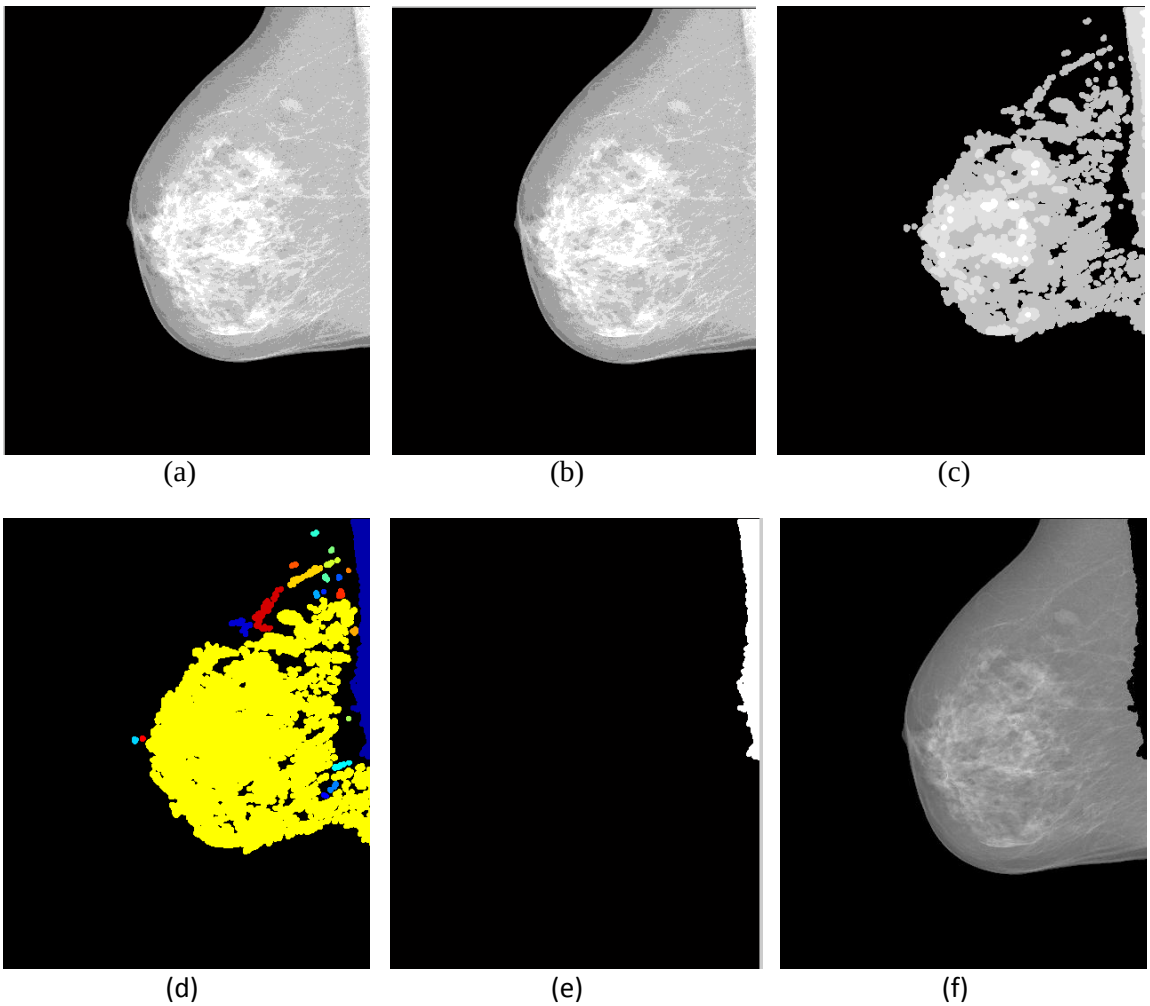
Once the database do not have a ground truth for the extraction of pectoral muscle and according to a radiologist technician that evaluate the images the accuracy is approximately 76%. The criteria of Doctor Sandra Ventura was to consider that the algorithm failed when there were at least 0.5cm or more muscle tissue still present in mammogram. The teacher thinks that the images are technically poor executed (for bad positioning, poor contrast and the presence of artifacts) may be condition the effectiveness of the algorithm. The advice given was to choose a set of images well executed because these factors will be relevant in measuring breast density, which is one of the goals of this work.

Hereupon, were used the images of INBreast database which had MLO projection that correspond to 201 images. This database has a specific fold about pectoral muscle where are files with XML and ROI format. Every image has a specific file associated that has multiple annotations including a list of contour points of the pectoral muscle. The annotations were made by a specialist in the field, and validated by a second specialist. Since only have some points of the contour of the pectoral muscle it was necessary create a little algorithm and use splines to connect all points, the final result can be seen in Figure 30.

(a) (b)

Figure 30 – (a) Segmentation of pectoral muscle using splines; (b) Extraction of pectoral muscle.

The 201 images in MLO projection of INBreast were tested with the algorithm of segmentation and extraction of the muscle. An example of the result obtained can be seen in Figure 31.

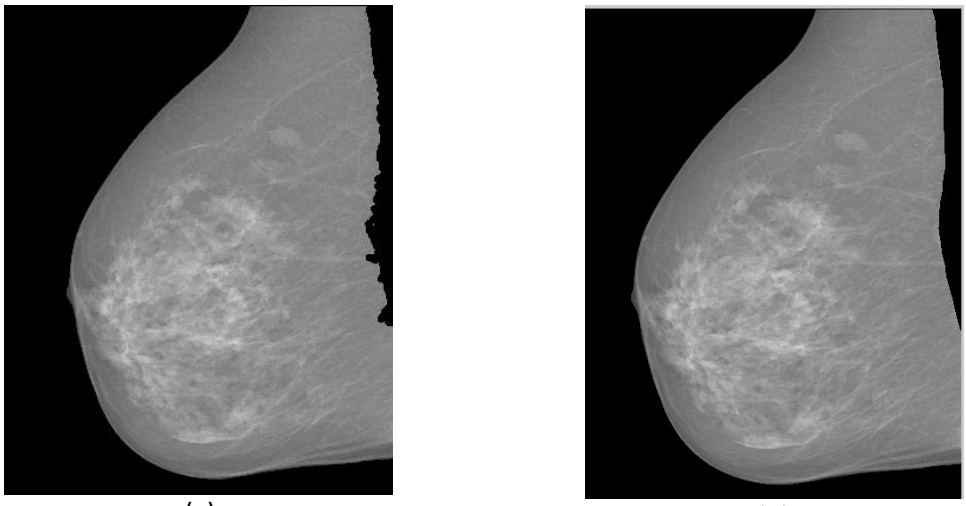


(a) (b) (c)

(d) (e) (f)

Figure 31 – (a) Original image; (b) Reduction to 9 gray levels; (c) Visualization of levels 4,5 and 6; (d) Labeling of regions; (e) Pectoral muscle; (f) Original image without pectoral muscle.

