UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA

Breast Cancer: Automatic Detection and Risk Analysis through Machine Learning Algorithms, using mammograms

João Pedro Velhinho Mendes

Mestrado Integrado em Engenharia Biomédica e Biofísica
Engenharia Clínica e Instrumentação Médica

Dissertação orientada por:
Professor Doutor Nuno Matela

2021

Acknowledgments

Em primeiro lugar, gostaria de agradecer ao meu orientador, Professor Nuno Matela: pelo apoio desde o primeiro dia em que mostrei interesse pelo tema que culminou neste trabalho, pelas respostas a todas as dúvidas e inseguranças que tive, por todas as contribuições valiosas que são parte integrante deste documento, e por tantas vezes me fazer ver que os problemas que tinha eram, invariavelmente, insignificantes.

Depois, queria agradecer a todos aqueles que, de uma forma ou de outra, se cruzaram comigo ao longo do decorrer deste trabalho e que me deixaram o seu apoio. De todos, destaco quatro: Quero agradecer ao Duarte, por tantas vezes se ter mostrado tão entusiasmado como eu pelo projeto, motivando-me ainda mais. Quero agradecer ao Vasco, pelas anotações que foi deixando, pelas sugestões que foi fazendo, e pelos abraços que foi dando quando as coisas corriam menos bem. Quero agradecer ao Manuel, pelo apoio que me dá a cada pequena vitória mesmo não fazendo ideia do que se está a passar. E, claro, quero agradecer à Sara por tantas coisas que um parágrafo não chega para as conter.. nomeio algumas: pelo interesse pelo meu trabalho, pelas revisões e anotações milimétricas, pela paciência que tem para mim, e por ser a aberta nas tempestades em copos de água que tantas vezes fiz.

Quero agradecer ao meu Tio Fernando, à minha Tia Angélica, e à minha Avó Joaquina, pelo apoio fundamental que me deram ao longo deste ano, e que tornou a minha vida muito mais tranquila.

Por fim, mas não menos importante, quero agradecer à minha família mais próxima: aos meus pais por, no meio de tantas dificuldades, nunca terem abdicado de investir na minha educação, por nunca terem exigido de mim mais do que dar o meu melhor, e por toda a compreensão e apoio que me deram em todas as decisões, mesmo que não fossem as melhores; ao meu irmão que, mesmo sem o verbalizar, me apoia todos os dias, sabendo que as vitórias, minhas e dele, serão sempre partilhadas… e por gozar comigo a toda a hora e me fazer perceber que não sou assim tão importante.

Quero, em última instância, agradecer à minha Avó Rosa. O amor, a amizade, a luta pelo que acreditamos, a resiliência, o rigor, e o "brio", são valores que ela me passava em todas as conversas que tínhamos, mesmo naquelas em que só ela é que falava. E a avó Rosa, mesmo não podendo assistir ao cumprir de mais um objetivo, está, de uma forma ou de outra, patente ao longo deste trabalho.

