Madalena Silva Ramos Madureira Simões

Bachelor of Science in Biomedical Engineering

Automatic Breast Density Classification on Tomosynthesis Images

Dissertation submitted in partial fulfilment of the requirements for the degree of

Master of Science in

Biomedical Engineering

Adviser: Professor Nuno Matela, Professor Auxiliar

Faculdade de Ciências da Universidade de Lisboa

Co-adviser: Professor Pedro Vieira, Professor Auxiliar

Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

November 2020

[Automatic Breast Density Classification on Tomosynthesis Images]

Acknowledgments

"No man is an island" (John Donne)

First and foremost, I would like to thank my adviser, Professor Nuno Matela, for his support and encouragement throughout the whole process of this study, and a special remark about the seamlessly easy manner in which he provided adaptation to an unprecedented time that broke when we had just started working on this thesis. I would also like to thank my co-adviser, Professor Pedro Vieira, for his continuous support and availability to discuss my doubts at all times.

To my college friends, Beatriz, Inês, Joana, Lourenço, Miguel and João, I would like to give thanks for motivating and reassuring me through such strange times that we were all living when developing our thesis and for keeping me grounded when times were hard.

To my childhood friends, Catarina, Filipa and Carolina Silva, I want to thank you for reminding me of who I am when I was unsure and feeling stressed, and for helping me unwind when rest was needed.

I would like to thank my family for pushing me and for supporting me unconditionally throughout this process. Thank you to my father, Nuno, for always being interested in what I was doing and for allowing me to explain it to him a thousand times. Thank you to my sister, Rita, for nurturing me and helping me by reading my work repeatedly and always providing an honest feedback.

Finally, I would like to express special gratitude towards the women in my life that inspired me to continue working towards a palpable result by being survivors of this disease. Thank you to my grandmothers, Avó Gette and Avó Né, and my mother, Teresa, for being true examples of determination, willpower and strength.

Abstract

Breast cancer (BC) is the type of cancer that most greatly affects women globally hence its early detection is essential to guarantee an effective treatment. Although digital mammography (DM) is the main method of BC detection, it has low sensitivity with about 30% of positive cases undetected due to the superimposition of breast tissue when crossed by the X-ray beam. Digital breast tomosynthesis (DBT) does not share this limitation, allowing the visualization of individual breast slices due to its image acquisition system. Consecutively, DBT was the object of this study as a means of determining one of the main risk factors for BC: breast density (BD). This thesis was aimed at developing an algorithm that, taking advantage of the 3D nature of DBT images, automatically classifies them in terms of BD. Thus, a quantitative, objective and reproducible classification was obtained, which will contribute to ascertain the risk of BC.

The algorithm was developed in MATLAB and later transferred to a user interface that was compiled into an executable application.

Using 350 images from the VICTRE database for the first classification phase – group 1 (ACR1+ACR2) versus group 2 (ACR3+ACR4), the highest AUC value of 0,9797 was obtained. In the classification within groups 1 and 2, the AUC obtained was 0,7461 and 0,6736, respectively. The algorithm attained an accuracy of 82% for these images. Sixteen exams provided by Hospital da Luz were also evaluated, with an overall accuracy of 62,5%.

Therefore, a user-friendly and intuitive application was created that prioritizes the use of DBT as a diagnostic method and allows an objective classification of BD. This study is a first step towards preparing medical institutions for the compulsoriness of assessing BD, at a time when BC is still a very present pathology that shortens the lives of thousands of people.

Keywords: breast density • breast tomosynthesis • automatic classification • breast cancer.

Resumo

O cancro da mama (CM) é o tipo de cancro que mais afeta mulheres globalmente sendo a sua deteção precoce fundamental para um tratamento eficaz. Apesar de a mamografia digital (MD) constituir o principal método de deteção do CM, apresenta baixa sensibilidade com aproximadamente 30% de casos positivos indetetados devido à sobreposição do tecido mamário quando atravessado pelo feixe de raios-X. A tomossíntese digital mamária (TDM) não partilha desta limitação, permitindo visualizar planos individuais da mama devido ao seu sistema de aquisição de imagens. Consecutivamente, a TDM foi objeto deste estudo enquanto meio para determinar um dos principais fatores de risco de CM: a densidade mamária (DM). Esta tese visou desenvolver um algoritmo que, aproveitando a natureza 3D das imagens de TDM, as classificasse automaticamente quanto à DM. Assim, obteve-se uma classificação quantitativa e reprodutível que contribuirá para averiguar o risco de CM.

O algoritmo foi desenvolvido em MATLAB e transferido para uma interface de utilizador que foi compilada para uma aplicação executável.

Utilizando 350 imagens da base de dados VICTRE para a primeira fase de classificação – grupo 1 (ACR1+ACR2) *versus* grupo 2 (ACR3+ACR4), obteve-se o valor de AUC mais elevado de 0,9797. Na classificação dentro dos grupos 1 e 2, a AUC obtida foi de, respetivamente, 0,7461 e 0,6736. O algoritmo apresenta uma precisão de 82% para estas imagens. Foram também avaliados 16 exames fornecidos pelo Hospital da Luz tendo sido obtida uma precisão de 62,5%.

Desta forma, desenvolveu-se uma aplicação acessível e intuitiva que prioriza o uso de TDM para diagnosticar CM e permite uma classificação objetiva da DM. Este estudo constitui um primeiro passo para preparar as entidades de saúde para a obrigatoriedade de avaliação da DM, numa altura em que o CM é ainda uma patologia muito presente que encurta a vida de milhares de pessoas.

Palavras-chave: densidade mamária • tomossíntese mamária • classificação automática • cancro da mama.

Contents

List of figures

List of tables

Abbreviations

FCT	Faculdade de Ciências e Tecnologia
UNL	Universidade Nova de Lisboa
BC	Breast Cancer
DM	Digital Mammography
DBT	Digital Breast Tomosynthesis
BD	Breast Density
3D	Three-dimensional
CM	Cancro da Mama
MD	Mamografia Digital
TDM	Tomossíntese Digital Mamária
DM	Densidade Mamária
FBP	Filtered back projection
BI-RADS	Breast Imaging Reporting and Dated System
2D	Two-dimensional
FCM	Fuzzy C-Means
GLCM	Grey-Level Co-occurrence Matrix
kNN	k-Nearest Neighbour
VICTRE	Virtual Imaging Clinical Trials for Regulatory Evaluation
ACR	American College of Radiology
ROI	Region of Interest
GLRM	Grey Level Run-length Matrix
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
DA	Discriminant Analysis
NB	Naïve Bayes
DT	Decision Tree
SVM	Support Vector Machine
RF	Random Forest

1

Introduction

In this Chapter, a context and motivation regarding the main issue approached in this work – breast cancer diagnosis – is presented, as well as an overview of important theoretical topics related to it and a review of the literature on the matter. In the end, the proposal of this work is disclosed.

1.1. Context

Breast cancer is the type of cancer that most greatly affects women globally. In Portugal it is the form of cancer responsible for the highest amount of mortality among women, accounting for approximately 16% of all cancer related deaths in 2018 [1]. Nevertheless, efficient treatment is possible, and its odds increase with early detection and consequent medical intervention. The main method of diagnosis is breast digital mammography (DM). However, this exam has many disadvantages such as its low sensitivity and specificity, resulting in about 30% undetected positive cases [2]. Additionally, this type of imaging method is also characterized by its high value of recall rate, remaining above the 5%-10% ideal target range [3]. These shortcomings are mostly due to DM's two-dimensional nature that usually results in the superimposition of healthy breast tissue when the x-ray beams pass through the patient's body, which can originate one of two unwanted effects: on the one hand, an abnormality indicating a possible cancer development can be obscured (constituting the false negative case) or, on the other hand,

this superposition can create pseudolesions known as summation artifacts, producing a false positive result and an unnecessary patient recall.

Nowadays, other imaging methods of breast cancer diagnosis have been considered such as digital breast tomosynthesis (DBT). DBT imaging has been progressively used more as an addition to DM (and in some cases actually replacing DM) to aid diagnosis as it enables the visualization of individual slices of the breast, reducing the impact of said superimposition. Conventional DM is a procedure in which the resulting image is created from a single x-ray exposure, providing a 2D view of the breast. DBT is a 3D medical exam considered as an extension to DM as it is performed using the same breast compression and positioning. The main improvements from DM to DBT are specificity and sensitivity. As a matter of fact, DM's 2D mammographic view is easily susceptible to tissue superimposition and thus masking effect, resulting in a decreased sensitivity value [4]. Hence, DBT is considered to be superior to DM given the fact that, by minimizing tissue superimposition, not only it enables a better differentiation between benign and malignant masses, thus revealing existing lesions and so increasing the efficiency of the diagnosis, particularly in denser breasts [5], but also reduces the amount of false-positive findings by discarding summation artifacts [6]. In Hovda et al [7], by comparing DM versus DBT breast images, it was found that for images obtained with a consecutive DBT, the recall value was maintained low and the amount of efficient detections rose from 4.6/1000 (for DM after DM) to 9.9/1000 (for DBT after DM). Moreover, it was found that the additional use of DBT increases cancer detection rate by 4% when compared to the single use of DM and reduces recall rate by 30-40% [6, 8]. The consequences listed might be due to the increase of lesion and background contrast ratio, that is, lesion conspicuity, and the reduction of the masking effect caused by tissue superimposition. In Figure 1.1, the visual differences between DM and DBT are illustrated and it is possible to verify a clear improvement in terms of image quality and distinctness from A) to B).

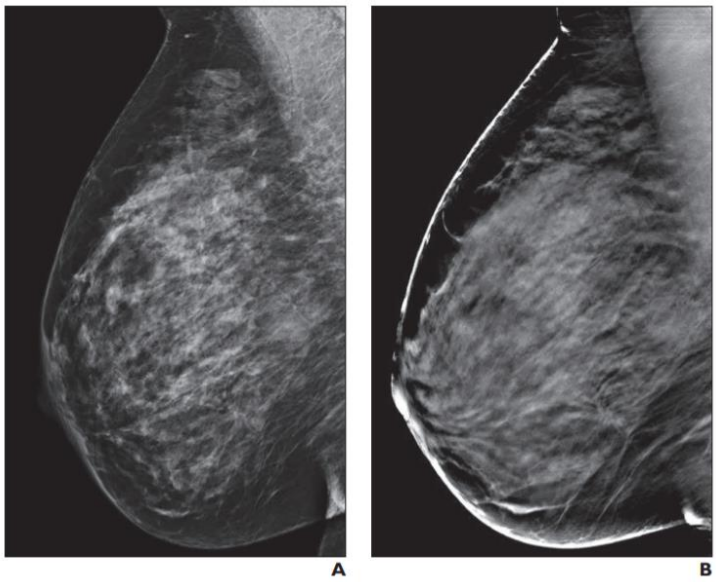

Figure 1.1: MLO view of breast in a A) DM vs. B) DBT. Adapted from [8].

1.2. Motivation

Breast density (BD) is one of the most relevant risk factors associated to breast cancer development [9]. As of April 2020, it is mandatory in 36 states in the USA to inform the patient about her BD status and the risk it poses, as it is seen as a breast cancer risk factor that can indicate the need for other complemental exams and subsequently patient recall [10].

In fact, it was determined that women in the two upper BD categories (ACR3 and ACR4, mentioned later in the BI-RADS metric segment) correspond to approximately 40% and 8% of DM performed, respectively [11]. Hence, almost half of examined women might beneficiate from having additional screening exams. Besides determining the patient's BD, it is also necessary to provide a clear and understandable communication of it and its implications so that the patient can make an informed decision about her options and whether additional exams are helpful, especially for women with dense breasts.

Despite its importance, BD classification in Portugal and most countries is determined through visual inspection by a specialist, thus being a very subjective methodology. Therefore, it is important to develop an algorithm that automatically classifies BD

on DBT images, taking advantage of DBT's 3D nature, resulting in a quantitative, objective and reproductible classification that allows for the ascertainment of breast cancer risk.

1.3. Breast Histology and Tissue Composition

All organs are mainly composed of four basic types of tissue: epithelium (epithelial tissue), connective tissue, muscle tissue and nerve tissue. Each of these has different functional properties and morphologic features [12]. More specifically, breast histology includes the following tissues: glandular tissue, which is a type of epithelial tissue that includes mammary lobes and ducts; adipose or fatty tissue, which is connective tissue made of fat cells that can be found between fibrous and glandular tissue; fibrous tissue, which is connective tissue that binds together adipose and glandular tissue and, finally, ductal tissue [13]. In Figure 1.2, it is possible to visualize the difference between adipose tissue, represented by the arrows, connective (dense) tissue (represented by A) and ducts (represented by B).

BD is intrinsically related to the ratio between the areas occupied by the nucleus (epithelial and non-epithelial tissue), collagen (connective tissue) and glandular structures. In fact, a correlation between BD and fibroglandular tissue, i.e. fibrous and glandular tissues, proliferation was found, adding to the fact that these two tissues are in fact the ones held accountable for the breast's radiological density [14]. Dense areas of the breast are histologically different from the less dense, with larger proportions of epithelial and stroma tissues and less fat [15, 16].

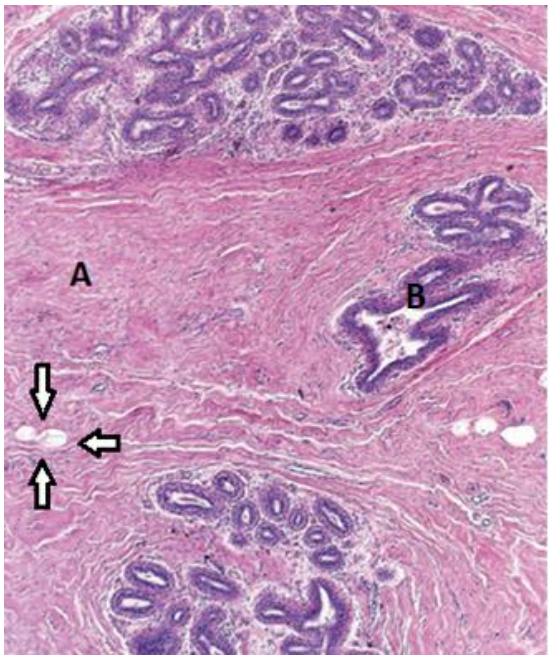

Figure 1.2: Histology of an inactive mammary gland. A) Dense connective tissue. B) Ducts. Adapted from [12].