When compared with the ground truth, Figure 32, the success rate is 79%.



(a) (b)

Figure 32 – Comparison of pectoral muscle extraction: (a) Result of application of the algorithm; (b) Ground truth.

By observing the results it can be said that the most notable difference between the two images is related to the delineation of a smooth contour, this happens because the morphological operators are used with high structural element. Even if the final image was resubmitted to morphological operators to try to smooth the part where was extracted the pectoral muscle the image would not be significant changes in this aspect. Note that although the success rate is not high image resulting from applying algorithm developed follows the direction of the curve of ground truth, which shows a good performance.

The segmentation of the pectoral muscle is useful in image processing mammography, since this muscle usually presents as a dense region incidences prevalent in medio lateral (MLO) mammograms may, therefore, affect the performance of methods for automatic detection of lesions.

The worst performance was observed in images where the edge of the pectoral muscle is presented poorly defined causing the density of this to be confused with the breast tissue.

Regarding the automatic selection of the gray level, this may not be the more appropriate due to variations of the gray level in the pectoral muscle. These variations arise from factors such as improper positioning of the breast, the presence of axillary fold and pectoral muscle minor. In these cases the error appears to be related to problems when the mammograms acquisition than actually processing. This algorithm should be part of a CAD system for mammography.

5.4 Classification

5.4.1 Breast Density

In the previous chapter were indicated all the characteristics used to classify breast density. All features extracted where normalized between -1 and 1, using the following equation:

$$y = \frac{(ymax - ymin) * (x - xmin)}{(xmax - xmin) + ymin}$$

Where *ymin* and *ymax* stand for the desired range for the new values and *xmin* and *xmax* stand for the smallest and largest feature value that is normalized, respectively.

As said previously, the classification was performed using 10-fold cross-validation scheme. Table 10 and Table 11 show the results for some combinations of features when the images have pectoral muscle and when do not have. The first table is related with MIAS database and the second with INBreast.

Table 10 – Accuracy results for MIAS database.

Features	MIAS			
	Accuracy with Pectoral Muscle and Labels		Accuracy without Pectoral Muscle and Labels	
	Cross Validation (%)	All images for training and all for test(%)	Cross Validation (%)	All images for training and all for test(%)
Statistical	67.4	83.2	72,7	74,5
Co-occurrence matrix	63.1	97.5	73.9	84,8
Statistical + Co-occurrence matrix	69.3	99.4	75,5	85,7
Spectrum	60.3	78.9	61,8	80,7
PCA+2DPCA (5 components)	52.5	77	50	77,9
Invariants moments of Hu	54.4	60.9	62.2	67,4
Statistical + Co-occurrence matrix + Invariants moments of Hu	67	86	74,8	86.5
Statistical + Co-occurrence matrix + Spectrum	69.6	96	65,5	97,5

The classification accuracy of the presented technique discriminating between 3 categories of density based on the ground truth. This measure is calculated as the percentage of correctly classified mammograms in a breast parenchymal density category over the ground truth total number of mammograms in that category. The results show and prove what was expected, namely that when the pectoral muscle is extracted as well as other artifacts existents the success rate increases. Would think that the more features it is used much higher would be the success rate, but taking into account the results it was ascertained that the statistical descriptors and the co-occurrence matrix together gets the best results with 75.5 in the cross-validation and 85.7 when all images are used for training and testing. In some cases, when it is used all images for training and for test the accuracy decreases when the pectoral muscle is not present. This phenomenon it is difficult to explain because in theory the percentage in this case should be 100%, after all being the same images used to train and classify.

Table 11 – Accuracy results for IN Breast database.

Features	INBreast			
	Accuracy with Pectoral Muscle and Labels		Accuracy without Pectoral Muscle and Labels	
	Cross Validation (%)	All images for training and all for test(%)	Cross Validation (%)	All images for training and all for test(%)
Statistical	61.2	76.6	75.6	83.9
Co-occurrence matrix	71	78.5	77.4	80.2
Statistical + Co-occurrence matrix	78.3	81.5	80.7	91.5
Spectrum	64	100	71.3	100
PCA+2DPCA (5 components)	63.2	70.4	65	70.8
Invariants moments of Hu	49.5	87.6	65.1	93.5
Statistical + Co-occurrence matrix + Invariants moments of Hu	70	86.3	71.9	88.3
Statistical + Co-occurrence matrix + Spectrum	66.6	88.3	77.1	90.3

In case of this base image it is important to note that the set of images is a bit higher than the previous base, in addition the classification of these images was made in four categories according to the ground truth. When the images without pectoral muscle are submitted to the classifier there is a significant increase in accuracy. Again, it is the same combination of features that get the best results, but in this case the values are higher. So for cross-validation the best accuracy is 80.7 and when all images are used for training and testing the accuracy is 91.5. An example of a confusion matrix obtained to this database is presented in Table 12.

Table 12 – Confusion Matrix.

	1	**2**	**3**	**4**
1	126	8	0	3
2	21	115	0	10
3	1	3	19	5
4	7	16	2	74

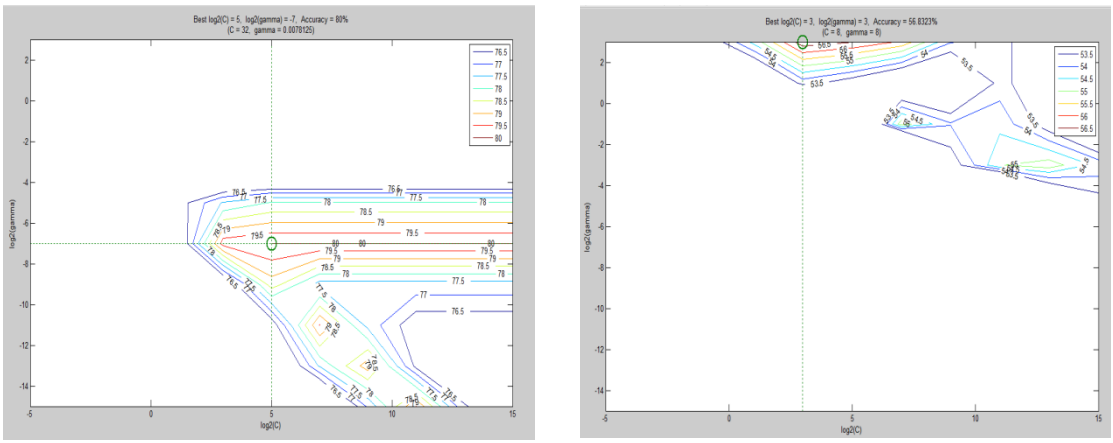
Another test was made in order to evaluate the accuracy in classification of breast density. The SVM usually are used to classify binary problems so the categories of density were divided into higher density or lower density and for all features the results are presented in Table 13.