Abstract

Two million and three hundred thousand Breast Cancer (BC) cases were diagnosed in 2020, making it the type of cancer with the highest incidence that year, considering both sexes. Breast Cancer diagnosis usually occurs during screening programs using mammography, which has some downsides: the masking effect due to its 2-D nature, and its poor sensitivity concerning dense breasts. Since these issues result in difficulties reading mammograms, the main part of this work aimed to verify how a computer vision method would perform in classifying mammograms into two classes: cancer and non-cancer. The 'non-cancer group' (N=159) was composed by images with healthy tissue (N=84) and images with benign lesions (N=75), while the cancer group (N=73) contained malignant lesions. To achieve this, multiple classifiers were optimized and trained ($N_{train} = 162, N_{test} = 70$) with a previously selected ideal sub-set of features that describe the texture of the entire image, instead of just one small Region of Interest (ROI). The classifier with the best performance was Support Vector Machine (SVM), (AUC = 0.875), which indicates a good-to-excellent capability discriminating the two defined groups. To assess if Percent Mammographic Density (%PD), an important risk factor, added important information, a new classifier was optimized and trained using the selected sub-set of texture features plus the %PD calculation. The classifier with the best performance was a Linear Discriminant Analysis (LDA), (AUC=0.875), which seems to indicate, once it achieves the same performance as the classifier using only texture features, that there is no relevant information added from %PD calculations. This happens because texture already includes information on breast density. To understand how the classifier would perform in worst image acquisition conditions, gaussian noise was added to the test images (N=70), with four different magnitudes (AUC= 0.765 for the lowest noise value vs. AUC ≈ 0.5 for the highest). A median filter was applied to the noised images towards evaluating if information could be recovered. For the highest noise value, after filtering, the AUC was very close to the one obtained for the lowest noise value before filtering (0.754 vs 0.765), which indicates information recovery. The effect of density in classifier performance was evaluated by constructing three different test sets, each containing images from a density class (1,2,3). It was seen that an increase in density did not necessarily resulted in a decrease in performance, which indicates that the classifier is robust to density variation (AUC = 0.864, AUC= 0.927, AUC= 0.905 ; for class 1, 2, and 3 respectively). Since the entire image is being analyzed, and images come from different datasets, it was verified if breast area was adding bias to classification. Pearson correlation coefficient provided an output of $\rho = 0.22$, showing that there is a weak correlation between these two variables. Finally, breast cancer risk was assessed by visual texture feature analysis through the years, for a small set of women (N=11). This visual analysis allowed to unveil what seems to be a pattern amongst women who developed the disease, in the mammogram immediately before diagnosis. The details of each phase, as well as the associated final results are deeply described throughout this document. The work done in the first classification task resulted in a state-of-the-art performance, which may serve as foundation for new research in the area, without the laborious work of ROI definition. Besides that, the use of texture features alone proved to be fruitful. Results concerning risk may serve as basis for future work in the area, with larger datasets and the incorporation of Computer Vision methods.

Keywords: Breast Cancer, Medical Imaging, Texture Features, Machine Learning, Risk Analysis

Resumo

Com 2.3 milhões de casos diagnosticados em todo o Mundo, durante o ano de 2020, o cancro da mama tornou-se aquele com maior incidência, nesse mesmo ano, considerando ambos os sexos. Anualmente, em Portugal, são diagnosticados aproximadamente sete mil (7000) novos casos de cancro da mama, com mil oitocentas (1800) mulheres a morrerem, todos os anos, devido a esta doença - indicando uma taxa de mortalidade de aproximadamente 5 mulheres por dia. A maior parte dos diagnósticos de cancro da mama ocorrem ao nível de programas de rastreio, que utilizam mamografia. Esta técnica de imagem apresenta alguns problemas: o facto de ser uma imagem a duas dimensões leva a que haja sobreposição de tecidos, o que pode mascarar a presença de tumores; e a fraca sensibilidade a mamas mais densas, sendo estas caraterísticas de mulheres com risco de cancro da mama mais elevado. Como estes dois problemas dificultam a leitura das mamografias, grande parte deste trabalhou focou-se na verificação do desempenho de métodos computacionais na tarefa de classificar mamografias em duas classes: cancro e não-cancro. No que diz respeito à classe "não-cancro" (N = 159), esta foi constituída por mamografias saudáveis (N=84), e por mamografias que continham lesões benignas (N=75). Já a classe "cancro" continha apenas mamografias com lesões malignas (N = 73). A discriminação entre estas duas classes foi feita com recurso a algoritmos de aprendizagem automática. Múltiplos classificadores foram otimizados e treinados ($N_{treino} = 162, N_{teste} = 70$), recorrendo a um conjunto de características previamente selecionado, que descreve a textura de toda a mamografia, em vez de apenas uma única Região de Interesse. Estas características de textura baseiam-se na procura de padrões: sequências de pixéis com a mesma intensidade, ou pares específicos de pixéis. O classificador que apresentou uma performance mais elevada foi um dos *Support Vector Machine* (SVM) treinados – AUC= 0.875, o que indica um desempenho entre o bom e o excelente. A *Percent Mammographic Density* (%PD) é um importante fator de risco no que diz respeito ao desenvolvimento da doença, pelo que foi estudado se a sua adição ao set de features selecionado resultaria numa melhor performance dos classificadores. O classificador, treinado e otimizado utilizando as features de textura e os cálculos de %PD, com maior capacidade discriminativa foi um *Linear Discriminant Analysis* (LDA) – AUC = 0.875. Uma vez que a performance é igual à obtida com o classificador que utiliza apenas features de textura, conclui-se que a %PD parece não contribuir com informação relevante. Tal pode ocorrer porque as próprias características de textura já têm informação sobre a densidade da mama. De forma a estudar-se de que modo o desempenho destes métodos computacionais pode ser afetado por piores condições de aquisição de imagem, foi simulado ruído gaussiano, e adicionado ao set de imagens utilizado para testagem. Este ruído, adicionado a cada imagem com quatro magnitudes diferentes, resultou numa AUC de 0.765 para o valor mais baixo de ruído, e numa AUC de 0.5 para o valor de ruído mais elevado. Tais resultados indicam que, para níveis de ruído mais baixo, o classificador consegue, ainda assim, manter uma performance satisfatória – o que deixa de se verificar para valores mais elevados de ruído. Estudou-se, também, se a aplicação de técnicas de filtragem – com um filtro mediana – poderia ajudar a recuperar informação perdida aquando da adição de ruído. A aplicação do filtro a todas as imagens ruidosas resultou numa AUC de 0.754 para o valor mais elevado de ruído, atingindo assim um desempenho similar ao set de imagens menos ruidosas, antes do processo de filtragem (AUC=0.765). Este resultados parecem indicar que, na presença de más condições de aquisição, a aplicação de um filtro mediana pode ajudar a recuperar informação, conduzindo assim a um melhor desempenho dos métodos computacionais. No entanto, esta mesma conclusão parece não se verificar para valores de ruído mais baixo onde a AUC após filtragem acaba por ser mais reduzida. Tal resultado poderá indicar que, em situações onde o nível de ruído é mais baixo, a técnica de filtragem não só remove o ruído, como acaba também por, ela própria, remover informação ao nível da textura