1.4. BD and Breast Cancer Risk

From the moment that the classification of BD was a point of interest in the scientific community, not only several metrics were created, but also the correlation between the degree of BD and breast cancer risk was inferred. The first speculation came from the work of Wolfe et al [17] where, correlated to the four parenchymal patterns, it was observed an increase of as much as 37 times higher future cancer risk from N1 (the lowest risk) to DY (the highest risk) patterns. Shortly after, Boyd et al [18] demonstrated a lower dependency between BD and breast cancer risk than the one presented by Wolfe, indicating a 4 to 5 times increase between N1 and DY categories.

Throughout the years, the correlation between the amount of fibroglandular tissue and breast cancer risk has been studied with the aim to corroborate BD as a relevant risk factor to be considered. More recently, in Ghosh, K. et al [15] it was stated that women with fibroglandular tissue occupying 60-75% of total breast tissue have a risk of developing breast cancer of about 3 to 6 times higher than women with mostly fatty breasts. Additionally, when compared to women with fatty breasts, women with dense breasts are at least 3.5 times more likely to have an interval cancer, diagnosed less than 1 year after a negative screening mammogram [11].

As is known, cancer originates from the fast proliferation of abnormal cells that are intrinsically characterized by their speedy rate of cell division. In the work of Vachon, C. M. et [19] al it was observed a correlation between BD changes, i.e. fibroglandular tissue, and hormones and growth factors , as such: menopausal hormones, tamoxifen and, lastly, IGF-1 factor in pre-menopausal women. In fact, IGF-1 was featured in the work of Diorio et al [20] as being associated with the risk of developing breast cancer due to its influence on the morphogenesis of breast density. Thus, this type of tissue, which is the one associated with BD, is highly responsive to factors that stimulate cell division and so incurs in a higher chance of suffering carcinogenesis than less dense tissues. As a matter of fact, breast cancer originates from epithelial cells, which indicates that fibroglandular tissue areas are composed of more cells that are characterized by faster proliferation rates than other types of tissue. Hence, the relative abundance of this type of tissue on the breast is a critical component to determine whether the patient is in fact at a higher risk of initiating and developing breast cancer [19, 21].

Finally, in Wengert et al [22], it is mentioned that breasts classified with higher density categories are more often linked to the existence of larger breast tumours and advanced stages with lymphatic involvement which corroborates the involvement of BD in cancer risk.

1.5. DBT Technique

Like mentioned before, DBT is an emerging diagnosis technique more progressively used to detect breast cancer in its early stages when its chances of being eradicated are the highest. Although it is not available in all hospitals and facilities, this imaging exam was approved by the Food and Drugs Administration in 2011 and is being prescribed as a complementary exam to DM, as it provides a better visualization of the breast tissue, especially for women with dense breasts [23].

Its functioning is essentially similar to DM except for some particularities. While in DM two x-ray images are taken of the breast while it is compressed between a clear plastic paddle and an imaging detector, in DBT an x-ray tube moves along an arc positioned over the compressed breast and it captures several low-dose full field exposures of it from different angles at pre-set intervals [24]. The setup differences between conventional DM and DBT are illustrated in Figure 1.3. The resulting stack of 2D projections

of the imaged breast, which is typically in the craniocaudal (CC) or mediolateral oblique (MLO) views, is then converted and reconstructed into a set of 3D tomographic thin slices parallel to each other and the total number of reconstructed slices varies according to the thickness of the compressed breast.

Just like the angle range, tube motion and time needed to obtain a whole set of projection images, there are several reconstruction algorithms and they depend on the manufacturer of the equipment [25]. Typically, the algorithm most used to reconstruct tomosynthesis images is filtered back projection – FBP – also commonly employed for CT images. This algorithm is an improvement from the basic backprojection method by adding in the filtering step before the backprojection step. Firstly, the set of projections is filtered with a high pass filter, namely a ramp filter, which sharpens the edges of the image by removing frequencies lower than the cut off threshold, and it takes place in the frequency domain. Next is the backprojection step which takes place in the spatial (image) domain, is repeated for all pixels and can be performed by 15° or 30° scan angles. Each row of the 2D projections represents the sum of all counts along a straight line through the depth of the object being imaged, i.e., the object's 2D Fourier Transform along that line [26]. The backprojection technique runs the obtained filtered projections back through the image, performing the 2D inverse Fourier Transform and so reconstructing the original object [27].

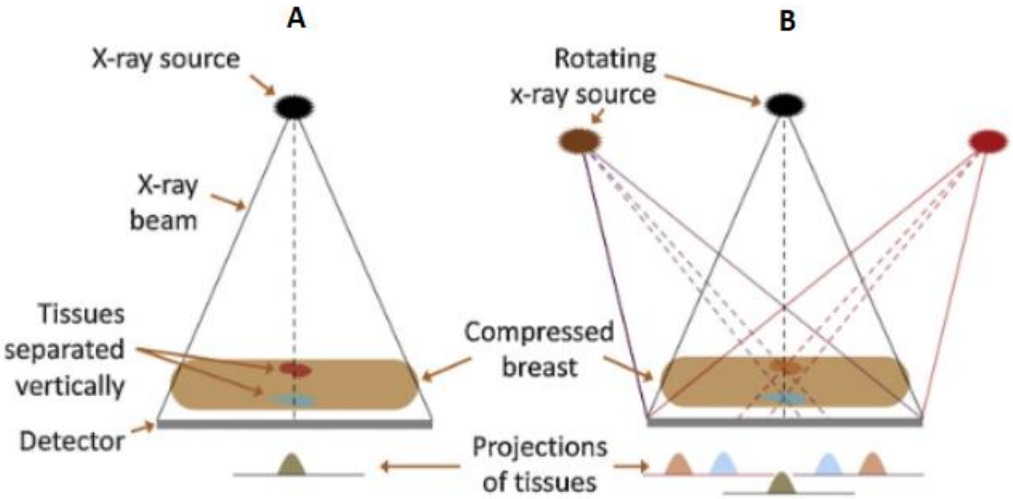

Figure 1.3: Setup differences between conventional DM (A) and DBT (B). Adapted from [28].

1.6. Linear Attenuation Coefficients and BD

Different parts of the body absorb the x-rays differently, resulting in different appearances in the image. Hence, breast tissue composition is easily distinguished given the fact that different tissues appear differently in radiographic images. This is due to their different linear attenuation coefficients, μ. This constant represents the fraction of a beam of x-rays or gamma rays that is absorbed per unit thickness of the given material. Most commonly, the constant determined to characterize attenuation is the mass attenuation coefficient, μ/ρ, easily converted to the linear attenuation coefficient given the tissue density. In the work of Chen et al [29], linear attenuation coefficient values were determined for three types of tissues – fatty, fibrous and tumoral - using synchrotron radiation computed tomography. For a *kVp* of 15 keV, it was obtained a mean μ value of 0,794 and 1,659 for adipose and fibrous tissue, respectively.

Following the exposed above, a large μ value means that the beam is easily attenuated throughout its path across the material, resulting in the latter absorbing most of the rays. For example, dense bone (higher value of μ) absorbs much of the radiation while soft tissue, such as organs, which have a lower value of μ, do not attenuate the x-rays as much. As a result, bones appear white on the x-ray and soft tissue shows up in shades of grey. As follows, the distinction between two main tissue types is done according to the intensity they display in the image where less dense areas appear darker corresponding to fatty tissue that has a smaller X-ray linear attenuation coefficient, as it is confirmable in Figure 1.4. Contrarily, fibroglandular tissue, similarly to tumoral tissue, has a higher density value so it appears brighter in the image, for the same value of peak Kilovoltage value, kVp. The relative prevalence of these bright areas portraying denser regions originates the degree of BD [30, 31].

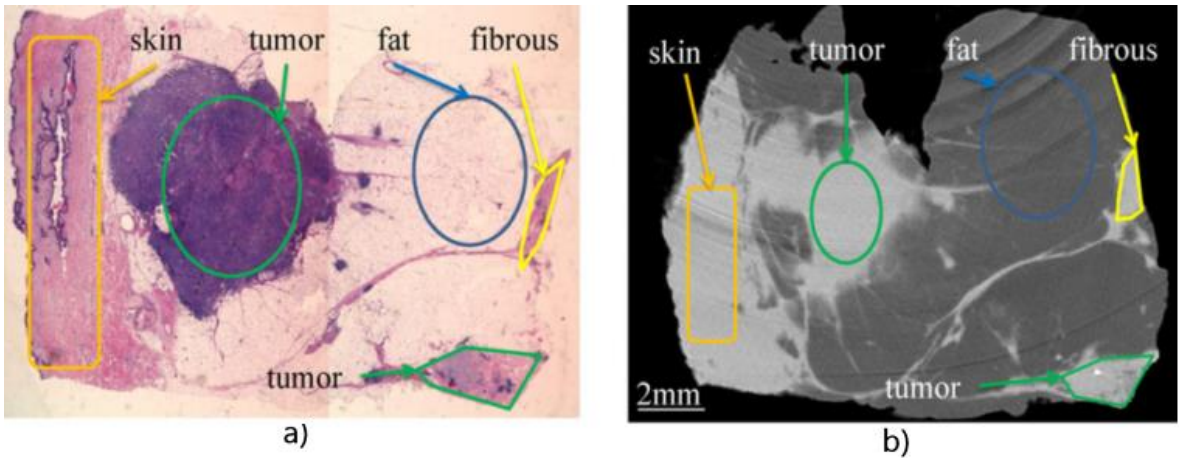

Figure 1.4: Breast histology and imaging. a) Histological image. b) SR CT image at 19 keV. Adapted from [18].

1.7. Masking effect

As stated before, breast fibroglandular tissue superimposition can cause masking, known as the consequence of obscuring tumours due to the fact that both types of tissue mentioned have the same x-ray attenuation properties [32, 33]. This effect is most greatly affected by the value of BD. In Figure 1.5 it is visible the effect mentioned where in a) it is depicted a DM image showing a mass that is barely distinguishable whereas in b) is represented a DBT image that provides a sharper distinction of the said mass [34]. A higher value of BD consequently means a higher probability of the occurrence of masking effect in some extent, hence reducing DM's sensitivity and limiting the possibility of an early detection. To add to this, it is also a given that women with denser breasts are at a higher risk of developing breast cancer thus aggravating the masking effect [32, 35]. As follows, DBT constitutes a relevant and more efficient diagnosis exam given the fact that it does not suffer masking effect, in opposition to DM.

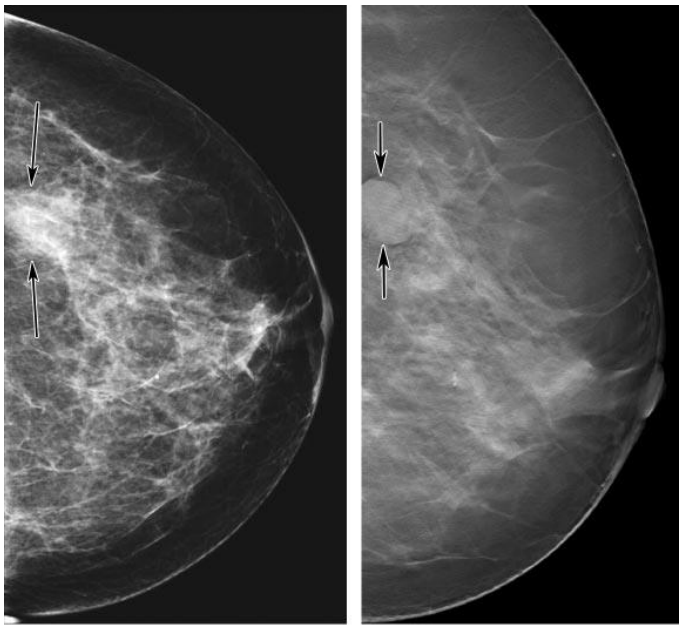

Figure 1.5: Comparison of mass circumscription in a) DM image, b) DBT image [34].

1.8. Literature Review

1.8.1. BD classification metrics

1. Wolfe scale

BD was first mentioned in Wolfe's work in 1976 where he focused on the correlation between breast parenchymal patterns and cancer risk. Later, Wolfe et al created a qualitative grading system that featured four different categories based on the visual interpretation of the breast parenchyma in a mammogram, namely the amounts of fat and other types of tissue present. The 4 categories are termed as such:

- N1, attributed to a breast constituted mainly of fat with a trabeculated appearance, considered to be "normal";
- P1, assigned to a breast composed mainly of fat with prominent ducts;
- P2, representing a breast with a more intense prominent duct pattern;
- DY, credited to a breast with a higher density of the parenchyma with possibly a minor component of prominent ducts [17, 36].

2. *Boyd scale*

Boyd et al introduced the first quantitative metric for BD after the strong correlation between BD and cancer risk was presented by Wolfe et al [37]. Boyd classification system was the first attempt at having an objective classification system for BD and it comprised six categories, also known as SCC (six-category classification) [38, 39], which is described in Table 1.1:

Table 1.1: Boyd BD classification metric.

Category	SCC_1	SCC_2	SCC_3	SCC_4	SCC_5	SCC_6
BD %	0%	<10%	10% ≤ BD < 25%	25% ≤ BD < 50%	50% ≤ BD < 75%	≥ 75%

3. *Tabár scale*

Later, inspired by Wolfe's work, Tabár and Dean et al [40] proposed a model to classify mammograms in five histological patterns based on four mammographic building blocks that made up the normal breast composition: nodular (N) areas mainly correspond to terminal ductal lobular units; linear (L) areas correspond to either ducts or fibrous or blood vessels; homogeneous (H) structureless areas correspond to fibrous tissues; radiolucent (R) areas related to adipose fatty tissues [18, 41]. Mammograms were then classified as one of five risk categories according to the percentage of each building block observed, according to Table 1.2:

Table 1.2. Tabár BD classification metric.

Category	T_I	T_{II}	T_{III}	T_{IV}	T_V
N area %	25%	2%	2%	49%	2%
L area %	15%	14%	14%	19%	2%
H area %	35%	2%	2%	15%	89%
R area %	25%	82%	82%	17%	7%

Through visual inspection of Table 1.2, it is clear that T_I, T_{II} and T_{III} represent the lowest risk due to their smaller percentage of H densities, i.e., fibrous tissue, when compared to T_{IV} and T_V. T_{II} corresponds to a mammogram with an increase of radiolucent fatty tissue. Moreover, although T_{III} has the same composition as T_{II}, it is characterized by retroareolar prominent ducts that are often associated with periductal fibrosis. The next category, $T_{IV,}$ is characterized by its mammogram being more difficult to read than the latter two and resistant to the process of involution. Finally, T_V represents the highest risk given the fact that it has the highest percentage of fibrous tissue present (H areas) and its mammogram presents more obscurity in terms of the appearance of pathological lesions.