Table 13 – Accuracy results for two categories of density.

Features	MIAS		INBreast	
	Cross Validation (%)	**All images for training and all for test(%)**	**Cross Validation (%)**	**All images for training and all for test(%)**
Statistical	82.9	85	82	88.5
Co occurrence-matrix	83.2	92.9	82.9	92.9
Statistical+ Co – occurrence	85.4	87.9	87.5	95
Spectrum	79.5	88.8	81.7	100
PCA+2DPCA(5 components)	67.1	78.9	78	90.3
Invariants moments of Hu	78	81.7	69.3	88
Statistical + Co-occurrence matrix + Invariants moments of Hu	83.6	95.7	75.8	89.4
Statistical + Co-occurrence matrix + Spectrum	81.7	90.7	82.7	93.9

It is clear that there is a very evident difference when used only two categories of density. In all features the accuracy increased compared to the tables 10 and 11, but again the combination of statistical descriptors and co-occurrence matrix may prove the most effective. This classification is simpler but the algorithm demonstrates the ability to differentiate tissues of high and low density, which is a satisfactory result.

The invariants moments of Hu shown to be not very effective in this type of problem, however as mentioned in the previous chapter some parameters are used to calibrate the algorithm using the grid search scheme, trying to minimize the classification error. In the figure below, Figure 33, provides two examples of a grid scheme, one represents the worst results and the other the best.



(a) (b)

Figure 33 – Graphic representation of grid scheme: (a) best result; (b) worst result.

It is important to note that the use of 2DPCA proved to be better than the usual PCA. In addition, when combined with the features of co-occurrence reasonable results obtained in the order of 70%. For both types of PDA were tested several components including, but differences were not significant.

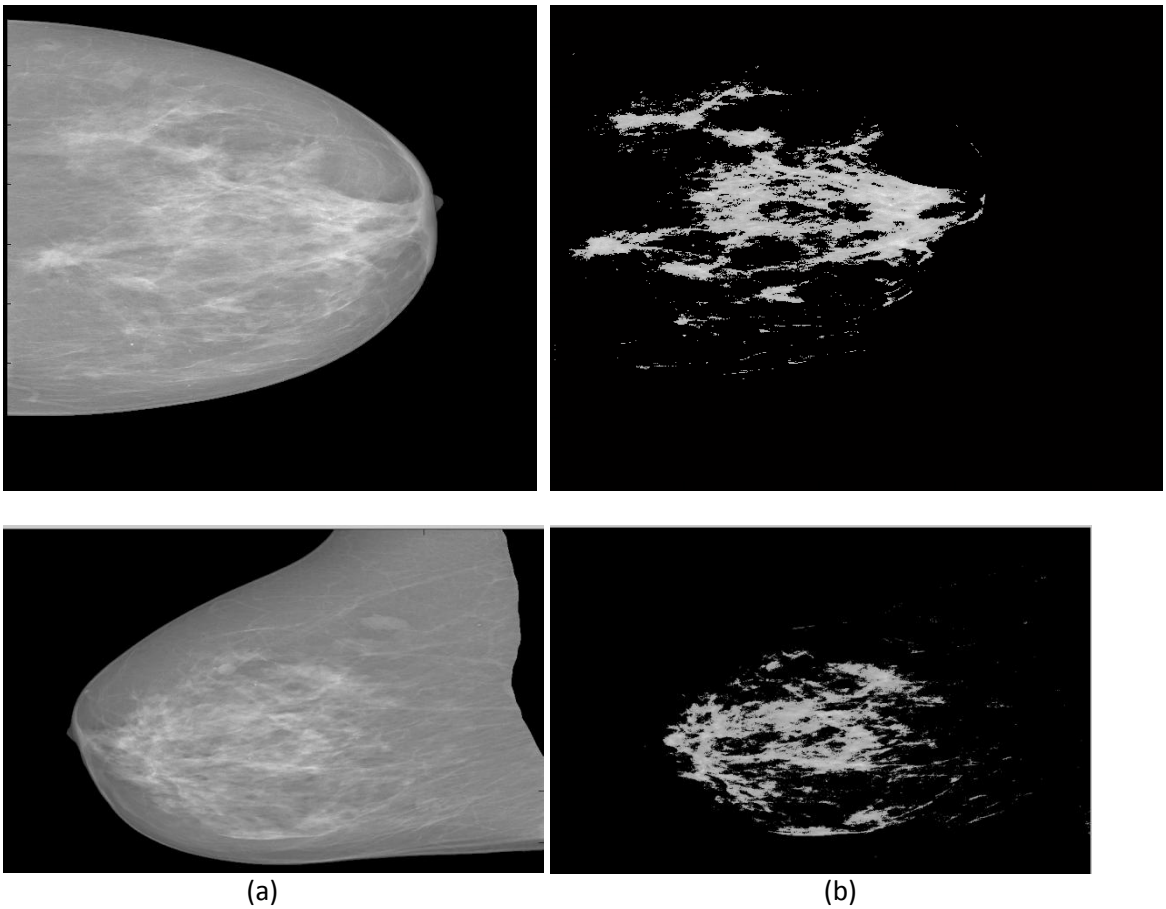
The proposed method is simple but at the same time promising because may allow contemplate another type of information that will improve them. The fact of the base MIAS have accuracies much lower than INbreast is not only related to the fact of the images were badly obtained, but also due to having only one intermediate class. The base INBreast is a good way of testing any algorithm because the images are of excellent quality and have all the necessary information.

5.4.2 Masses

For the detection algorithm masses were subjected classified as normal images and images with masses of both projections. However the MLO projection images, first underwent segmentation technique pectoral muscle and hence the removal of that of the image.

Although the entire process is automated using techniques from low computational cost, the image size directly influences the running time of the algorithm.