da imagem. De modo a verificar se mamas com diferentes densidades afetavam a performance do classificador, foram criados três sets de teste diferentes, cada um deles contendo imagens de mamas com a mesma densidade (1, 2, e 3). Os resultados obtidos indicam-nos que um aumento na densidade das mamas analisadas não resulta, necessariamente, numa diminuição da capacidade em discriminar as classes definidas (AUC = 0.864, AUC = 0.927, AUC= 0.905; para as classes 1, 2, e 3 respetivamente). A utilização da imagem integral para analisar de textura, e a utilização de imagens de datasets diferentes (com dimensões de imagem diferentes), poderiam introduzir um viés na classificação, especialmente no que diz respeito às diferentes áreas da mama. Para verificar isso mesmo, utilizando o coeficiente de correlação de Pearson, $\rho = 0.3$, verificou-se que a área da mama (e a percentagem de ocupação) tem uma fraca correlação com a classificação dada a cada imagem. A construção do classificador, para além de servir de base a todos os testes apresentados, serviu também o propósito de criar uma interface interativa, passível de ser utilizada como ficheiro executável, sem necessidade de instalação de nenhum software. Esta aplicação permite que o utilizador carregue imagens de mamografia, exclua *background* desnecessário para a análise da imagem, extraia features, teste o classificador construído e dê como output, no ecrã, a classe correspondente à imagem carregada. A análise de risco de desenvolvimento da doença foi conseguida através da análise visual da variação dos valores das features de textura ao longo dos anos para um pequeno set (N=11) de mulheres. Esta mesma análise permitiu descortinar aquilo que parece ser uma tendência apresentada apenas por mulheres doentes, na mamografia imediatamente anterior ao diagnóstico da doença. Todos os resultados obtidos são descritos profundamente ao longo deste documento, onde se faz, também, uma referência pormenorizada a todos os métodos utilizados para os obter. O resultado da classificação feita apenas com as features de textura encontra-se dentro dos valores referenciados no estado-da-arte, indicando que o uso de features de textura, por si só, demonstrou ser profícuo. Para além disso, tal resultado serve também de indicação que o recurso a toda a imagem de mamografia, sem o trabalho árduo de definição de uma Região de Interesse, poderá ser utilizado com relativa segurança. Os resultados provenientes da análise do efeito da densidade e da área da mama, dão também confiança no uso do classificador. A interface interativa que resultou desta primeira fase de trabalho tem, potencialmente, um diferenciado conjunto de aplicações: no campo médico, poderá servir de auxiliar de diagnóstico ao médico; já no campo da análise computacional, poderá servir para a definição da *ground truth* de potenciais datasets que não tenham legendas definidas. No que diz respeito à análise de risco, a utilização de um dataset de dimensões reduzidas permitiu, ainda assim, compreender que existem tendências nas variações das features ao longo dos anos, que são especificas de mulheres que desenvolveram a doença. Os resultados obtidos servem, então, de indicação que a continuação desta linha de trabalho, procurando avaliar/predizer o risco, deverá ser seguida, com recurso não só a datasets mais completos, como também a métodos computacionais de aprendizagem automática.