4. BI-RADS

In 1995, a new classification metric for BD was presented by the American College of Radiology, integrated in the BI-RADS - Breast Imaging Reporting and Dated System – intended for DM images. The metric is constituted by four different categories that are represented in Figure 1.6, such as: A) fatty breast – composed mainly of fatty tissue (density<25%); B) scattered breast – exhibiting fibroglandular tissue areas disperse throughout the mammary tissue (26% ≤ density < 50%); C) heterogeneous breast – characterized by a heterogenous density (51% ≤ density < 75%); D) dense breast – characterized by the highest percentage of fibroglandular tissue present (density ≥ 76%) [18].

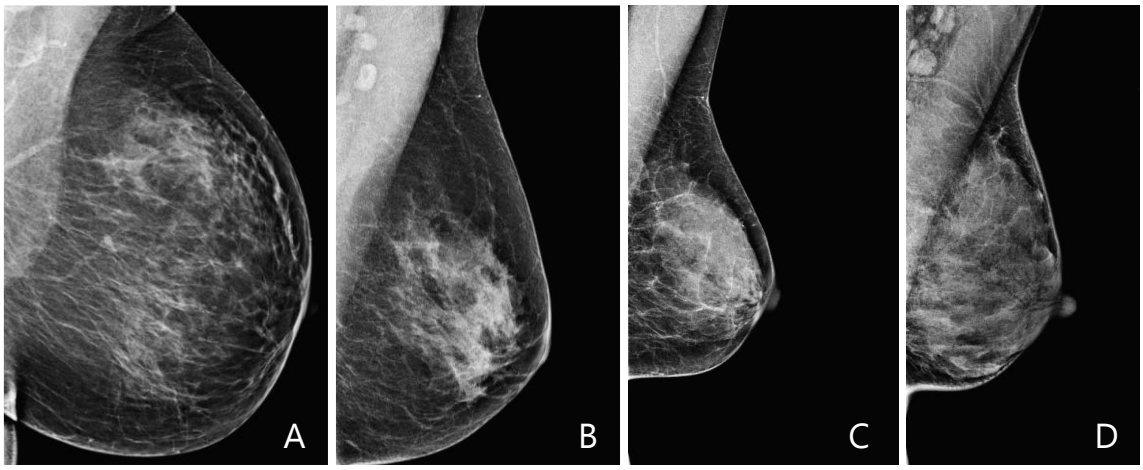

Figure 1.6: BI-RADS categories. Adapted from [18].

5. CAD systems

Recently, the fifth edition of BI-RADS atlas has been revised and the four categories, A, B, C and D, now correspond to ACR-MG-a, ACR-MG-b, ACR-MG-c and ACR-MG-d, respectively [22]. However, the most common notation used is, respectively, ACR1, ACR2, ACR3 and ACR4.

Several algorithms have been developed to determine BD quantitatively in DM images, commonly referred to as CAD, or Computer-Aided Detection.

There are essentially two types of CAD methods: semi-automated and automated [35]. Semi-automated algorithms are based on segmentation and thresholding techniques that aim to determine BD in terms of the percentage of existing dense tissue in the totality of the breast. These algorithms are characterized by being labour-intensive and time consuming and they include planimetry and interactive thresholding methods [31]. Planimetry consists of the direct measurement of the areas of dense tissue identified on the DM by tracing it with a planimeter. Interactive thresholding software, for example the Cumulus software [42], features the identification of desired areas of the breast and consequent selection of their threshold grey levels with a pointing device. The threshold levels typically correspond, on a first level, to the separation between the breast and the background and then, on the second selection, segmentation between dense and fatty tissue. These selections are progressively highlighted and thus the contrast between regions is accentuated.

Automated methods comprise several categories such as: calibration techniques, texture techniques, model-based methods and more specific algorithms like MedDensity and AutoDensity [35]. The first two categories calculate BD using mathematical and physical modelling that contemplates calibration and texture approaches. Model-based techniques examples are Volpara, that calculates the volumetric BD by inferring the volume of fibroglandular tissue present in the totality of the breast, represented in Figure 1.7, and Quantra, which operates similarly to Volpara with the exception that it first automatically segments the breast from the background, then determines the energy deposited at the detector in each pixel and, with that, calculates the thickness of fibroglandular tissue and the Standard Mammographic Form (SMF). With this information, Quantra calculates the height of non-fatty tissue that corresponds to each pixel by using image information like height of compression and exposure time [43, 44]. Lastly, MedDensity and

Autodensity are automated thresholding techniques. MedDensity is a software that was produced by Tagliafico et al [45], which performs automatic thresholding based on spatial information and then segments the breast into adipose and fibroglandular tissue. Consequently, BD is calculated as the percentage of fibroglandular tissue pixels present in the total breast area. AutoDensity, developed by Nickson C. et al [46], is a software that works similarly to Cumulus except that it automatically segments the breast area in the DM and then identifies distinctly white tissue that is perceived as dense and this way classifies breast density.

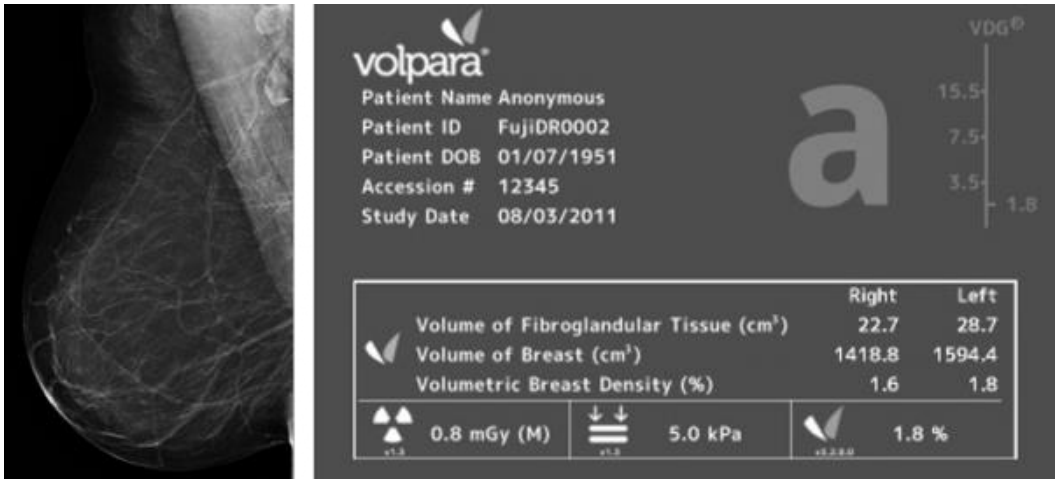

Figure 1.7: Volpara software. Adapted from [47].

These methods are mainly praised for their reproducibility quality, although presenting a major disadvantage due to the fact that, because their calculations are based in 2D DM images, there is a limitation in terms of depth. This drawback is accounted for by modelling the variation in breast tissue thickness, which may result in errors and assumptions that are not true for every patient. Hence, the algorithms presented are not yet translatable to DBT images given their 3D nature.

1.8.2. Classification methods

The quantitative applications formerly presented encompass various methods which commonly have a similar first step which is segmentation. Segmentation algorithms aim to not only segment the breast from the background but also to properly distinguish the various types of tissue present in the breast, namely fibroglandular and adipose tissues. One of the most used segmentation techniques in breast imaging is the

k-Means method, an unsupervised clustering algorithm that divides the intensities in the image based on cluster centroids, creating different clusters that express similarity between pixels [48].

An improved version of k-Means that is increasingly gaining more highlight in breast segmentation is fuzzy C-means clustering (FCM), which mainly differs from k-Means in terms of enabling each pixel to belong to more than one cluster. Effectively, the FCM method has been regularly used in various studies to segment breast tissue in DM like in the work of Moon et al [49] where it was used to distinguish fibroglandular tissue from fat to compute the breast density, much like the presented work. Although FCM is greatly used in the field of breast segmentation, it can also be applied in other areas such as in the work of Caldas et al [50] where FCM was used to evaluate variations in walking parameters by clustering groups with similar gait patterns so as to detect possible shifts in physical capacity that might occur with age. Lastly, another example that illustrates the usage of the FCM algorithm is the work of Rundo et al [51] which depicts FCM techniques as the means to cluster three different types of tumoral tissues.

Subsequentially, a texture analysis is typically computed and it consists in calculating a set of textural features from the resulted segmented tissues that best characterize each type of tissue mostly based on grey-level variations presented in the image [52]. The most used types of textural features in this frame of work are Haralick features, extracted from the GLCM matrix; statistical features; histogram-based features and run-length features [53]. The goal is to first calculate the maximum number of features possible while also not compromising computing time and then perform feature selection to ascertain which features are most relevant to each classification phase and prevent overfitting. There are several feature selection methods that fall in one of three types: supervised, semi-supervised or unsupervised. The first type identifies the relevant features for the classification problem having access to labelled data. On the contrary, unsupervised methods score features without the knowledge of classes/labels, like Principal Component Analysis. Finally, in semi-supervised methods both labelled and unlabelled data are used for learning. In terms of evaluation criterion, there are mainly four types: embedded methods, that involve a learning algorithm and model fitting; wrapper methods, which choose to add/eliminate features by evaluating the gain in accuracy; filter methods, which infer the relevance of features based only on their intrinsic characteristics

and, finally, hybrid methods, that originate from either combining two of the approaches formerly mentioned, two different techniques from the same criterion methods or two different feature selection methods [54].

Lastly, these algorithms culminate in a classification phase which typically is composed of one or a set of machine learning classifiers such as kNN, Naïve Bayes, Random Forest and Neural Networks. These classifiers go through a specific process [55], depicted in Figure 1.8, which involves them being trained with a set of images characterized by their class/label and features and then tested with never seen images in order to determine the category of BD, out of the four possible ones described by BI-RADS: fatty (ACR1), scattered (ACR2), heterogeneous (ACR3) and dense (ACR4) breast.

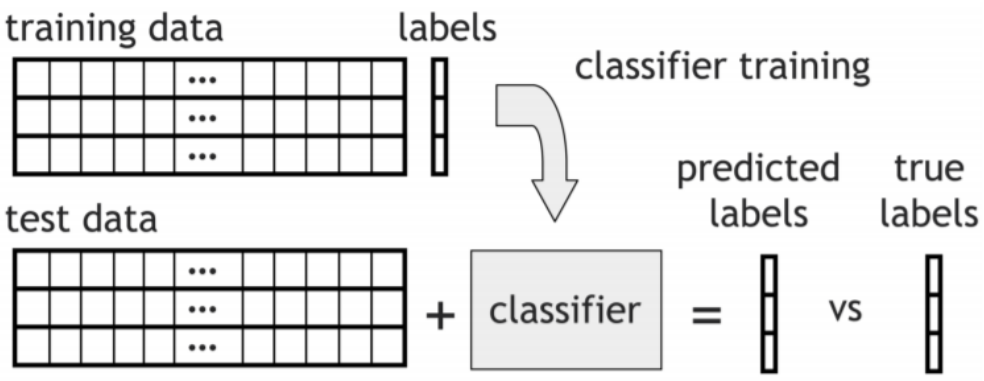

Figure 1.8: Training and testing a Machine Learning classifier. Adapted from [55].

1.9. Proposal

The proposal of this work is to develop an application that is destined to specialists with the aim to objectively determine the patient's BD value and use it as a proper breast cancer risk factor, aiding in the decision to run additional examinations and recall the patient. As it was stated before, currently the only objective methods that determine BD are intended for DM images only. On the contrary, this interface is destined to be used for DBT images as they provide a clearer understanding of the breast's anatomy without the disadvantage of having tissue superimposition and so aggravating the specialist's diagnosis. The interface built in MATLAB Guide will enable the specialist to load the set of DICOM images corresponding to the several DBT slices that make up the whole exam's

volume, visualize them and then compute the several steps to reach an objective BD classification.

2

Materials and Methods

In this Chapter, a description of the materials used to develop the present work is made, as well as a depiction of all the methods and techniques that encompass it and that were used to create the final product - the application.

2.1. Materials

All of the initial training and testing of the interface were performed using a specific database named VICTRE, Virtual Imaging Clinical Trials for Regulatory Evaluation, provided by The Cancer Imaging Archive (TCIA), formerly known as National Biomedical Imaging Archive open source (NBIA), developed by the National Cancer Institute in 2005 [56].

The main goal of the VICTRE trial was not to replace traditional clinical trials but in turn assess whether in-silico trials could complement the latter mentioned in evaluating new medical imaging systems, in an attempt to reduce human trial size, length and expenses. Since this specific trial is completely simulated, it is based on a comparative human trial whose patient cohort consisted of 326 asymptomatic women, where 21 patients had extremely dense breasts (5 positives), 156 had heterogeneously dense breasts (46 positives), 130 had scattered densities (41 positives), and 19 present almost entirely fatty breasts (10 positives), selected from 7 clinical sites. Based on this comparative human trial, the VICTRE trial cohort is composed of virtual female patients whose breasts were

generated using a procedural analytic model in which major anatomical structures are randomly generated within a predefined breast volume bounded by skin and chest wall. In this trial, two forms of medical imaging were conducted, DM and DBT. The synthetic images were obtained using an in-silico version of the Siemens Mammomat Inspiration DM and DBT systems and using a customized version of the MC-GPU Monte Carlo transport code. The Mammomat Inspiration is characterized by the following specifications: 50º scanning angle, 25 projections, scanning time of 24 seconds, continuous tube motion and FBP reconstruction algorithm [25]. Moreover, the DBT system used a compressed thickness from 3,5 to 6 cm. These images were generated in raw format and then converted to DICOM format, incorporating metadata such as patient sex, patient comments, compression thickness and others [57].

The result of the trial was the in-silico imaging of 2986 virtual patients in both DM and DBT systems, with representative breast sizes and radiographic densities which computational readers analysed and inferred the following BI-RADS densities: 286 extremely dense, 1200 heterogeneously dense, 1200 scattered and 300 almost entirely fat.