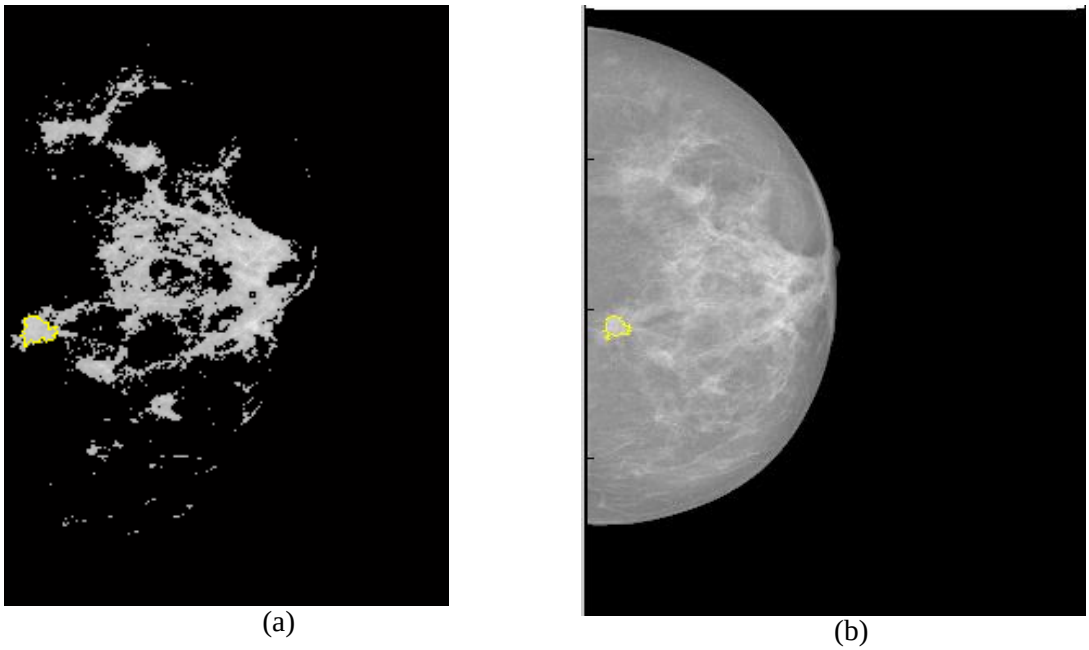
As explained in the previous chapter it was found a value that allows get only certain parts of the image, between which where is located the masses, Figure 34.



(a) (b)

Figure 34 – (a) Original Image; (b) Images after processing.

It is clear that there is still a large part of the image to explore, but it is easier to analyze this small part especially when is sure that the mass is there, as shown in Figure 35.



(a) (b)

Figure 35 – (a) Image after processing with a mass; (b) Original image with a mass.

Although the classification is not able to identify the exact location of the mass, only this data the classifier can distinguish satisfactorily a normal breast and a breast with a mass, as can be seen by analyzing Table 14.

Table 14 – Results of classification in normal breast and breast with masses.

Features	Accuracy of cross validation (%)	Accuracy using all images for training and test (%)
Statistical	70	100
Texture	85	94.6
Statistical+ Texture	76.7	94.6
Spectrum	84.2	97.3

It is possible to observe that when the image is complete, the accuracies rates are not so high, Table 15.

Table 15 – Results of classification in normal breast and breast with masses with the original image.

Features	Accuracy of cross validation (%)	Accuracy using all images for training and test (%)
Statistical	76.7	97.3
Texture	72.5	97.3
Statistical+ Texture	80	97.2
Spectrum	67.5	91.9

As improvement of this work we intend to further divide the first image processing in small ROI and analyze each one to identify which one is the injury. When the image suffers a small pre-processing the results are a little better than when the original image is directly submitted to the classifier. Thus, the features extracted from the co-occurrence matrix shown to be most effective in classification with a success rate of 85%. Given that this type of injuries are very common and involve a special attention by the technicians and doctors, the fact of can reduce the image to a region where is the lesion can be considered a great help.

5.5 Summary

Presented various algorithms reached the objectives proposed and shown to be promising in respect to a complete CAD system. After extensive research it was verified that no method of automatic image analysis distinguishes between left and right breast and between CC and MLOprojection. The removal of the pectoral muscle is a very important step in any system analysis and image processing mammography, since it can influence the results. Finally, breast density has been much studied because it is believed to be a relationship between it and between the onsets of cancer classification should therefore be the first step of this stage. The distinction between a normal and a breast mass is a considerable advance in the detection of lesions.

CHAPTER VI – Conclusions and Future Perspectives

6. Conclusions and Future Perspectives

Early detection of breast cancer is of utmost importance, since only localized cancer is deemed to be treatable and curable, as opposed to metastasized cancer.

Mammography is a widely used screening tool and is the gold standard for the early detection of breast cancer. However, many suspicious findings on mammograms are benign. The most important mammographic signs of malignancy are masses and microcalcifications. Yet, the sensitivity of screening mammography is affected by image quality and the radiologist's level of expertise.

CAD systems have been developed to improve the results of mammographic screening programs by assisting radiologists in decision between follow up and biopsy, thereby decreasing the number of false- positive cases.

Some CAD systems have reasonable results on detecting breast lesions, mainly microcalcifications and masses. But, some abnormalities are very difficult to detect in mammograms even as some structures of breast can be confused with abnormalities and cause an increase of false positives.

Adjustment of CAD algorithms would increase its sensitivity and specificity and consequently, would increase the trust from the radiologists in these systems, as the role of CAD is still emerging. The development of the digital mammography may also increase the CAD effectiveness due to the potential to improve contrast resolution compared with film-screen imaging.

In this work it was proposed a method for breast density classification that consists in use combination of features and also segment and extract the pectoral muscle in order to prove that can influence the results of classification. Once the main goal is create a completely automatic method of mammography images analysis, the firsts steps were distinct the directions and the projections of breast. There are only few methods that are dedicated to make that distinction soon after image acquisition. It is clear that analyzing image without resorting to computational methods is easy to see if it is a right or left breast, and also what the projection. However, in a clinic or hospital where there are many images to analyze monetize it takes time, so do these distinctions at the beginning of the process of automatic analysis is a useful step. Moreover, since there is not a standard naming this method becomes an advantage in this aspect. Given that the images of MIAS database were considered unsuitable to be used as a test algorithm and even than achieved an accuracy rate of 85% for the distinction between left and right breast may be considered that the result was very positive. To emphasize the fact that the method of distinguishing between left and right breast is well developed it was obtained a hit rate of 100% for INBreast database. On the other hand, the results obtained from the distinction between CC and MLO projection were not as satisfactory as the first since the success rate decreased a little.

Note that if the first distinction was always made correctly to the images of the MIAS database, this algorithm would also have a 100% rate of success. These differences that occur between these image databases may be due to the fact that the conditions for acquisition and formats that stored are different, and then they are not being compared equally.

It is visible that the presented method for segmentation and extraction pectoral muscle can be an important phase in any system of automatic analysis of mammography images. Although in the case of image database MIAS not have a ground truth with the contour points of the pectoral muscle in order to compare the results obtained, it was found another way validation have been obtained satisfactory results. In respect to the base INBreast the result was relatively similar to the above. In fact these results are slightly below compared to other literature studies but even when the pectoral muscle is only partially removed already positively influence the classification results.

The combination of features has proved to be a methodology that can be extended to other features that were not included in this work. Compared to other studies the classifier used proved to be very effective, inclusive possessed the advantage of seeking the parameters that allow optimizing the process and getting better results. It has been proven that the algorithm PCA two dimensions that are better than the usual PCA. However it should be noted that as the base INBreast images have a size in the order of a few Megs all this method has a very high computational cost.