Palavras-Chave: Cancro da Mama, Imagem Médica, Características de Textura, Aprendizagem Automática, Análise de Risco

Table of Contents

List of Figures

List of Tables

List of Abbreviations

%PD	Percent Mammographic Density
ACR	American College of Radiology
AUC	Area Under the Curve
BC	Breast Cancer
BI-RADS	Breast Imaging-Reporting and Data System
BMI	Body Mass Index
CBIS-DDSM	Curated Breast Imaging Subset of Digital Database for Screening Mammography
CC	Cranio-Caudal
DA	Discriminant Analysis
FDA	Food and Drug Administration
FMP	First Moment of the Power Spectrum
FPR	False Positive Rate
GLCM	Gray Level Co-Occurrence Matrix
k-NN	k Nearest Neighbor
LBP	Local Binary Pattern
LCIS	Lobular Carcinoma *in situ*
LR	Logistic Regression
MI	Mutual Information
ML	Machine Learning
MLO	Medio-lateral Oblique
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
NGTDM	Neighbourhood Gray-Tone Difference Matrix
pSNR	Peak Signal-to-Noise Ratio
RLM	Run-Length Matrix
RMS	Root Means Square of the Power Spectrum
ROC	Relative Operating Characteristic
ROI	Region of Interest
SS	Structural Similarity
SVM	Support Vector Machine
TPR	True Positive Rate
US	Ultrasound

1. Introduction

Section 1 is divided into nine subsections. The first gives a motivation for the developed work, showing worldwide data concerning Breast Cancer (BC). Section 1.2 offers a clinical background, relate to breast anatomy, which is complemented in section 1.3, where breast diseases are analyzed. In subsection 1.4, genetic and environmental risk factors for the development of this disease are described, while in subsection 1.5 screening methodologies, and their advantages and disadvantages, are examined. The information studied in the previous referred subsection is deepened in subsection 1.6, where mammography is profoundly explored, in term of its physical principles, as in terms of the most common findings in this type of exam. Subsections 1.7 and 1.8 are more related to the purpose of these work, with feature extraction and machine learning (ML) introductions being made, respectively. Finally, subsection 1.9 presents the primary goals of this thesis.

1.1 – Motivation

One in eight women will be diagnosed with BC in their lifetime, with one in thirty-nine women dying from this disease, only in the United States of America. In the same country, in 2020, approximately 42 170 women were expected to die from BC and it was anticipated that approximately 30% of the cancers detected in women were BC [1]. Around 95% of cancers are due to genetic mutations that result from environmental or lifestyle factors, where the remaining percentage is related to inherited genes – with BRCA1/BRCA2 genes being responsible for most of cases of BC [2, 3].

BC diagnosis occurs either during a common screening program, before symptoms appear, or after women noticing breast changes. Screening programs are important for an early detection of BC - that is, in a more treatable stage - resulting in a decrease in mortality [1, 4].