Finally, the robustness of the interface was put to the test using DBT images in both CC and MLO views provided by Hospital da Luz in Lisboa, amounting to a total of 16 images classified. Each of these images were classified by a single specialist conducting visual inspection of the exam. These classifications could be as objective as one of the degrees of BD being written out in the comments or, in some cases, the classification was up to my evaluation when confronted with specific language and anatomical evidence pointed out by the specialist in the comments.

2.2. Methods

In the present work, a MATLAB interface was designed to automatically classify DBT images that are loaded into it by the specialist. Let us consider there is a certain DBT exam to classify. Firstly, it is loaded into the interface. Then, the user must proceed to the pre-processing phase where he/she will choose a region of interest (ROI) that will then be applied to all the images in the exam. After the computation of the ROIs, each of them is clustered into two different clusters that make up the ROI: a cluster corresponding to the fibroglandular tissue and another cluster which represents the adipose tissue,

according to SFCM2D Toolbox and the calculations of features occurs for each clustered ROI. Afterwards, the phase of feature analysis takes place and it basically finishes off the calculations of all the features retrieved in the previous phase and stores them. Finally, it is time for the classification process, which features the election particularity throughout (later mentioned in 2.2.5.7). The first classification stage places the data in one of two groups: group 1 (Fatty – ACR1 - and Scattered type – ACR2) or group 2 (Heterogeneous – ACR3 - and Dense type – ACR4). Then, depending on which group the exam was placed, a further number of features (thoroughly described in Chapter 3) is analysed and the final label classification is presented to the user. An example of the process is illustrated in Figure 2.1.

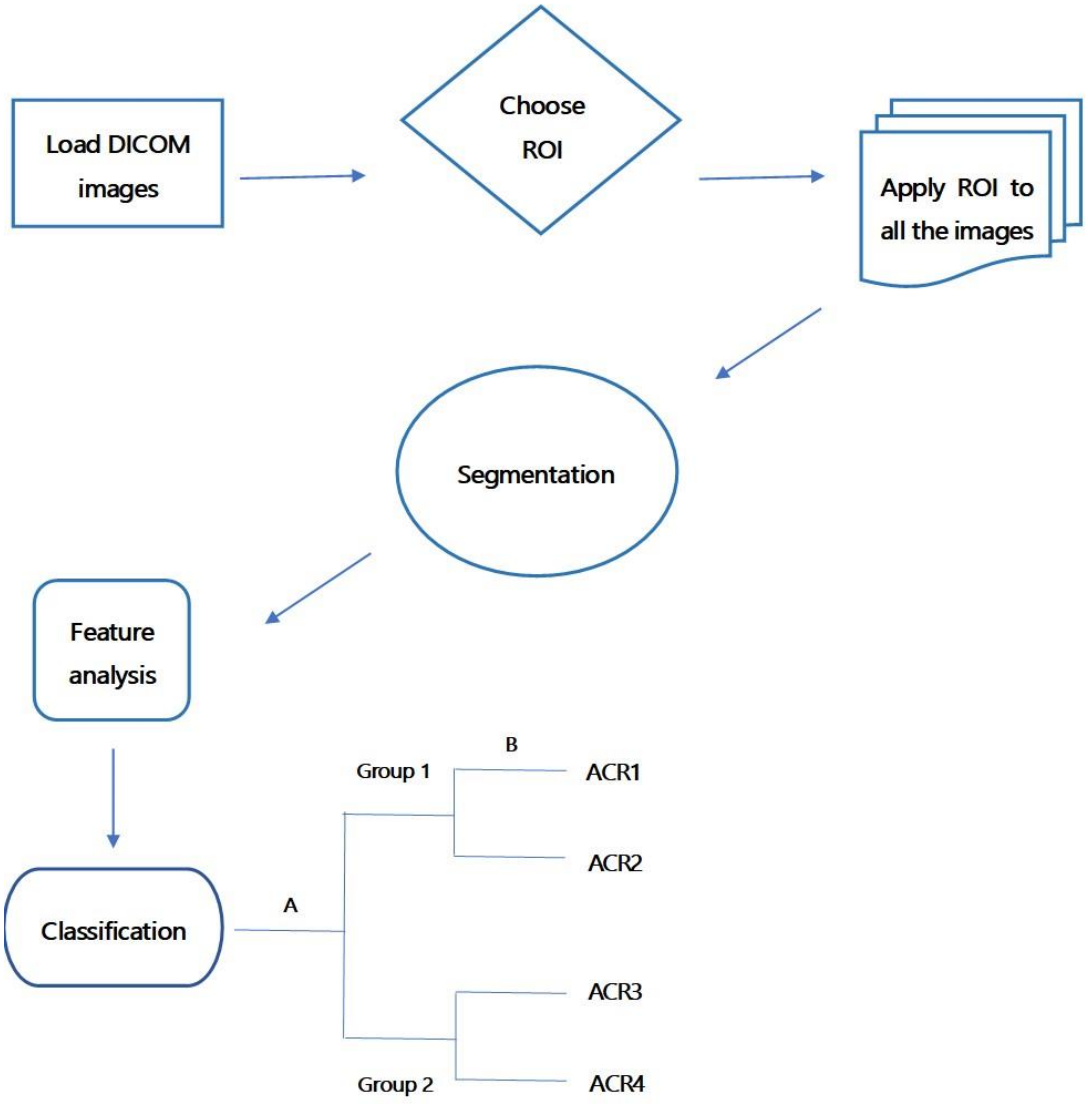

Figure 2.1: Flowchart of the interface.

2.2.1. Image Pre-Processing

The database used to develop the interface is composed of various DBT reconstructed slices corresponding to DICOM images that present not only the breast imaged but also the background. So as to only take into account breast tissue pixels, the user is asked to either draw a rectangular region of interest or use the automatic feature to do so. This way, a ROI is calculated and applied to all the DICOM images of the patient, allowing for a more comprehensive analysis of the breast tissue. After this processing, the result is several ROI images corresponding to the exact same number of DICOM initial images which will then be segmented and analysed in the following steps.

2.2.2. Image Segmentation

After pre-processing the slices, to properly classify BD, one must first execute the process of segmentation. Indeed, this step enables the separation and posterior quantization of the ratio between fibroglandular and adipose tissues. Many techniques have been used to segment the breast tissue into two different tissue types but one of the most recurrent is the k-Means clustering method. This unsupervised learning technique is a machine learning algorithm which attributes a certain cluster to each pixel given its grayscale intensity, so that it promotes intra-cluster variance minimization that translates to data points in the same cluster being similar and in turn differ from the ones in other clusters. Hence, each cluster corresponds to a set of data that was grouped according to their similarities and each represents a different grey level in the image. An improved technique that stemmed from k-Means is fuzzy C-Means which follows the same rational as the former but utilizes a fuzzy membership function that enables each pixel to belong to more than one cluster, having various membership degrees to each cluster [49]. The mean grey intensity of each cluster is then used to establish the cluster density hierarchy in which a higher value of mean corresponds to a higher probability of that cluster being composed of dense tissue. These various degrees of membership of each data points are represented by the main operation of the FCM algorithm which is

the minimization of the cost/objective function, commonly known as *J*, across each iteration, which is calculated using Equation 1:

$$J = \sum_{i=1}^{N}\sum_{j=1}^{C} \delta_{ij} \,||\, \chi_i - c_j \,||^2 \qquad [1]$$

where *N* is the number of data points (pixels), *C* is the number of clusters, χ_i is a pixel value, c_j is the centroid vector of cluster *j*, δ_{ij} is the degree of membership of the ith pixels χ_i in cluster *j* and the term enclosed in the module represents the closeness of the pixels χ_i to the centroid vector c_j of cluster *j*.

On the contrary to what happens in k-Means, in the FCM method the centroid vector of each cluster is the mean of all data points weighted by their corresponding membership degrees to each cluster and it is computed with Equation 2:

$$c_j = \frac{\sum_{i=1}^{N} {\delta_{ij}}^m \cdot \chi_i}{\sum_{k=1}^{C} \left[\frac{||\, \chi_i - c_j \,||}{||\, \chi_i - c_k \,||}\right]^{\frac{2}{m-1}}} \qquad [2]$$

where *m* is the fuzziness coefficient that controls the degree of fuzzy overlap with *m* > 1. Hence, this component represents the number of pixels that have a degree of membership in more than one cluster.

Finally, δ_{ij} is the degree of membership computed in the previous iteration which is randomly initialized for each pixel and is calculated according to Equation 3:

$$\delta_{ij} = \frac{1}{\sum_{k=1}^{C} \left[\frac{||\, \chi_i - c_j \,||}{||\, \chi_i - c_k \,||}\right]^{\frac{2}{m-1}}} \qquad [3]$$

This technique was implemented by using the SFCM2D file in the work of ABing [58] and the result of this process is two images corresponding to the two programmed different clusters – fibroglandular tissue and adipose tissue – represented in Figure 2.2.

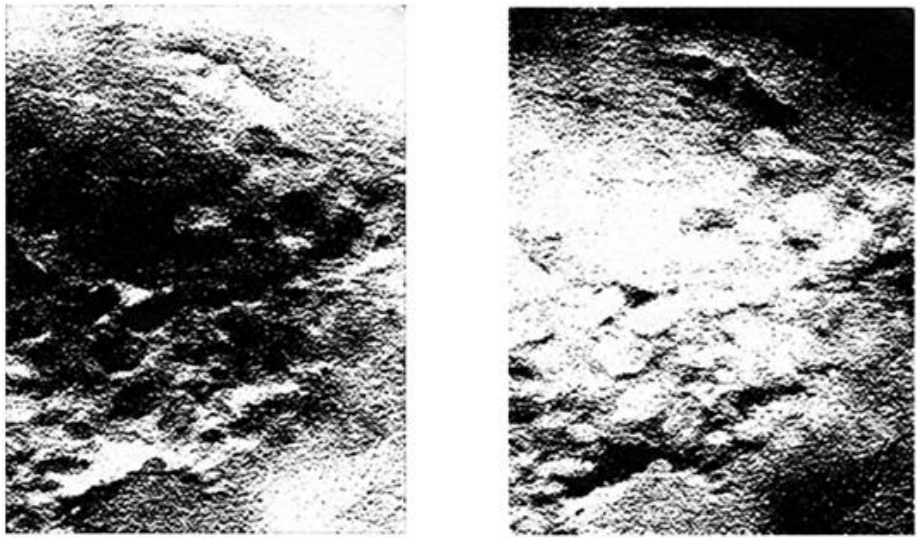

Figure 2.2: Example of a clustered ROI using the FCM method.

The choice of two clusters is due to the fact that the interface is mainly destined to CC images, whereas if a whole MLO breast image was to be classified then three clusters would be necessary in order to include the pectoral muscle. However, as it will be described later in the overall functioning of the interface, the user has a choice of drawing a rectangular region of interest (ROI) which, if drawn in the main part of the breast tissue, would enable as well the use of MLO images without the need to include the third cluster.

For texture analysis and feature calculation reasons, it was considered that the white pixels corresponded to fibroglandular tissue cluster datapoints and black pixels belonged to the adipose tissue cluster.

2.2.3. Feature Analysis

Texture analysis is a technique that started to be used in the 1970s as a tool to describe the spatial distribution of intensities that characterize different images to classify them. Moreover, it proved to be particularly useful to discriminate similar regions in different images. Therefore, given the different types of texture present in the segmented images, this approach was used in the present work to assess the different textural features that characterized each clustered ROI image.

Thus, like it was mentioned in the first Chapter, to distinguish the two types of clusters, it is necessary to compute features that discriminate them as clear as possible.

A feature analysis was computed by means of calculating histogram features and textural features, besides calculating the percentages of fibroglandular and adipose tissue present in the segmented breast ROI. And so, histogram features and two types of textural features (Haralick features and run-length features) were calculated for each cluster and averaged throughout all of the ROI's of the same patient, amounting to a total of 88 features, summed up in Table 2.1.

2.2.3.1. Histogram Features

The first group of features – histogram features - is also referred to as first-order statistics and it provides different statistical properties extracted from the image's intensity histogram. While the second group – texture features – depends on the co-occurrence, i.e., interaction between a pixel value and its neighbour's, histogram features depend only on individual pixels and their values. Hence, these features measure the probability of a grey level occurring at a randomly chosen location and so the distribution of grey levels in the image.

Six histogram derived features were extracted from each clustered ROI image [59], referring to different statistical and geometrical properties of the image's histogram, and those are:

- **Mean**: measures average brightness;
- **Variance**: measures average contrast brightness i.e. how much the grey levels differ from the mean;
- **Skewness**: measures the asymmetry of the image's histogram around the mean value (how the data points fall on both sides of the mean);
- **Kurtosis**: measures the combined weights on both sides of the tails in relation to the mean;
- **Energy**: provides a measure of information;
- **Entropy**: quantifies the degree of randomness of the histogram.

The six histogram features and their mathematical expressions are featured in Table 7.1.

2.2.3.2. Haralick Texture Features

Haralick features calculation was proposed by Haralick et al [60] in 1973 as a method for image classification and it consisted in a number of calculations derived from a grey-level co-occurrence matrix (GLCM) that describes the correlation between pixels. Firstly, four different co-occurrence matrixes are derived from the whole ROI image using the *graycomatrix* MATLAB function. This command calculates how regularly a pixel with greyscale intensity value *i* occurs horizontally adjacent to a pixel with value *j*. Hence, each element (*i,j*) that constitutes the matrix represents the number of times that pixel *i* and pixel *j* occurred horizontally adjacent. In computing the GLCM, it is necessary to choose in which direction should this "proximity" evaluation occur, for a constant distance between the pixel of interest and its neighbour of 1, which was shown to produce more accurate results as one can argue that the probability of a pixel being correlated to a pixel located near is higher than to one located far away in the image. Thus, calculations for four different possible directions were made, expressed in angles: 0º, 45º, 90º and 135º, expressed in Figure 2.3. For instance, for an angle of 135º, the offset direction is [-1 -1], where the first value corresponds to the number of rows between the pixel of interest and its neighbour and the second value corresponds to the number of columns between them. The same rational applies to the other three directions.

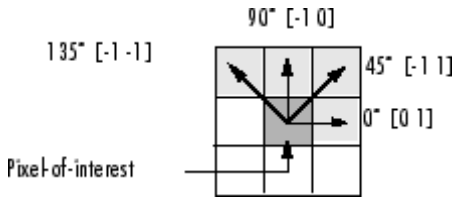

Figure 2.3: Four different directions used to calculate the GLCM [61].