On the overall, it is very important to continue the development of systems that can improve and help in the study of the breast cancer, creating algorithms capable of aiding the expert, and reducing their examination time and subjectivity. Thus the method showed can be very promising and it is suitable to continue working on, until has also a classification and identification of lesions.

References

References

Ackerman, L.V., Gose, E.E. (1972) Breast lesion classification by computer and xeroradiograph, *Cancer* 30 (4) 1025-1035.

Akay, M. (2006) *Wiley of Encyclopedia of Biomedical Engineering* (1st edition). John Wiley & Sons.

Alvin, Siverstein, V., & Nunn, L. S. (2006) *Cancer*. (L. of congress cataloging in Publication, Ed.).

Ayres, F. J., & Rangayyan, R. M. (2003) Characterization of architectural distortion in mammograms. *IEEE engineering in medicine and biology magazine : the quarterly magazine of the Engineering in Medicine & Biology Society*, *24*(1), 59-67.

Bajger, M., Ma, F., Bottema, M. J. (2005) Minimum Spanning Trees and Active Contours for Identification of the Pectoral Muscle in Screening Mammograms. *Proceedings of the Digital Imaging Computing: Techniques and Applications*, pp. 323-329.

Baker, J.A., Rosen, E.L., Lo, J.Y., Gimenez, E.I., Walsh, R., & Soo, M.S. (2003) Computer-aided detection (CAD) in screening mammography: Sensitivity of commercial CAD systems for detecting architectural distortion. *American Journal of Röentgenology*, vol. 181, pp. 1083-1088.

Banik, S., Rangayyan, R. M., & Desautels, J. E. L. (2009) Detection of Architectural Distortion in Prior Mammograms of Interval-cancer Cases with Neural Networks. *31st Annual International Conference of the IEEE EMBS Minneapolis, Minnesota, USA.*, 6667-6670.

Banik, S., Rangayyan, R. M., & Desautels, J. E. L. (2011) Detection of architectural distortion in prior mammograms. *IEEE transactions on medical imaging*, *30*(2), 279-94.

Birdwell, R.L., Ikeda, D.M., O'Schaughnessy, K.F., & Sickles, E.A. (2001) Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. *Radiology*, vol. 219, no. 1, pp. 192-202.

Bluekens, A. M. J., Karssemeijer, N., Beijerinck, D., Deurenberg, J. J. M., van Engen, R. E., Broeders, M. J. M., & den Heeten, G. J. (2010) Consequences of digital mammography in population-based breast cancer screening: initial changes and long-term impact on referral rates. *European radiology*, *20*(9), 2067-73.

Boyer, B., Balleyguier, C., Granat, O., & Pharaboz, C. (2009) CAD in questions/answers Review of the literature. *European journal of radiology*, *69*(1), 24-33.

Bozek, J., & Dumic, E. (2009) Bilateral asymmetry detection in digital mammography using B-spline interpolation. *Systems, Signals and Image*, 1-4.

Brake, G. M., & Karssemeijer, N. (1999) Single and multiscale detection of masses in digital mammograms. *IEEE transactions on medical imaging*, 18(7), 628-39.

Bruce, L. M., & Adhami, R. R. (1999) Classifying mammographic mass shapes using the wavelet transform modulus-maxima method. *IEEE transactions on medical imaging*, *18*(12), 1170-7.

Byng JW, Yaffe MJ, Jong RA, et al. (1998) Analysis of mammographic density and breast cancer risk for digitized mammograms. *RadioGraphics*. **18**:1587- 1598.

Byrne C, Schairer C, Wolfe J, Parekh N, Salane M, Brinton LA, Hoover R, Haile R. (1995) Mammographic features and breast cancer risk: effects with time, age, and menopause status**.** *J Natl Cancer Inst.*, 1;87(21):1622-9.

Cardenosa, Gilda (2004) *Breast Imaging*. Lippincott Williams & Wilkins.

Chan, H.P., Doi, K.., Galhotra, S., Vybomy, C. J., Jokich, H., & MacMahon, P. M. (1987) Image feature analysis and computer-aided diagnosis in digital radiography. I: Automated detection of microcalcifications in mammography. *Med. Phys*., vol. 14, no. 4, pp. 538-548.

Chan, T., & Huang, H. K. (2008) Effect of a computer-aided diagnosis system on clinicians' performance in detection of small acute intracranial hemorrhage on computed tomography. *Academic radiology*, *15*(3), 290-9.

Chang, R. F., Huang, S. F., Wang, L. P., Chen, D. R., & Moon, W. K. (2005) Microcalcification detection in 3-d breast ultrasound. *Conference proceedings :Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, *6*(1), 6297-300.

Chang, C-C and Lin, C-J (2011) LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol 2,27.

Cheng, H., Luy, Y.M., & Freimanis, R.I. (1998) A novel approach to microcalcification detection using fuzzy logic technique, *IEEE Trans. Med. Imaging* 17 (3) 442-450.

Chianyama Catherine N. (2004) *Breast Diseases: Radiology, Pathology, Risk Assessment.* Springer-Verlag Berlin Heidelberg New Yorkublishing, Ed.

Crammer, K., Singer, Y. (2000) On the learnability and design of output codes for multiclass problems, *Computer Learning Theory* , 35–46.

Dengler, J., Behrens, S., & Desaga, J. F. (1993) Segmentation of microcalcifications in mammograms. *IEEE transactions on medical imaging*, *12*(4), 634-42.

Dheeba, J. (2011) Classification of malignant and benign microcalcification using SVM classifier. *Emerging Trends in Electrical and Computer Technology (ICETECT)*, 686-690.

Dixon, J. M. (2006) *ABC of breast diseases*. (3rd ed.) B. Publishing, Ed.

Doi, K. (2007) Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Comput Med Imaging Graph.* , 31, No. 4-5, pp. 198-211.

Duin, R. P. W. and Mao Jain, A. K., J. statistical pattern recognition. (2000) *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4– 37.

Engeland, S. V., Snoeren, P., Hendriks, J., & Karssemeijer, N. (2003) Mammogram Registration, *22*(11), 1436-1444.

Estevez, L. W., & Kehtarnavaz, N. D. (1995) Computer assisted enhancement of mammograms for detection of microcalcifications. *Proceedings Eighth IEEE Symposium on Computer-Based Medical Systems*, 16-23. IEEE Comput. Soc. Press.

Evans, W.P., Burhenne, L.J.W., Laurie, L., O'Shaughnessy, K.F., & Castellino, R.A. (2002) Invasive lobular carcinoma of the breast: Mammographic characteristics and computer-aided detection. *Radiology*, vol. 225, no. 1, pp. 182-189.

Ferrari, R.J., Rangayyan, R.M., Desautels, J.E.L., & Frere, A.F. (2001) Analysis of Asymmetry in Mammograms via Directional Filtering with Gabor Wavelets. *IEEE Transactions on Medical Imaging*, Vol. 20, No. 9, pp. 953-964

Ferrari, R. .J., Rangayyan, R. M., Desautels, et al. (2004) Automatic Identification of the Pectoral Muscle in Mammograms. *IEEE Transactions on Medical Imaging*, v. 23, n. 2, pp. 232-245.