The criterion that defines if a woman is eligible for screening is, normally, her age. Different countries have different recommendations on which age is the best to start screening; the USA states that women from age 45 to 54 should have a mammography once a year, while 55+ plus women should have a mammography once every two years. On the other hand, the UK National Health System says that only women between 50 to 71 should be screened, and only once every three years [5, 6].

Incidence rates of BC increased highly during the decades of 1980 and 1990, mostly due to the increase in mammography screening programs available – studies even point out a prevalence increase from 29% to 70% resulting from screening within the time frame from 1987 to 2000 [7]. This increase can be seen, for the case of invasive BC, in Figure 1.1. The recent increase in BC incidence is thought to be related to a higher Body Mass Index (BMI) and a diminished number of births per woman [8].

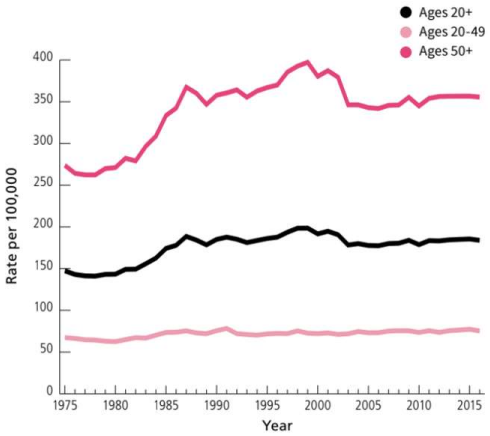

Figure 1.1 - Incidence Rate of Invasive Breast Cancer in the United States of America. [1]

In what concerns to the mortality of this disease, an extensive decline has been observed, with a drop of nearly 40% from 1989 to 2017, although having slowed down in recent years. This decline in mortality may be related not only to better treatments but also with best preventive tools [9]. Nonetheless, not all women were affected by this improvement in both diagnostic and treatment options, as it can be seen in Figure 1.2, with the descending trend in Non-Hispanic Black Woman not being has leaned as it happens in Non-Hispanic White Women. There are a great number of factors contributing for this difference, and while one might argue that there are unfortunate tumor characteristics specific to this ethnicity, social disadvantages must not be overlooked. Black women tend to have less access to high-quality diagnostic and treatment options, having a higher probability of being screened in institutions that don't have as many resources or are even nonaccredited at all. This can lead not only to a long period of time between examination and getting the results, as also to a poor assessment and consequently to a bad follow-up [1].

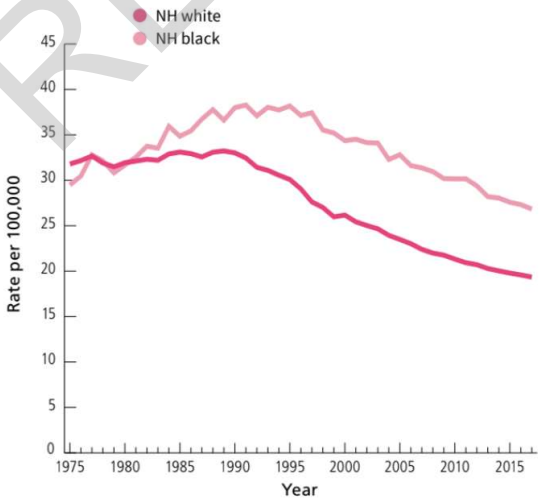

Figure 1.2 - Mortality Rates of Breast Cancer between Non-Hispanic white Women and Non-Hispanic Black Women. [1]

The treatment for invasive BC involves a panoply of options that drive from surgery, like breast removal (mastectomy), to radiation therapy, that is normally done after surgery to eliminate abnormal cell that remained in the breast area. Besides that, chemotherapy; hormonal therapy, to block the effects of estrogen, that stimulates BC; and Immunotherapy, that is the use of drugs to stimulate one's immune system to recognize and destroy cancer cells, are other possible lines of treatment [1].

Although there are multiple screening programs, they might not serve all women. Some younger women may be at higher risk of developing BC than women in their fifties and, despite that, these women are not eligible for screening. With that in mind, the perfect screening program should not consider age as the *only* risk factor that determines when to screen women.