After calculating the four GLCM, it is necessary to normalize them to only have values ranging from 0 to 1 so that they can be understood as probabilities. Following, nineteen Haralick features that represent various characteristics about the image are computed for each normalized GLCM [62]. Out of the total nineteen features, there are six main features from which the rest are derived which are:

- **Angular Second Moment**, which is a measure of textural uniformity;

- **Contrast**, also known as standard deviation, which accounts for the number of local pixel intensity variations that occur in the image;
- **Correlation**, which represents the linear dependency of grey tones in the image;
- **Variance**, which measures the heterogeneity in values present in an image, so that it increases when the grey level values differ from the mean;
- **Inverse Difference Moment** measures image homogeneity by assigning larger values for smaller grey tone differences in pair pixel elements;
- **Entropy**, directly opposing the Angular Second Moment feature, affers the complexity or disorder of an image.

All the nineteen features and their mathematical formulas are explicit in Table 7.2.

2.2.3.3. Run-length Texture Features

To understand the concept of the grey level run-length matrix (GLRM), one must first grasp what a grey-level "run" is. A GLRM explores the spatial connectivity between colinear pixels that, because of being so close in grey level, originate a grey level run [63]. Thus, a "run" corresponds to the length of consecutive pixels characterized by a specific grey level. Considering the definition of a grey level run and by visual observation of an image presenting textural diversity, one may say that there are two types of grey level "runs" correlated to the coarseness of the image. In fact, in a coarse textured image, there is the main occurrence of long grey-level runs and, on the contrary, in fine textures, the majority of grey-level runs are characterized as short [64].

The concept of the GLRM was first proposed by Galloway et al [65] in 1974 with the aim to be used for texture feature extraction. Moreover, a GLRM is typically represented by the notation $p(i, j \mid \theta)$ that constitutes the number of runs or times j that pixels of grey level i appear consecutively in a specified direction θ. Four GLRM were computed in four different directions - 0º, 45º, 90º and 135º - that were then combined in a single resulting GLRM from which six features were computed [66]. These six features correspond to four of the original proposed in the work of Galloway (SRE, LRE, GLN, RP) and two novel features (LGRE and HGRE) presented by Chu et al [67], corresponding to the following statistics:

- **Short Run Emphasis** (SRE): measures the distribution of short runs;
- **Long Run Emphasis** (LRE): measures the distribution of long runs;
- **Grey Level Non-Uniformity** (GLN): measures the similarity of grey level values throughout the image;
- **Run Percentage** (RP)**:** measures the homogeneity and the distribution of runs of an image;
- **Long Grey Level Run Emphasis** (LGRE): measures the distribution of low grey level values;
- **High Grey Level Run Emphasis** (HGRE): measures the distribution of high grey level values.

All run-length features extracted and corresponding mathematical expressions are featured in Table 7.3.

Table 2.1: Features calculated.

Feature Class	Number of features	Features
Histogram Features	6	Mean Variance Skewness Kurtosis Energy Entropy
Haralick Features	76	Angular Second Moment Contrast Correlation Variance Inverse Difference Moment Sum Average Sum Variance Sum Entropy Entropy Difference Variance Difference Entropy Information Measure of Correlation I Information Measure of Correlation II Maximal Correlation Coefficient Autocorrelation Cluster Prominence Cluster Shade Dissimilarity Maximum Probability
Run-Length Features	6	Short Run Emphasis Long Run Emphasis Grey Level Non-Uniformity Run Percentage Low Grey Level Run Emphasis High Grey Level Run Emphasis

2.2.4. Feature Selection Methods

After calculating the total number of features and adding the two corresponding to the percentages of adipose and fibroglandular tissues, it is necessary to perform feature selection so as to infer which features out of the total ninety are the most relevant for the classification task in question. This way, computational costs and time are smaller

as well as the risk of overfitting the classification model. According to Occam's razor, models should be as simple as one can explain them clearly and so, when we have a large number of features, explainability is lost.

So, the feature selection process was divided into three phases according to the classification process. Firstly, a distinction between the two main groups of labels and so the most important features to do it would be computed, represented by the two classes ACR2 and ACR3, as these two BD degrees are the most difficult to distinguish. This way, the image to be classified would be first inserted in either Group 1 (where it could be classified as ACR1 or ACR2) or Group 2 (where it could be classified as ACR3 or ACR4). Then, feature selection would again be performed to ascertain which features would provide the best distinction within each group.

Several methods of feature selection were attempted using MATLAB, such as forward and backward stepwise regression. However, these methods proved to be extremely slow and time-consuming and thus not effective. Hence, in the present work it was used a more efficient and straight forward feature selection technique which featured the use of Orange.

Orange is an open-source toolkit commonly used for data analysis and visualization. This software allows for an Excel file containing all the features and corresponding labels to be loaded on to it and then several widgets can be used to analyse it. In this study, the version of Orange used was 3.23.1 and it was provided by the Anaconda platform, more specifically Anaconda3 interface [68]. The widgets that were used to perform feature selection were Rank, Test and Score and Confusion Matrix, and they are featured in Figure 2.4. The first application ranks each feature according to assorted criteria such as Information Gain, Gain Ratio, ANOVA and Chi-Square, among others; Test and Score computes the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) accuracy of several machine learning classifiers given the set of features provided and, finally, the latter presents the confusion matrix of each of the machine learning classifiers.

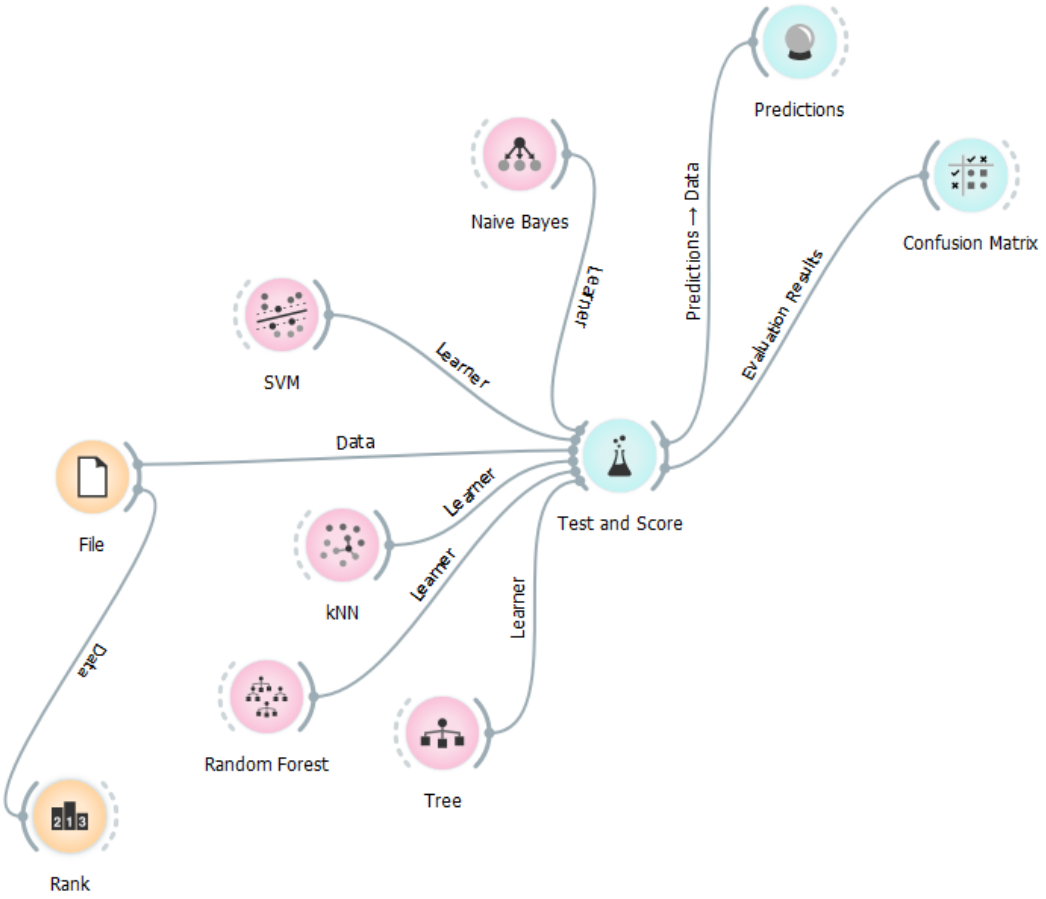

Figure 2.4: Scheme used in Orange workspace.

The AUC metric measures the entire 2D area underneath the ROC curve which plots two parameters: the true positive rate (*TPR*, in the *yy* axis) and the false positive rate (*FPR*, in the *xx* axis), represented in Equations 4 and 5 respectively.

$$TPR = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad [4]$$

$$FPR = \frac{False\ Positives}{False\ Positives + True\ Negatives} \quad [5]$$

Therefore, AUC is different from the basic model accuracy computation as it takes into account the entire range of class distributions and error costs of the classifier model

in order to examine its performance, thus constituting a better measure for evaluating and comparing learning algorithms [69].

Moreover, an additional criterion was used to attest the relevance of each feature: the Cohen's kappa coefficient, k. This coefficient measures intra-rater reliability for categorical classes, considering the possibility of the classification occurring by chance. Hence, k was computed to each of the classifiers according to their confusion matrix, using Equation 6:

$$k = \frac{p_o - p_e}{1 - p_e} \quad [6]$$

Where p_o is the relative observed agreement representing the number of accurate classifications out of all the classifications performed and p_e is the hypothetical probability of chance which represents the number of right and wrong classifications performed for each of the two classes. Let us take for example the following confusion matrix for a certain classifier, represented in Table 2.2:

Table 2.2: Example of a confusion matrix.

Predicted / Real	ACR1	ACR2
ACR1	2	2
ACR2	1	3

$$p_o = \frac{correct\ cases}{all\ cases} = \frac{5}{8} \quad [7]$$

$$p_e = P(Class1) + P(Class2) \quad [8]$$

Considering ACR1 as Class1 and ACR2 as Class2:

$$P(Class1) = \frac{all\ cases\ of\ real\ class1}{all\ cases} \cdot \frac{all\ cases\ of\ predicted\ class1}{all\ cases} \quad [9]$$

Using Equation 9 to calculate P(Class1) and P(Class2) we have:

$$P(Class1) = \frac{2+2}{8} \cdot \frac{2+1}{8} = 0{,}1875 \quad [10]$$

Calculating the same way for Class2, P(Class2) = 0,3125. Finally, $p_e = 0{,}5$ and k = 0,25.

According to Landis and Koch's Cohen's kappa Coefficient analysis [70], the corresponding strengths of reliability are as follows:

- $k < 0$: no agreement between predicted and real classes
- $0 \leq k < 0{,}2$: slight agreement
- $0{,}21 \leq k < 0{,}4$: fair agreement
- $0{,}41 \leq k < 0{,}6$: moderate agreement
- $0{,}61 \leq k < 0{,}8$: substantial agreement
- $0{,}81 \leq k \leq 1$: almost perfect agreement

So, a value of k close to 0 represents that most classifications made by the classifier have been performed randomly by the specific classifier.

To ascertain the relevance of each feature, the two scoring factors present in Rank that were taken into account were Information Gain and Gain Ratio. The first one corresponds to the expected amount of information that is gained with that specific feature and its impact in the reduction of entropy. The second element is the ratio of the information gain and the attribute's intrinsic information, which reduces the bias towards multivalued features that occurs in information gain.

The procedure of feature selection included three steps that were repeated for each classification phase: verifying the values of k and Area under Curve accuracy (AUC) of each classifier; removing all the features that had an Information Gain smaller than 1 and a Gain Ratio smaller than 0,5 and, finally, examining how the two factors, k and AUC, changed according to the features that were eliminated.

2.2.5. Classification

After collecting the specific features for each phase of classification, six supervised machine learning algorithms were implemented in order to constitute a multi classifier structure (MCS) used to decide the class for the two phases of classification. This system is widely used to improve the classification phase in various streams of work for example, in the work of Beevi et al [71], a MCS was employed to accurately detect mitotic cells. In the said work, the MCS is described as a way of heightening the performance of the various individual classifiers that make up the structure. Moreover, in a study implementing multi stage classification for breast cancer diagnosis, it was inferred that the multi classifier system accuracy was in fact larger that the individual classifiers' accuracies [72]. So, the six machine learning classifiers used in the MCS are: k-Nearest Neighbour (kNN), Discriminant Analysis (DA), Naïve Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM) and Random Forest (RF), further described below.

2.2.5.1. k-Nearest Neighbour

The k-Nearest Neighbour algorithm can be summarized in three simple steps: first, it finds the distances between the datapoint to be classified and all the examples in the data; then, it selects the previously set number of examples closest to the datapoint (in this case, $k = 3$) and, finally, it votes for the final classification to be equal to the most frequent label those k examples belong to [72].This technique is illustrated in Figure 2.5, where the green circle represents the unknown sample to be classified. Depending on the value chosen for k, i.e., the diameter of the circumference in black and thus the neighbours that are encompassed by it, the unknown sample can be classified as belonging to the red class or the blue class. In this specific case, the black solid line represents $k = 3$ and so the three nearest neighbours will decide the label attributed to the green circle. In the same way, for $k = 5$, which is represented by the black dotted line, the five nearest neighbours to the unknown sample will decide which class should be assigned to the green circle.

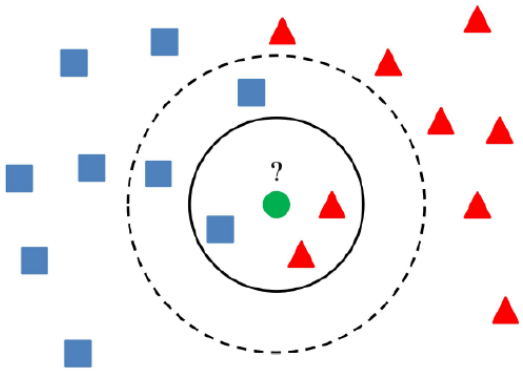

Figure 2.5: kNN illustration [73].