Fujita, H., Uchiyama, Y., Nakagawa, T., Fukuoka, D., Hatanaka, Y., Hara, T., & Lee, G. N. (2008) Computer-aided diagnosis: the emerging of three CAD systems induced by Japanese health care needs. *Computer methods and programs in biomedicine*, *92*(3), 238-48.

Georgsson, F. (2003) Differential analysis of bilateral mammograms. International *Journal on Pattern Recognition and Artificial Intelligence*, Vol. 17, pp. 1207-1226.

Good, W.F., Wang, X.H., & Maitz, G. (2003) Feature- based Differences Between Mammograms. *Medical Imaging 2003: Image Processing, Proceedings of SPIE*, Vol. 5032, pp. 919-929.

Gray, H. (2000) *Anatomy of the Human Body*. New York: Bartleby.

Guyton, A. C., & Hall, J. E. (2000) *Textbook of Medical Physiology* (10th ed.). (W. S. Company, Ed.).

Hadjiiski, L., Sahiner, B., Chan, H.-P., Petrick, N., Helvie, M.A., & Gurcan, M. (2001) Analysis of Temporal Changes of Mammographic Features: Computer-Aided Classification of Malignant and Benign Breast Masses, *Medical Physics*, Vol. 28, No. 11, pp. 2309-2317.

Haykin, S. (1999) *Neural Networks*. Prentice Hall.

Hajnal, S. , Taylor, P , Dilhuydy, M-H, Barreau, B. and Fox, J.(1993) Classifying mammograms by density: rationale and preliminary results, *Proc. SPIE* 1905, *Biomedical Image Processing and Biomedical Visualization*, 478.

Hashimoto, Beverly (2008) *Practical Digital Mammography.* Thieme Medical Publishers.

Hatanaka Y., Hara, T., Fujita, H., et al. (2001) Development of an Automated Method for Detecting Mammographic Masses With a Partial Loss of Region. *IEEE Transactions on Medical Imaging*, v. 20, n. 12, pp. 1209-1214.

Hu, M-K, (1962) Visual pattern recognition by moment invariants. *IRE Trans. on Information Theory*, 179–187.

Huo, Z., Giger , M.L., Vyborny , C.J., Bick, U., Lu, P., Wolverton, D.E., & Schmidt, R.A.(1995) Analysis of Spiculation in the Computerized Classification of Mammographic Masses, *Medical Physics*, Vol. 22, No. 10, pp. 1569-1579.

Huo, Z., Giger , M.L., Vyborny , C.J., Wolverton, D.E., Schmidt, R.A., & Doi, K. (1998) Automated Computerized Classification of Malignant and Benign Masses on Digitized Mammograms, *Academic Radiology*, Vol. 5, No. 3, pp. 155-168.

Hsu, C.W.,Lin, C.J. (2002) A comparison of methods for multiclass support vector machines, *IEEE Transactions on Neural Networks*, 13, 415–425.

Jatoi, I., & Kaufmann, M. (2010) *Management of breast diseases*. Springer, Ed..

Jiang, J., Yao, B., & Wason, A.M. (2005) Integrating of fuzzy logic and structural tensor towards mammogram contrast enhancement, *Comput. Med. Imaging Graph* 29 (1) 83-90.

Kalbhen, C.L., McGill, J.J., Fendley, P.M., Corrigan, K.W., & Angelats, J. (1999) Mammographic determination of breast volume: comparing different methods. *American Journal of Roentgenology*, vol. 173:1643-1649.

Karssemeijer, N. (1995) Detection of stellate distortions in mammograms using scale space operators. Y. Bizais, C. Barrilot, R.D. Paola (Eds.). Information Processing in Medical Imaging, Kluwer, Dordrecht, pp. 335-346.

Karssemeijer, N. (1998) Automated Classification of Parenchymal Patterns in Mammograms. *Physics in Medicine and Biology*, v. 43, pp. 365-378.

Katariya, R. N., Forest , A.P.M., & Gravelle, I.H. (1974) Breast volumes in cancer of the breast. *Br. J. Cancer*, 29:270-273.

Kavitha, K., & Kumaravel, N. (2007) A Comparitive Study of Various MicroCalcification Cluster Detection Methods in Digitized Mammograms. *2007 14th International Workshop on Systems, Signals and Image Processing and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services*, 405-409. IEEE.

Kendall, J. (1980) *Statistical pattern recognition. Charles Griffin & Company, London*

Kilday, J., Palmieri, F., & Fox, M. D. (1993) Classifying mammographic lesions using computerized image analysis. *IEEE transactions on medical imaging*, *12*(4), 664-9.

Kim, J. K., & Park, H. W. (1999) Statistical textural features for detection of microcalcifications in digitized mammograms. *IEEE transactions on medical imaging*, *18*(3), 231-8.

Kobatake, H., & Hashimoto, S. (1999) Convergence index filter for vector fields. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, *8*(8), 1029-38.

Kobatake, H., Murakami, M., Takeo, H., & Nawano, S. (1999) Computerized detection of malignant tumors on digital mammograms. *IEEE transactions on medical imaging*, *18*(5), 369-78.

Kopans, Daniel B. (2007) *Breast Imaging* (3rd edition). Lippincott Williams & Wilkins.

Kwok, S. M., Chandrasekhar, R., Attikiouzel, Y., et al. (2004) Automatic Pectoral Muscle Segmentation on Mediolateral Oblique View Mammograms. *IEEE Transactions on Medical Imaging*, v. 23, n. 9, pp. 1129-1140.

Laine, A., Fan, J., & Yang, W. (1995) Wavelets contrast enhancement of digital mammography, *IEEE Eng. Med. Biol*. 14 (1995) 536-550.

Lau, T.-K., & Bischof, W.F. (1991) Automated Detection of Breast Tumors Using the Asymmetry Approach. *Computers and Biomedical Research*, Vol. 24, No. 3, pp. 273-295.

Li, H. D., Kallergi, M., Clarke, L. P., Jain, V. K., & Clark, R. A. (1995) Markov random field for tumor detection in digital mammography. *IEEE transactions on medical imaging*, *14*(3), 565-76.

Li, H., Liu, K.J.U., & Lo, S.C.B.(1997) Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms, *IEEE Trans. Med. Imaging* 16 (6) 785-798.

Li, B., & Wang, X. (2009) The bilateral information asymmetry on insurance market. *Industrial Engineering and*, (70772057), 750-752.

Lodwick, G.S., Haun, C.L., & Smith,W.E. (1963) Computer diagnosis of primary bone tumor, *Radiology* 80 273–275.