The current medicine paradigm is one of preventive and personalized care and, for that reason, methodologies that allow the prevention, or an early diagnosis, of this disease, in a personalized fashion, are highly valuable. Although screening programs aim to prevent further effects of developing BC by making an early diagnosis, different factors like screening periodicity, may disrupt this goal. A technique that allowed an automatic detection of the disease, using a mammogram, would aid the doctors in the diagnostic process, making it easier and faster. An application like this one would be of great importance, since time is imperative in what concerns to cancerous diseases diagnosis and treatment. Besides that, a technology that allowed, based on a mammogram and other epidemiological factors, to retrieve a risk of future development of BC, would enable doctors to, in a personalized fashion, determine how and when to screen women, and if some preventive course of treatment should be made.

1.2 – Breast Anatomy

The breasts are an accumulation of tissue that is located in the pectoral region, on the anterior thoracic wall, overlying the *pectorales major* – a thick muscle located on the chest - ranging from the second to the sixth rib. The breast is composed of mammary glands, skin, and connective tissue. The mammary glands are modified sweat glands that develop during pregnancy, are maintained during lactating period and atrophy after this period ends. These mammary glands consist of fifteen to twenty lobes, that divide into lobules – the secretory part of the glands – that are located around the nipple. Each lobe is emptied by lactiferous ducts that dilate, forming lactiferous sinus that drain onto the nipple. Internally, the breast is composed by adipose and glandular tissue. In non-lactating women, the breast is mainly composed of fat while glandular tissue occupies a great part of the breast in lactating women. The breast is connected to the superimposed skin and to the pectoralis muscle through Suspensory (Cooper) ligaments that are fibrous bands of connective tissue. Vascularization is extensive across breast tissue - both from blood and lymph vessels – and the dermal blood capillaries and nerves are closer to the surface in the region that surrounds the nipple - the areola [10, 11].

Figure 1.3 depicts a general view of breast anatomy.

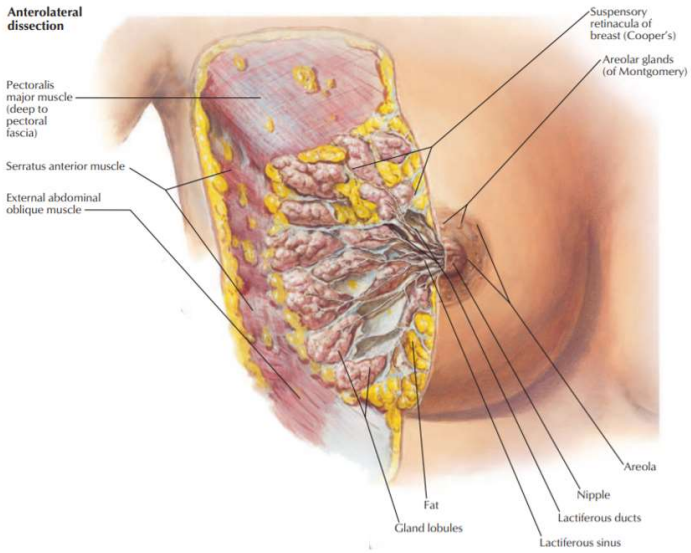

Figure 1.3- Breast Anatomy. Adapted from [12]

Concerning evolution of breast anatomy, at birth, only the main lactiferous ducts are fully developed. The mammary glands will only developed at its fullest during puberty, where the breast enlarges due to hormones like estrogen and progesterone, that stimulate the development of epithelial and connective tissue. Not only do the mammary glands develop during puberty, but during this time period there is also the deposition of adipose tissue, which contributes to breast enlargement. The breast will only complete its development during pregnancy, where the organ increases not only in volume but also in density, under the influence of various hormones. Besides breast enlargement, there are other anatomical modifications, which include vein dilation and a darkening in the pigment that constitutes the nipple and the areola. In addition, glandular tissue will start to occupy a great portion of the breast itself. Immediately after the end of the lactating period, the breast involutes, with glandular components and ducts starting to degenerate, resulting a diminished breast size. Finally, during menopause, ducts and glandular components will further deteriorate, making a menopausal breast to be primarily constituted by adipose tissue. However, over the years, there will be a decrease in this adipose tissue, which diminishes even further breast size and, moreover, the relaxation of the cooper ligaments may occur, resulting in breast ptosis [13].