2.2.5.2. Discriminant Analysis

DA classifier was first proposed by Fisher in 1936 to solve various classification problems, as it assesses which features are more relevant for the classification task in hand and then uses them to classify the unknown sample. To do so, the classifier is represented by discriminant functions ($\{f_1, f_2, \ldots, f_c\}$) also known as decision functions, where c corresponds to the existing number of classes. These functions aid to circumscribe the classes' region and compute decision boundaries to discriminate between different classes into different regions. The algorithm of the DA model can be summed up as such [74]:

1. Calculate the mean, mean-centring data and covariance matrices of each class;
2. Compute the discriminant function for each class;
3. The discriminant functions generate decision boundaries;
4. For all discriminant functions computed, substitute the value of the unknown sample in the discriminant function;
5. Classify the unknown sample by assigning it the class label corresponding to the class that has the maximum discriminant function.

Thus, the class which produces the maximum value of discriminant function is the one predicted as the label of the unknown sample.

2.2.5.3. Naïve Bayes

The Naïve Bayes model is a probabilistic classifier whose functioning is based on the Bayes theorem [75]. This theorem is used to calculate the probability of an event A happening granted that another event B has happened with Equation 11:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad [11]$$

An important particularity about this postulate is the assumption of independency between the predictors (features), i.e., A and B. Now applying this theorem to the context of a machine learning classifier, in the NB model A represents a certain class or label to be predicted and B stands for the features that enable the prediction to occur. Using Y to represent the label and X the feature vector, we have Equation 12:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad [12]$$

where $X = (x_1, x_2, \ldots, x_n)$ and n represents the number of predictors used to predict Y. By substituting X in Equation 12 and expanding it using the chain rule, we reach Equation 13:

$$P(Y|x_1, x_2, \ldots, x_n) = \frac{P(x_1|Y)P(x_2|Y) \ldots P(x_n|Y)P(Y)}{P(x_1)\,P(x_2) \ldots P(x_n)} \quad [13]$$

Equation 13 represents the probability of the datapoint belonging to label Y given each feature of X. In this work's classification task, where each phase comprehends only two classes each time, Y is composed of y_1 and y_2, representing the two possibilities of labels.

Depending on the classification task in hand, one of three types of NB classifiers can be chosen: Multinomial NB, Bernoulli NB and Gaussian NB. Taking into account the continuous nature of the predictors' values, the NB classifier chosen was Gaussian and

the conditional probability formula becomes the following Equation 14, as described in [72]:

$$P(x_i|Y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}) \quad [14]$$

where μ and σ are, respectively, the mean and variance of the feature vector X and x_i represents the elements in X.

Finally, the label that has the highest value of probability is the one assigned to the datapoint being classified.

2.2.5.4. Decision Tree

This type of algorithm is characterized by arriving to a certain classification by continuously splitting the data depending on the feature being analysed. The decision tree (DT) is made up by three elements, featured in Figure 2.6:

- Nodes: test for the value of a certain attribute;
- Branches: outcome of a node (test) that connects to the next node or leaf;
- Leaf nodes: terminal nodes that represent the predicted label.

A decision tree can be one of two types: a classification tree, when the labels are discrete, or a regression tree, when the labels are continuous. Seeing as in both classification phases the labels are categorical i.e. discrete, the decision tree implemented is of the classification type [76]. This particular type of tree is built using the iterative process of binary recursive partitioning which basically consists of clustering the data into partitions (leaves) repeatedly. The algorithm starts at the tree root and splits the data on the feature that results in the largest information gain so as to reduce the uncertainty throughout the process of arriving to the final decision. Then, this splitting procedure is repeated on each of the branches until all elements of each leaf have the same class label. The number of times the data is split is previously defined and set to 50 in this case, so that it prevents overfitting.

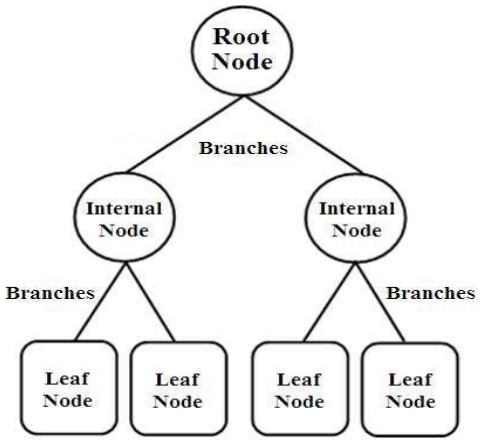

Figure 2.6: Decision Tree components [77].

2.2.5.5. Support Vector Machine

SVM is a linear machine learning model that can be used for classification or regression tasks. The main principle of SVM is to determine a hyperplane (line) that best separates the datapoints into two different classes, as it is more suited for binary problems [78], which is illustrated in Figure 2.7. In order to find the hyperplane that most clearly allows for a distinction between classes, there are three essential steps, as follows:

1. Find the points closest to the line from both classes, which are named support vectors;
2. Compute the distance between the line and the support vectors, which is called the margin;
3. Maximize the margin;
4. The hyperplane chosen is the one which margin is maximum.

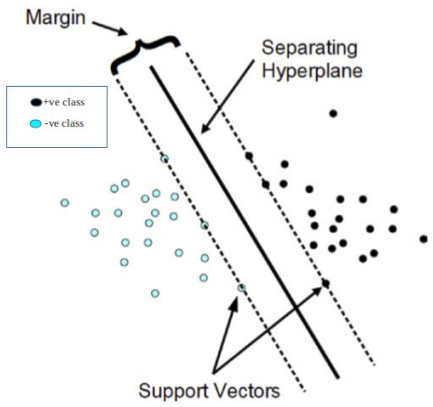

Figure 2.7: SVM algorithm geometric elements [78].

This all works well for clearly linear cases, i.e., when the set of training data is linearly separable. However, most often than not, the dataset doesn't behave as described and there is the need to map the non-linear separable dataset into a higher dimensional space that allows for the definition of a hyperplane that can accurately separate the samples. This process is done by using a kernel function most suited for the classification problem. Hence, a kernel function allows the projection of the dataset onto a higher dimension where it is possible to fit a plane to separate it. There are three types of kernels: linear kernel, polynomial kernel and Gaussian kernel. The latter was the one used for the SVM implemented and it is calculated with Equation 15:

$$K(X_1, X_2) = e^{-\gamma||X_1 - X_2||^2} \quad [15]$$

where $||X_1 - X_2||$ is the Euclidean distance between the two points and γ is a variable that indicates the degree of the model overfitting (it increases as the model overfits). Thus, the Gaussian kernel constitutes a reasonable measure of X_1 and X_2's similarity, granted that its value is close to one when X_1 and X_2 are close, and near zero when the points are far apart.

Given a training set TS = {($\overrightarrow{x_1}$, L_1), ($\overrightarrow{x_2}$, L_2), ⋯, ($\overrightarrow{x_M}$, L_M)}, where $\overrightarrow{x_i}$ (i = 1, 2,⋯,M) is the training data and L_i ($L_i \in \{-1, 1\}$) is the class label, the test vector $\overrightarrow{x_{test}}$ is classified according to the sign of function c, defined in Equation 16:

$$c\left(\overrightarrow{x_{test}}\right) = \sum\{\alpha_i \cdot L_i \cdot (\overrightarrow{x_i^T} \cdot \overrightarrow{x_{test}}) + b\} \quad [16]$$

where α_i (i = 1, 2, …, M) are the nonzero quadratic coefficients and ($|b| / ||\vec{w}||$) is the perpendicular distance between the hyperplane and the origin, whereas w is the normal vector of the hyperplane [72]. So, if c is positive, the unknown sample from $\overrightarrow{x_{test}}$ will belong on the right side of the hyperplane and if c is negative it will be classified as the left side class.

2.2.5.6. Random Forest

The RF classifier is a tree-based learning algorithm, composed of a set of decision tree classifiers that are selected randomly from the training set, as it can be seen in Figure 2.8. So, as it combines more than one algorithm of the same kind, it is considered an ensemble algorithm. This algorithm, akin to some of the other algorithms, can be used for regression or classification problems, such as the present. Thus, the predicted label corresponds to the majority vote of all predicted classes over the number of trees previously set (in this case fifty) [72].

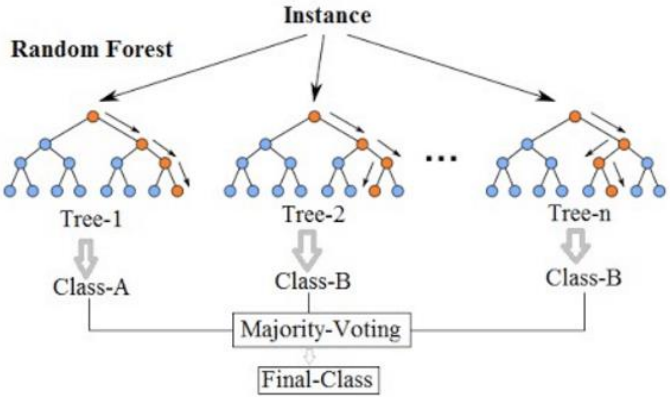

Figure 2.8: Random Forest classifier [79].

2.2.5.7. Election style classification decision

Taking into account the labels predicted by each classifier on each classification phase, an election system takes place. In fact, the label that has the most votes amongst

all the predicted labels is the one selected. However, in the event of the two possible labels having the same number of votes (both having three each), the label selected is the one belonging to the classifier that has the highest accuracy value, so as to prevent randomness in the prediction. This procedure is illustrated below where Process A and Process B are the ones defined in the flowchart present at the beginning of the Chapter as the two moments of classification that take place in the interface.

Process A:

Label kNN: **Scattered**
Label DA: **Scattered**
Label NB: **Scattered**
Label DT: **Scattered**
Label SVM: **Hetero**
Label RF: **Hetero**

Data will continue to be classified in **Group 1**

Process B:

Label kNN: **Scattered**
Label DA: **Scattered**
Label NB: **Scattered**
Label DT: **Fatty**
Label SVM: **Fatty**
Label RF: **Fatty**

The final classification corresponds to the label predicted by the classifier with the highest accuracy value

For illustrative reasons, let us assume that the classifier with the highest accuracy was Decision Tree. Then, the final label predicted by the interface would be ACR1 (most commonly used among healthcare professionals than "Fatty breast type").

3

Results and Discussion

In this Chapter, results regarding some of the methods described in the previous chapter are presented as well as comments with respect to certain decisions made throughout the present work. Moreover, the most important result of all – the application – is displayed in terms of its graphics and overall robustness and accuracy when tested with the materials used.

3.1. Segmentation

When applying the SFCM2D toolbox to compute FCM in MATLAB, there are 3 other parameters that had to be chosen beside the number of clusters, namely: the number of iterations (*max_iter*), the exponent of the fuzzy partition matrix (*expo*) and the value of threshold to binarize the images (*binary threshold value*). These parameters should be selected so that, besides minimizing computational time and achieving good results i.e. clear images, they attain a good compromise between two validity functions that were calculated during the segmentation process: the Partition Coefficient *Vpc* and the Partition Entropy *Vpe*, which are defined in Equations 17 and 18, respectively [80]:

$$Vpc = \frac{\sum_{j=1}^{N}\sum_{i=1}^{C}{\delta_{ij}}^{2}}{N} \quad [17]$$

$$Vpe = -\frac{\sum_{j=1}^{N}\sum_{i=1}^{C}\{\delta_{ij}\log\delta_{ij}\}}{N} \quad [18]$$

The value of *expo* controls the degree of fuzzy overlap between clusters and according to the literature it should be greater than 1, granted that smaller values create more crisp cluster boundaries. Hence, the value for *expo* chosen was 2. Moreover, the value of *max_iter* and *binary threshold value* were computed so that the following criteria are obtained in order to achieve the best interpretation of the samples considered:

- Minimize the value of the Objective Function, *J*
- Minimize the value of *Vpe*
- Maximize the value of *Vpc*

The results are described in the graphs below and they were obtained by following the rational of firstly fixing the value of *max_iter* for 100 and analysing which value of *binary threshold* is most suited. Then, for that value of *binary threshold,* select the number of *max_iter* most agreeable resulting from a compromise between *Vpc*, *Vpe* and *J*.

For a fixed value of *max_iter* = 100:

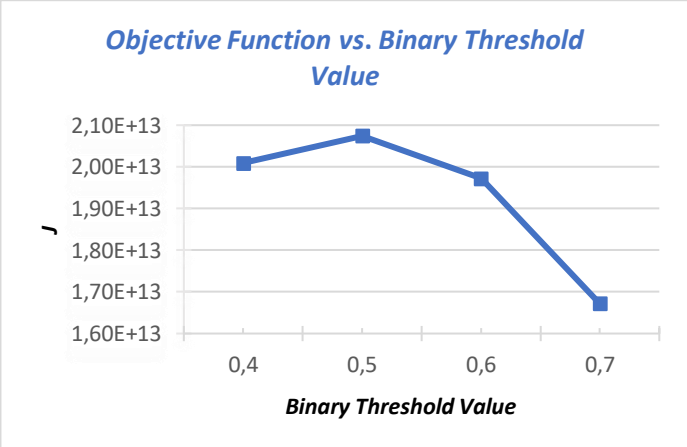

Figure 3.1: *J* as a function of the *Binary Threshold Value*.

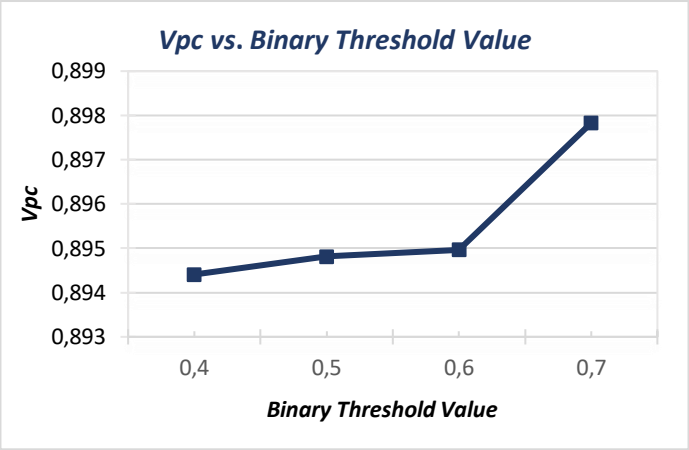

Figure 3.2: *Vpc* as a function of the *Binary Threshold Value*.