Malik, M. A. N. (2010) Breast Diseases ; *17*(3), 366-372.

Manni, Andrea (1999) *Endocrinology for Breast Cancer.* Humana Press.

Marrocco, C., Molinara, M., D'Elia, C., & Tortorella, F. (2010, September) A computer-aided detection system for clustered microcalcifications. *Artificial intelligence in medicine*. Vol. 50, pages 23-32.

Masek, M., Chandrasekhar, R., De Silva, C. J. S., et al. (2001) Spatially Based Application oh the Minimum Cross-Entropy Thresholding Algorithm to Segment the Pectoral Muscle in Mammograms. *Seventh Australian and New Zealand Intelligent Information Systems Conference*, pp. 101-106.

Matsubara, T., Ichikawa, T., Hara, T., Fujita, H., Kasai, S., Endo, T., & Iwase, T. (2003) Automated detection methods for architectural distortions around skinline and within mammary gland on mammograms. *International Congress Series: Proc. of the 17th International Congress and Exhibition on Computer Assisted Radiology and Surgery*, London: Elsevier, pp. 950-955.

Méndez, A.J., Tahoces, P.G., Lado, Souto, M.J., & Vidal, J.J. (1998) Computer-Aided Diagnosis: Automatic Detection of Malignant Masses in Digitized Mammograms. *Medical Physics*, Vol. 25, No. 6, pp. 957-964.

Meyers, P.H., Nice, C.M., Becker, H.C., Nettleton, W.J., Sweeney, J.W., & Meckstroth, G.R. (1964) Automated computer analysis of radiographic images, *Radiology* 83 1029-1033.

Miller, P., & Astley, S. (1993) Detection of Breast Asymmetry Using Anatomical Features. *R.S. Acharya, C.B. Goldgof (Eds.), Biomedical Image Processing and Biomedical Visualization, Proceedings of SPIE*, Vol. 1905, pp. 433-442.

Miller, P., & Astley, S. (1994) Automated detection of breast asymmetry using anatomical features. *K.W. Bowyer, S. Astley (Eds.), State of the Art in Digital Mammographic Image Analysis, Series in Machine Perception and Artificial Intelligence*, Vol. 9, World Scientific, River Edge, NJ, pp. 247-261.

Moinfar, F. (2007) *Essentials of diagnostic breast pathology*. Springer.

Moore, K. L., Agur, A. M., & Dalley, A. F. (2004) *Essential Clinical Anatomy*. (4nd ed.). (W. Kluwer, Ed.).

Mudigonda, N. R., Rangayyan, R. M., & Desautels, J. E. (2000) Gradient and texture analysis for the classification of mammographic masses. *IEEE transactions on medical imaging*, *19*(10), 1032-43.

Mudigonda, N.R. & Rangayyan R.M. (2001) Texture flow-field analysis for the detection of architectural distortion in mammograms, *Proc. of Biovision*, Bangalore, India, pp. 76-8.

Nandi, R.J., Nandi, A.K., Rangayyan, R.M., & Scutt, D. (2006) Classification of Breast Masses in Mammograms Using Genetic Programming and Feature Selection, *Medical and Biological Engineering and Computing*, Vol. 44, No. 8, pp. 683-694.

Naylor, S. M., York, J. (1999) An Evaluation of the Use of Pectoral Muscle to Nipple Level as a Component to Asses the Quality of the Medio-Lateral Oblique Mammogram. *Radiography*, v. 5, pp. 107-110.

Nees, A. V. (2008). Digital mammography: are there advantages in screening for breast cancer? *Academic radiology*, *15*(4), 401-7.

Netsch, T., Peitgen, H.-O. (1999) Scale-Space Signatures for the Detection of Clustered Microcalcifications in Digital Mammograms. *IEEE Transactions on Medical Imaging*, Vol. 18, No. 9, pp 774-786.

Nishikawa, R.M., Giger, L.M., Doi K., Vyborny C.J., & Schmidt R.A. (1994) Computer-aided detection and diagnosis of masses and microcalcifications from digital mammograms. *State of the Art in Digital Mammographic Image Analyis, Bowyer and Astley S, Eds*., pp. 82-102, Singapore: World Scientific.

Panno, J. (2005) *Cancer: the role of genes, lifestyle and environment*. Library of congress cataloging in Publication.

Paquerault, S., Hardy, P. T., Wersto, N., Chen, J., & Smith, R. C. (2010) Investigation of optimal use of computer-aided detection systems: the role of the "machine" in decision making process. *Academic radiology*, *17*(9), 1112-21. Elsevier Ltd.

Pasqualette, Henrique; Koch, Hilton; Soares-Pereira, Paulo; Kemp, Cláudio (1998) *Mamografia Actual*. Rio de Janeiro: Revinter.

Perez, Reinaldo J. (2002) *Design of medical electronic devices*. Academic Press

Petrick, N., Chan, H.P., Sahiner, B., & Helvie, M. (1999) Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms, *Med. Phys*. 26 (8), 1642-1654.

Pimentel, F. V. (2004) "Pré-Processamento de Imagens Mamográficas e Extraçãode Parâmetros das Microcalcificações", Tese de M. Sc., Programa de Engenharia Biomédica, COPPE/UFRJ, Rio de Janeiro, Brasil, 96 p.

Pisano, Etta D.,Yaffe, Martin J., Kuzmiak, & Cherie M.(2004) *Digital Mammography*. Lippincott Williams & Wilkins.

Pohlman, S., Powell, K. A., Obuchowshi , N. A., Chilote , W. A., & Grundfest-Broniatowski, S. (1996) Quantitative classification of breast tumors in digitized mammograms. *Med. Phys.,* vol. 23, pp. 1337–1345.

Qian, W., Li, L., & Clarke, L.P. (1999) Image Feature Extraction for Mass Detection in Digital Mammography: Influence of Wavelet Analysis, *Medical Physics*, Vol. 26, No. 3, pp. 402-408.

Raba, D., Oliver, A., Martí, J., et al. (2005) Breast Segmentation With Pectoral Muscle Suppression on Digital Mammograms. *Proceedings of the Second Iberian Conference: Pattern Recognition and Image Analysis*, Part II.

Rangayyan, R.M., Ferrari, R.J., & Frere, A.F. (2007) Analysis of Bilateral Asymmetry in Mammograms Using Directional, Morphological, and Density Features. *Journal of Electronic Imaging*, Vol. 16, No. 1.

Rovere, G. Q. della, Warren, R., & Benson, J. R. (2006) *Early Breast Cancer: From Screening to Multidisciplinary Management* (2nd edition). Taylor & Francis.

Sahiner, B., Chan, H.P., Wei , D., Petrick, N., Helvie., M.A., Adler, D.D., & Goodsitt, M.M. (1996) Image Feature Selection by a Genetic Algorithm: Application to Classification of Mass and Normal Breast Tissue, *Medical Physics*, Vol. 23, No. 10, pp. 1671-1684.