1.3 – Breast Conditions

BC can be defined as a group of diseases where there is an uncontrolled cell division in breast tissue, usually beginning in the ducts or the lobules. There are two main types of BC: *In Situ* Carcinoma and Invasive Carcinoma.

As for In Situ Carcinoma, initially there were two classes: ductal carcinoma and lobular carcinoma; however, lobular carcinoma in situ is considered benign, although correlated with high risk BC development. Nonetheless, this condition does not have the potential to advance to an invasive stage. In contrast, Ductal in Situ Carcinoma is, generally, directly related to a development of invasive cancer. Nonetheless, while in an *in situ* stage, there is no proliferation of cells outside the location from where they were originated.

On the other hand, Invasive Carcinomas, which represents roughly 80% of all breast cancers, are diseases where cells emerge from the ducts or lobules where they first started to proliferate

uncontrollably. Invasive Carcinoma is a group of diseases with a vastly number of subtypes, depending on molecular characteristics [1].

Besides BC there are also benign conditions that can be related to a risk of developing malignancy. Some authors [14] divide these diseases in terms of risk of developing for BC:

- with no increased risk and minimal cell proliferation there are, for example, benign tumors like fibroadenomas, solitary papilloma, and sarcoidosis.

- with a small increase in risk and having cell proliferation without cell abnormalities there are diseases like ductal hyperplasia and sclerosing adenosis.

- with a moderate increase in BC with abnormal cell proliferation there are only two benign conditions: atypical ductal hyperplasia and atypical lobular hyperplasia.

1.4 – Breast Cancer Risk Factors

The question resides in what risk factors are not being considered when choosing the best screening option. Age is one of the best documented risk factors, with the incidence of BC being extremely low before the age of 30 and having a linearly increase until the age of 80 [15]. Body Mass Index has also been shown to be a potential risk factor for the development of BC but only in post-menopausal [15, 16]. Prior history of neoplastic or hyperplastic breast disease also presents itself as a risk factor for the development of BC. When it comes to family history, a woman who had a first-degree relative with BC when they were 50 years or older, is almost twice at risk of developing BC than a woman with no family history of BC [15]. Early menarche, late first full-term pregnancy and late menopause are three major risk factors for BC [17]. Normally, the earlier the age of the first menarche, the higher the cancer risk. The fact that both women with early menarche and later menopause are at higher risk of BC, can lead to the conclusion that prolonged exposure to estrogen is also a risk factor for this disease [17]. Longer duration of the breastfeeding period is associated with a diminished risk of BC, in comparison with women that had shorter breastfeeding periods.

Use of oral contraceptives also puts women at higher risk of developing BC [18]. As it was previously discussed, the existence of the BRCA1/BRCA2 mutated gene in women genotype puts them at higher risk of BC, compared to women who do not possess that gene [19].

Besides these risk factors, in 1976, John Wolfe, started studying the association between breast parenchyma patterns and BC. Wolfe showed that a prominent duct pattern helps to classify a woman as having higher risk than average for developing BC. Wolfe also stated that it is possible to predict which women will develop BC and which are less likely to develop it based only on the parenchymal pattern [20-23]. The studies conducted by Wolfe helped to define a classification of BC risk, based only on breast composition:

- N1, lowest risk, parenchyma composed primarily of fat, no ducts visible.
- P1, low risk, parenchyma chiefly fat with prominent ducts (< a quarter of the breast volume).
- P2, High risk, severe involvement of prominent duct pattern - occupying more than a quarter of the breast volume.
- DY, Highest risk. Severe involvement with dysplasia.

Many descriptors of these texture patterns have been documented. Mammographic density is one of those descriptors, normally represented numerically by percent mammographic density (%PD), that is also highly associated with an increased risk of BC [24-26]. In