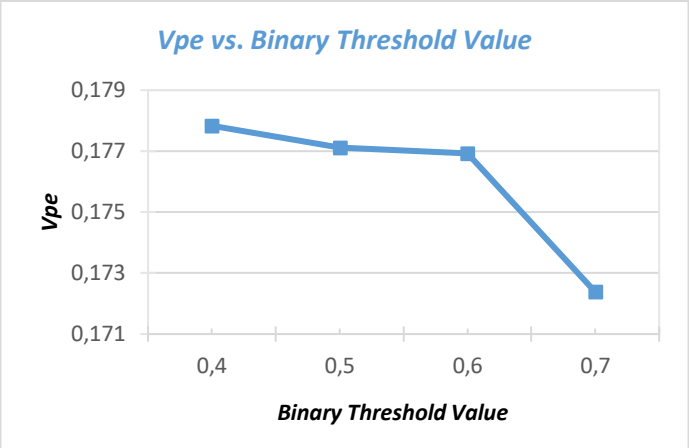

Figure 3.3: *Vpe* as a function of the *Binary Threshold Value*.

By visual analysis of the three figures 3.1, 3.2 and 3.3 that feature the graphs regarding the choice of *threshold* value, the value that satisfies the three criteria is 0,7.

Fixing this value and now evaluating the number of iterations, *max_iter*, we obtain the three following graphs:

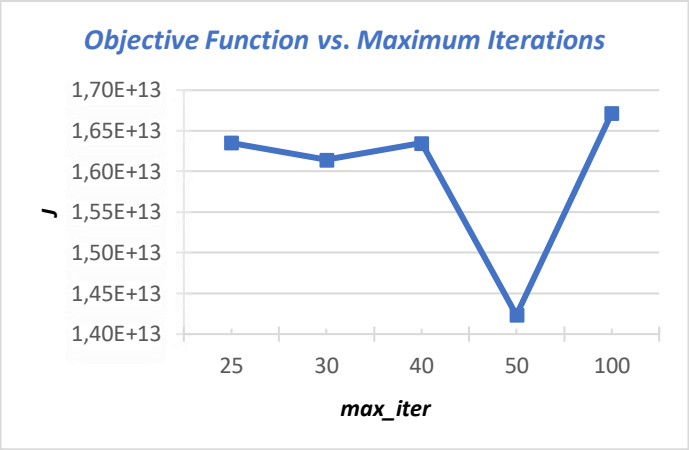

Figure 3.4: *J* as a function of *Maximum Iterations* value.

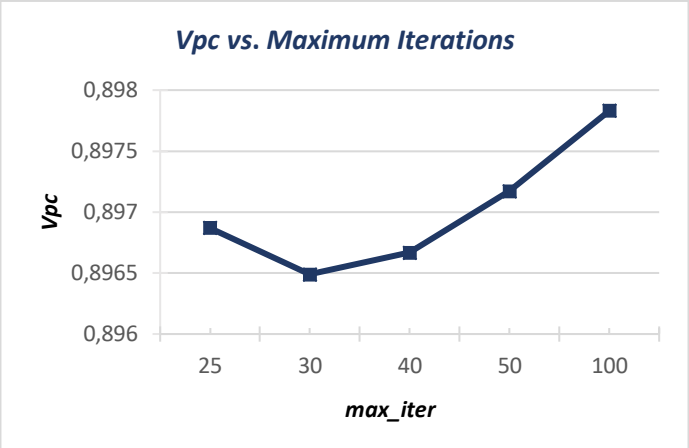

Figure 3.5: *Vpc* as a function of *Maximum Iterations* value.

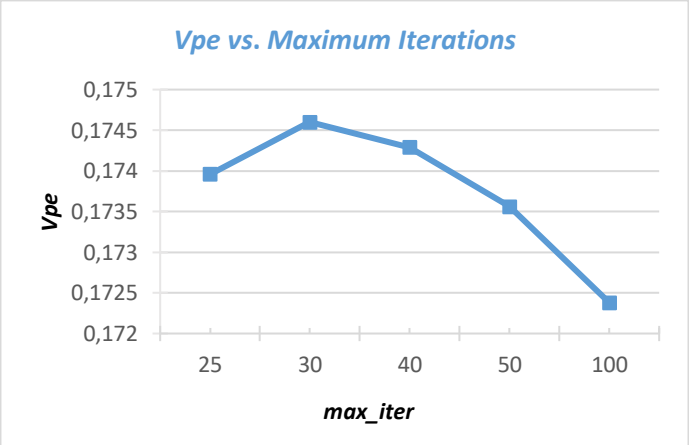

Figure 3.6: *Vpe* as a function of *Maximum Iterations* value.

Looking at Figure 3.4 that portrays the graph regarding the Objective Function, we observe that the two best options for the value of *max_iter* are 50 and 30, given the fact that they attain the smallest values for *J*. Although the best value for *max_iter* should appear to be 50 when compared to the second-best option of *max_iter* = 30 in terms of the *Vpc* and *Vpe* values (Figures 3.5 and 3.6), it posed a larger computational cost and time.

Therefore, the values chosen for each parameter were *max_iter* = 30, *expo* = 2 and *binary threshold value* = 0,7.

3.2. Feature Selection

Like it was mentioned in the previous chapter, feature selection was conducted using the Orange tool, more specifically the Rank widget. The results obtained for each of the classification phases are illustrated in Tables 3.1A, 3.1B, 3.2A, 3.2B, 3.3A and 3.3B. It is important to note that Discriminant Analysis classifier is not yet featured in the Orange version used, so that is why it is not present in the tables below.

The A versions of each table correspond to the initial values of AUC and k for each classifier using all the features computed. The B versions illustrate the change in value of AUC and k for each classifier after excluding the features that have values below the ones mentioned in Chapter 2 in terms of Information Gain and Gain Ratio. A clear improvement in both factors is observable when comparing tables A and B.

Phase 1: Group 1 vs. Group 2

Table 3.1A: k and AUC values with the initial number of features.

	90 features				
	kNN	**RF**	**NB**	**DT**	**SVM**
k	0,75	0,75	1	0,75	0,75
AUC	0,875	0,938	1	0,875	0,938

Table 3.1B: k and AUC values after excluding irrelevant features.

	34 features				
	kNN	**RF**	**NB**	**DT**	**SVM**
k	1	0,75	1	1	1
AUC	1	0,969	1	0,875	1

Phase 2A: ACR1 vs. ACR2

Table 3.2A: k and AUC values with the initial number of features.

	90 features				
	kNN	**RF**	**NB**	**DT**	**SVM**
k	-0,5	0,5	0,5	0,5	0,25
AUC	0,25	0,688	0,812	0,75	0,25

Table 3.2B: k and AUC values after excluding irrelevant features.

	23 features				
	kNN	**RF**	**NB**	**DT**	**SVM**
k	0,75	0,75	0,5	0,75	0,75
AUC	0,75	0,75	0,938	0,875	0,75

Phase 2B: ACR3 vs. ACR4

Table 3.3A: k and AUC values with the initial number of features.

	90 features				
	kNN	**RF**	**NB**	**DT**	**SVM**
k	-0,25	0,25	0	0,75	-0,5
AUC	0,375	0,656	0,312	0,875	0,625

Table 3.3B: k and AUC values after excluding irrelevant features.

	14 features				
	kNN	**RF**	**NB**	**DT**	**SVM**
k	0,25	0,5	0,75	0,75	1
AUC	0,625	0,875	0,938	0,875	0,688

In each table B, the values highlighted in green illustrate improvements achieved by reducing the dimensionality of features in relation to the previous situations that include all the initial features. To not overfit the models, only the two elements Information Gain and Gain Ratio from the Rank widget were evaluated and used to decide which features to exclude.

Therefore, different features were selected for the three different classification tasks and this enables a faster computation of the final label for each phase because of the reduced number of features that is analysed. All features collected and corresponding classification tasks are summed up in Table 3.4:

Table 3.4: Features collected and analysed for each classification task.

Classification Task	Features		
Group 1 vs. Group 2	% Adipose tissue % Fibroglandular tissue Variance_0 SumAverage_0 SumVariance_0 IMC2_0 Autocorrelation_0 ClusterProminence_0 ClusterShade_0 Variance_45	SumAverage_45 SumVariance_45 Autocorrelation_45 ClusterProminence_45 ClusterShade_45 Variance_90 SumAverage_90 SumVariance_90 Autocorrelation_90 IMC2_90 ClusterProminence_90 ClusterShade_90	Variance_135 SumAverage_135 SumVariance_135 Autocorrelation_135 ClusterProminence_135 ClusterShade_135 RP HGRE GLN Mean Skewness Kurtosis
ACR1 vs. ACR2	Correlation_0 InverseDifference-Moment_0 IMC1_0 MCC_0 Dissimilarity_0 Correlation_45 InverseDifference-Moment_45	IMC1_45 IMC2_45 MCC_45 Dissimilarity_45 Correlation_90 InverseDifference-Moment_90 IMC1_90 MCC_90	Correlation_135 InverseDifference-Moment_135 IMC1_135 IMC2_135 MCC_135 GLN LGRE Variance
ACR3 vs. ACR4	Variance_0 SumVariance_0 Autocorrelation_0 Variance_45	SumVariance_45 Autocorrelation_45 Variance_90 SumVariance_90 Autocorrelation_90	SumVariance_135 Autocorrelation_135 RP HGRE GLN

3.3. Accuracy and Robustness

In order to evaluate the quality and effectiveness of the algorithm developed, a dataset of images unknown to the MCS was tested and the AUC of each classification phase was computed. Firstly, the algorithm was tested using images from the database that was used to train the MCS – VICTRE – and then, so as to also evaluate the algorithm's robustness, images from Hospital da Luz were also classified.

3.3.1. VICTRE Dataset

For a total of 350 images classified from the VICTRE database, three different confusion matrixes and AUC values were obtained for the three classification phases.

Starting by evaluating the very first classification phase where the image is classified into one of two groups: either group 1 (ACR1 and ACR2) or group 2 (ACR3 and ACR4). As it was mentioned before in Chapter 2, this classification aims to basically classify the image as belonging to either group 1 and group 2 by using features that best distinguish between ACR2 and ACR3 as these degrees are the most similar and difficult to discriminate. The confusion matrix obtained is illustrated in Table 3.5.

Table 3.5: Confusion matrix of Group 1 vs. Group 2 classification phase.

Predicted / Real	Group 1	Group 2
Group 1	134	3
Group 2	4	209

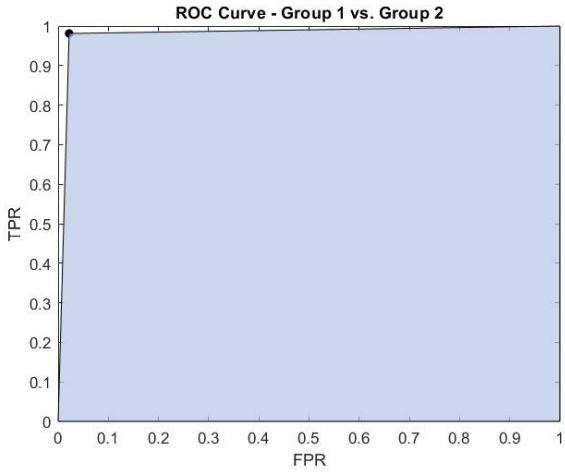

Figure 3.7: ROC curve obtained in Group 1 vs. Group 2 classification phase.

Using the results in Table 3.5, the corresponding values of TPR and FPR were calculated according to Equations 4 and 5 featured in Chapter 2, the corresponding ROC curve was plot using a maximum threshold of 1, illustrated in Figure 3.7. Finally, the AUC value was determined using the MATLAB function *trapz* and it came at a value of 0,9797.

Next, the same computations were performed for the following classification phase, depending on the previous result (either group 1 or group 2). Within group 1, the classifications of ACR1 vs. ACR2 obtained are summed in the Table 3.6:

Table 3.6: Confusion matrix of ACR1 vs. ACR2 classification.

Predicted / Real	ACR1	ACR2
ACR1	5	5
ACR2	1	126

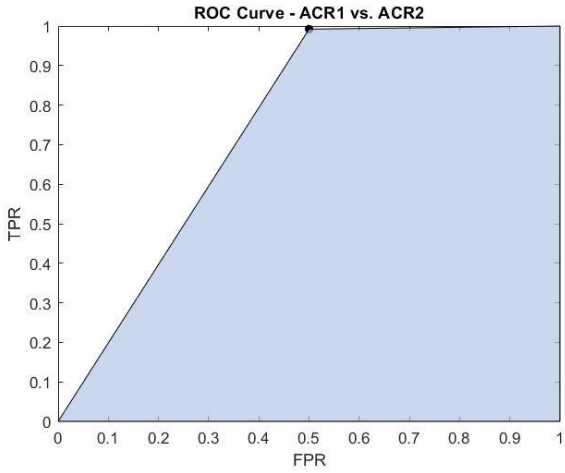

Figure 3.8: ROC curve obtained in ACR1 vs. ACR2 classification phase.

Thus, the resulting ROC curve is represented in Figure 3.8 and this classification phase amounted to an AUC of 0,7461.

Finally, if the result in the first phase of classification was group 2, the image would proceed to being classified as either ACR3 or ACR4. In this classification phase, the confusion matrix and ROC curve obtained are illustrated by Table 3.7 and Figure 3.9, respectively:

Table 3.7: Confusion matrix of ACR3 vs. ACR4 classification.

Predicted / Real	ACR3	ACR4
ACR3	141	0
ACR4	57	15

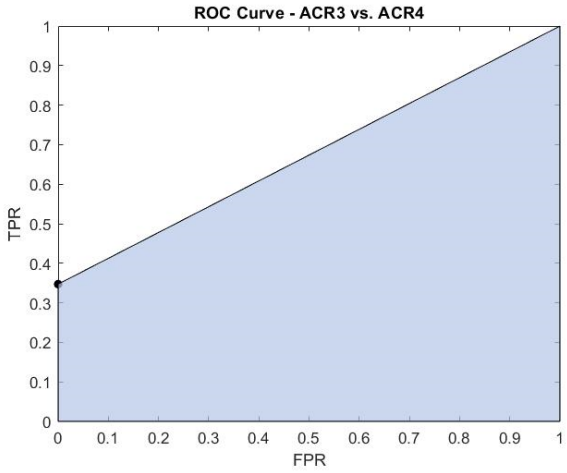

Figure 3.9: ROC curve obtained in ACR3 vs. ACR4 classification phase.

Hence, the AUC obtained for this phase came to 0,6736.