Sahiner, B., Chan, H.P., Petrick, N., Helvie, M.A., & Goodsitt, M.M. (1998) Computerized Characterization of Masses on Mammograms: the Rubber Band Straightening Transform and Texture Analysis, *Medical Physics*, Vol. 25, No. 4, pp. 516-526.

Saki, F., & Tahmasbi, A. (2010) A novel opposition-based classifier for mass diagnosis in mammography images. *Engineering (ICBME)*, 3-4.

Sampat, M., Markey, M., & Bovik, A. (2005) Computer-Aided Detection and Diagnosis in Mammography. In A. Bovik, Handbook of Image and Video Processing (pp. 1195-1217). *Elsevier*.

Seeley, R., Stephens, T., & Tate, P. (2004) *Anatomy and Physiology*. The McGraw−Hill Company.

Shutler, J. (2002)Statistical moments. *CV Online: On-Line Compendium of Computer Vision*

Sivaramakrishna, R., & Gordon, R. (1997) Detection of Breast Cancer at a Smaller Size Can Reduce the Likelihood of Metastatic Spread: A Quantitative Analysis. *Acad Radiol, Association of University Radiologists*, Canadá, v. 4, n. 1, p. 8-12.

Sonka, Milan, Fitzpatrick, & J. Michael (2000) *Handbook of Medical Imaging, Vol 2. Medical Image Processing and Analysis*, SPIE Press.

Stamatakis, E.A., Ricketts, I.W., Cairns, A.Y., Walker, T., & Preece, P.E. (1996) Detecting Abnormalities on Mammograms by Bilateral Comparison. *IEE Colloquium on Digital Mammography*, London, UK, pp. 12/1-12/4.

Strickland, R.N., & Hahn, H.(1996) Wavelet transforms for detecting microcalcifications in mammograms, *IEEE Trans. Med. Imaging* 15 (2) 218-229.

Subashini, T., Ramalingam, V., and Palanivel,S. (2010) Automated assessment of breast tissue density in digital mammograms. *Computer Vision and Image Understanding*, vol. 114, no. 1, pp. 33–43.

Sukling, J., Parker, J., Dance, D. R., et al. (1994) The Mammographic Image Analysis Society Database. *Exerpta Medica International Congress Series*, v.1069, pp. 375-378.

Sun, L., Li, L., Xu, W., Liu, W., Zhang, J., & Shao, G. (2010) A Novel Classification Scheme for Breast Masses Based on Multi-View Information Fusion. *2010 4th International Conference on Bioinformatics and Biomedical Engineering*, 1-4. IEEE.

Tahmoush, D., & Samet, H. (2006) Image Similarity and Asymmetry to Improve Computer-Aided Detection of Breast Cancer. *Proceedings of the International Workshop on Digital Mammography (IWDM) 2006*, pages 221- 228, Manchester, UK.

Tang, J., Rangayyan, R. M., Xu, J., El Naqa, I., & Yang, Y. (2009) Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, *13*(2), 236-51.

Tartar, Marie, Comstock, Christopher E., Kipper, & Michael S. (2008) *Breast Cancer Imaging:A Multidisciplinary, Multimodal Approach.* Mosby Elsevier.

Taylor, P., Hajnal, S., Dilhuydy, M-H and Barreau, B. (1994) Measuring image texture to separate difficult from easy mammograms, *The British Journal of Radiology*, Vol. 67, pp. 456–463.

Taylor, J. S. and Cristianini,N. (2000) *Support Vector Machines and other kernel-based learning methods.* Cambridge University Press.

Ursin, G., & Qureshi, S. A. (2009) Mammographic density – a useful biomarker for breast cancer risk in epidemiologic studies, *Norsk Epidemiologi*, *19*(1), 59-68.

Webster, J. (2006) *Encyclopledia of Medical Devices and Instrumentation* (2nd ed., Vol. 4). U.S.: Wiley Insterscience.

Winsberg, F., Elkin, M., Macy, J., Bordaz, V., & Weymouth, W. (1967, Aug.) Detection of Radiographic Abnormalities in Mammograms by Means of Optical Scanning and Computer Analysis. *Radiology* , 89, pp. 211-215.

Wirth, M., Fraschini, M., Lyon, J., & Università, M. (2004) Contrast enhancement of microcalcifications in mammograms using morphological enhancement and non-flat structuring elements. *IEEE Computer Society*.

Woods, Richard E., Gonzalez, Rafael C., and Eddins, Steven L.. (2002) *Digital image processing using matlab*, Pearson Education.

Wróblewska, A., Boniński, P., Przelaskowski, A., & Kazubek, M. (2003) Segmentation and feature extraction for reliable classification of microcalcifications in digital mammograms. *Opto - Electron*, *11*(3), 227-235.

Wu, Y., Doi, K., Giger, M. L., & Nishikawa, R. M. (1992) Computerized detection of clustered microcalcifications in digital mammograms: Applications of artificial neural networks. *Med. Phys*., vol. 19, pp. 555-560.

Yaffe, M. J. (2008) Mammographic density. Measurement of mammographic density. *Breast cancer research : BCR*, *10*(3), 209.

Yaghjyan, L., Pinney, S. M., Mahoney, M. C., Morton, A. R., & Buckholz, J. (2011) Mammographic Breast Density Assessment: A Methods Study. *Atlas Journal of Medical and Biological Sciences*, 8-14.

Yam, M., Brady, M., Highnam, R., et al. (2001) Three-Dimensional Reconstruction of Microcalcification Clusters from Two Mammographic Views. *IEEE Transactions on Medical Imaging*, v. 20, n. 6, pp. 479-489.

Yang, J., Zhang, D., Frangi, A. F., and Yang, J. (2004) Two-dimensional PCA: A new approach to appearance-based face representation and recogni-tion. *IEEE Transactions on Pattern Analysis Machine and Intelligence*, vol. 26, no. 1, pp. 131–137.

Yu, S., Brown, S., Xue, Y., & Guan, L. (1996) Enhancement and Identification of Microcalcifications in Mammqgram Images Using Wavelets, 1166-1171.

Yu S, Guan L. (2000) A CAD System for the Automatic Detection of Clustered Microcalcifications in Digitized Mammogram Films. *IEEE Transactions On Medical Imaging*, Vol. 19, No. 2, pp 115-126.

Zhao, H., Xu, W., Li, L., & Zhang, J. (2011) Classification of Breast Masses Based on Multi-View Information Fusion Using Multi-Agent Method. *2011 5th International Conference on Bioinformatics and Biomedical Engineering*, 1-4. IEEE.

Zheng, B., Chang, Y. H., & Gur, D. (1995) Computerized detection of masses in digitized mammograms using single-image segmentation and a multilayer topographic feature analysis. *Academic radiology*, *2*(11), 959-66.