By observing the AUC values of the three different classifications and corresponding ROC curves, it is clear that the first phase, Group 1 versus Group 2, was the one that produced the most accurate results when compared to the following two. This result is due to the difference in the dimensionality of training data files available from each BD category. Indeed, as it is referred in Chapter 2, the amount of ACR2 and ACR3 exams available in the VICTRE database is almost four times than that of ACR1 and ACR4. Hence, given the fact that ACR1 and ACR4 images exist in lower quantity, when comparing them to ACR2 and ACR3, respectively, the AUC of the classification is not as high and the MCS fails. Moreover, looking at the small difference in AUC value between the classifications within each group, one may argue that it comes down to the number of images evaluated from the categories ACR1 and ACR4. Effectively, more ACR4 exams were evaluated on the testing phase hence why its AUC value is lower than the one computed for classification within Group 1.

In conclusion, when comparing the number of exams successfully classified to the total, the algorithm's accuracy comes at 82% when testing the VICTRE database.

3.3.2. Hospital da Luz Dataset

When it comes to evaluating the algorithm using the dataset provided by Hospital da Luz, a subjective component was involved. Each of the exams had the doctor's

observations and comments that reported the morphology of the breast and featured explicitly the BD degree. However, in four out of the exams classified, the doctor did not convey which BD degree he attributed and so some assumptions were made concerning it given the doctor's notes on the breast physiology. Hence, these reports were particularly subjective as not only they relied on a visual observation made by the doctor but also the final result was not always explicit.

Thus, for the 16 classified images, the confusion matrixes obtained for each classification phase are represented in Tables 3.8, 3.9 and 3.10:

Table 3.8: Confusion matrix of Group 1 vs. Group 2 classification.

Predicted / Real	Group 1	Group 2
Group 1	4	2
Group 2	3	7

Table 3.9: Confusion matrix of ACR1 vs. ACR2 classification.

Predicted / Real	ACR1	ACR2
ACR1	0	0
ACR2	0	4

Table 3.10: Confusion matrix of ACR3 vs. ACR4 classification.

Predicted / Real	ACR3	ACR4
ACR3	6	1
ACR4	0	0

So, out of the 16 exams, the algorithm "agreed" with 10 of the doctor's classifications. As it was mentioned earlier, given the fact that some exams didn't have a direct BD classification and so it had to be deduced from the doctor's notes, one may argue that 2 of the 6 exams that the algorithm failed to classify could be different BD degrees, depending on the interpretation of the doctor's comments. Hence, for the Hospital da Luz dataset, the algorithm's accuracy comes at 62,5% for the images analysed. More importantly than testing the algorithm, these images were paramount to evaluate its robustness, as the images with which it was trained differ not only in terms of file format but also regarding image composition. Thus, the algorithm proved to be robust enough that it can properly classify DBT DICOM images that are different appearance-wise from the ones it was trained with.

3.4. ABC Application

As it was mentioned before, the main result of the present work is the application. In fact, an interface destined to be used by specialists with the aim to classify objectively BD with DBT images was successfully constructed and compiled into an executable application, available for the user to download and posteriorly run without the need of Internet.

When the user runs the application, the logo pops up and is followed by the display of the application's environment, presented below in Figure 3.10.

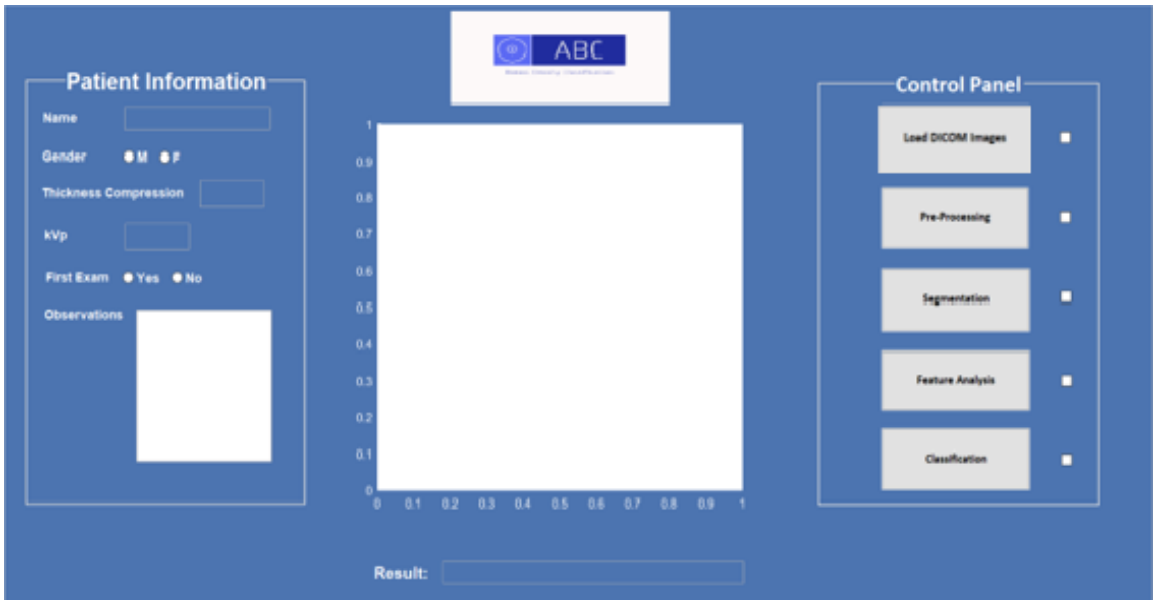

Figure 3.10. Application environment.

A brief instruction manual that displays the application's graphics and presents what all the interface's features do and how to use them can be found in Appendix A.

A quick overview of the application's features on each side can be summarized as such:

- On the left hand-side is where the Patient Information component is located and it entails several personal data about the patient that are relevant to a better assessment of breast cancer risk when combined with the final result regarding the patient's BD displayed by the application;
- On the right hand-side the user can find the Control Panel which contains all the application's features, displayed in order of use so as to compute the patient's BD: firstly, the user is asked to load the folder that contains the patient's DBT images; then, the user has to proceed to performing pre-processing in order to select a ROI to be evaluated; the next two steps, Segmentation and Feature Analysis, are mathematical calculation steps that, when completed, lead to the final command, "Classification", that presents the BD classification.

After conducting all the commands displayed in the application (accordingly to the instruction manual), the user is presented with the final breast density classification next to "Result", as it can be seen in Figure 3.11.

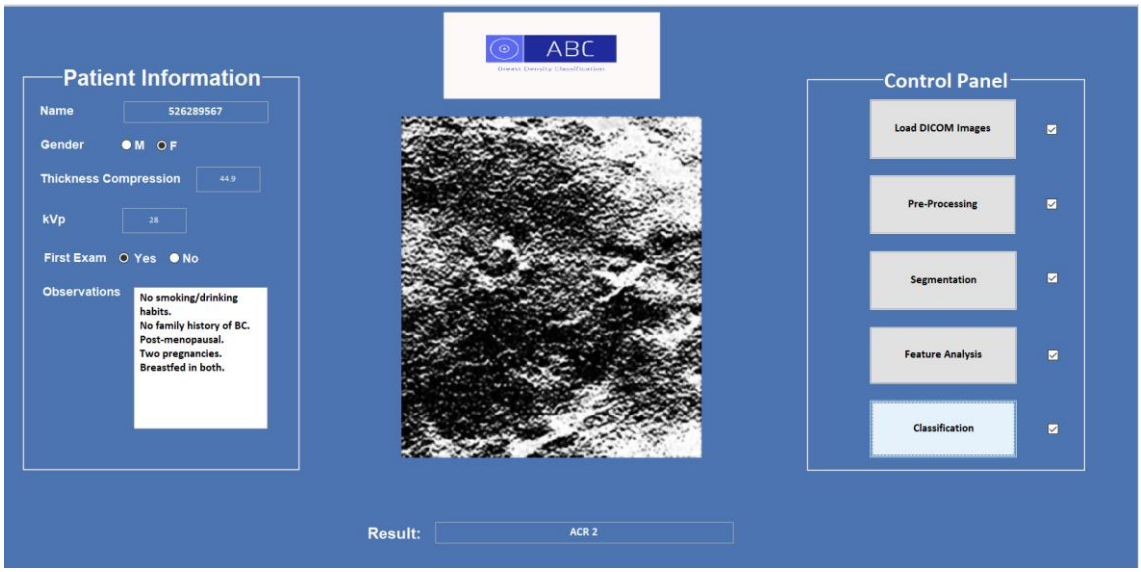

Figure 3.11: Application environment with the final result displayed.

4

Concluding Remarks

In this Chapter, a brief overview of this study is presented regarding the objectives reached and aspects to be improved in the future.

In this study, an application destined for doctors to automatically classify breast density and so enable them to inform the patient about her density status and consequently breast cancer risk was produced. Hence, the main goal of creating an algorithm that would take advantage of the 3D nature of digital breast tomosynthesis images to automatically classify BD, as defined in Chapter 1, was successfully achieved with good results in terms of the algorithm's accuracies and user-friendly design. One of the biggest advantages associated with the way the application was created is the fact that it can run without the use of Internet, which is more economically friendly and valuable during a Wi-Fi shortage.

However, there are certain shortcomings associated with the application that, by being overcome, would improve its quality like, for example, its AUC. If the multi classifier structure was to be trained with a larger dataset, enriched with more examples of ACR1 and ACR4 exams, it would more easily recognize these two "rarer" BD degrees and so the AUC for each of the classification phases where these degrees are featured would increase, leading to a more accurate algorithm. Another minor inconvenient that largely depends on several factors like the device's CPU characteristics and RAM value and

whether other programmes are being executed at the same time would be the segmentation's execution time. In fact, if the application was to be executed in a device that was running other programmes simultaneously, the segmentation command could take up to 5 minutes to be concluded.

Besides improving the factors mentioned earlier and thus making the application more attractive to be used and so possibly getting it to medical facilities, other steps need to be pursued to combat breast cancer. In fact, working towards standardizing the use of DBT exclusively (or in addition to digital mammography) and mandatory BD classification in order to achieve an earlier diagnose and so make the treatment of breast cancer have higher chances of success is imperative as this disease is still an ongoing issue and a life-limiting factor in this day and age.

5

References

[1] Óbitos (N.º) por Local de residência (NUTS - 2013), Sexo, Grupo etário e Causa de morte [Online].

[2] T. Wu, R. H. Moore, E. A. Rafferty, and D. B. Kopans, "A comparison of reconstruction algorithms for breast tomosynthesis," (in English), *Medical Physics,* Article vol. 31, no. 9, pp. 2636-2647, Sep 2004.

[3] J. A. Baker and J. Y. Lo, "Breast Tomosynthesis: State-of-the-Art and Review of the Literature," (in English), *Academic Radiology,* Review vol. 18, no. 10, pp. 1298-1310, Oct 2011.

[4] R. Fusco *et al.*, "Evaluation of average glandular dose and investigation of the relationship with compressed breast thickness in dual energy contrast enhanced digital mammography and digital breast tomosynthesis," (in English), *European Journal of Radiology,* Article vol. 126, p. 8, May 2020, Art. no. 108912.

[5] N. Sharma, "Virtual special issue on digital breast tomosynthesis," (in English), *Clinical Radiology,* Editorial Material vol. 74, no. 12, pp. 901-902, Dec 2019.

[6] C. Mandoul *et al.*, "Breast tomosynthesis: What do we know and where do we stand?," (in English), *Diagnostic and Interventional Imaging,* Review vol. 100, no. 10, pp. 537-551, Oct 2019.

[7] T. Hovda *et al.*, "Screening outcome for consecutive examinations with digital breast tomosynthesis versus standard digital mammography in a population-based screening program," (in English), *European Radiology,* Article vol. 29, no. 12, pp. 6991-6999, Dec 2019.

[8] D. B. Kopans, "Digital Breast Tomosynthesis From Concept to Clinical Care," (in English), *American Journal of Roentgenology,* Review vol. 202, no. 2, pp. 299-308, Feb 2014.

[9] J. A. Tice *et al.*, "Validation of the breast cancer surveillance consortium model of breast cancer risk," (in English), *Breast Cancer Research and Treatment,* Article vol. 175, no. 2, pp. 519-523, Jun 2019.

[10] E. N. Marcus, M. Yepes, and N. Dietz, "Perception of Breast Density Information Among Women in Miami, FL: a Qualitative Study," (in English), *Journal of Cancer Education,* Article; Early Access p. 8.

[11] O. Alonzo-Proulx, J. G. Mainprize, J. A. Harvey, and M. J. Yaffe, "Investigating the feasibility of stratified breast cancer screening using a masking risk predictor," (in English), *Breast Cancer Research,* Article vol. 21, no. 1, p. 9, Aug 2019, Art. no. 91.

[12] M. H. Ross and W. Pawlina, L. W. Williams, Ed. *Histology:A Text and Atlas With Correlated Cell and Molecular Biology*, 6th ed. Wolters Kluwer, 2011.

[13] W. Zhu and C. M. Nelson, "Adipose and mammary epithelial tissue engineering," *Biomatter,* vol. 3, no. 3, Jul-Sep 2013.

[14] T. Li *et al.*, "The association of measured breast tissue characteristics with mammographic density and other risk factors for breast cancer," (in English), *Cancer Epidemiology Biomarkers & Prevention,* Article vol. 14, no. 2, pp. 343-349, Feb 2005.

[15] K. Ghosh *et al.*, "Tissue composition of mammographically dense and non-dense breast tissue," (in English), *Breast Cancer Research and Treatment,* Article vol. 131, no. 1, pp. 267-275, Jan 2012.

[16] M. Mullooly *et al.*, "Application of convolutional neural networks to breast biopsies to delineate tissue correlates of mammographic breast density," (in English), *Npj Breast Cancer,* Article vol. 5, p. 11, Nov 2019, Art. no. 43.

[17] J. N. Wolfe, "BREAST PATTERNS AS AN INDEX OF RISK FOR DEVELOPING BREAST-CANCER," (in English), *American Journal of Roentgenology,* Article vol. 126, no. 6, pp. 1130-1139, 1976.

[18] W. D. He, A. Juette, E. R. E. Denton, A. Oliver, R. Marti, and R. Zwiggelaar, "A Review on Automatic Mammographic Density and Parenchymal Segmentation," (in English), *International Journal of Breast Cancer,* Review p. 31, 2015, Art. no. 276217.

[19] C. M. Vachon *et al.*, "Marnmographic breast density as a general marker of breast cancer risk," (in English), *Cancer Epidemiology Biomarkers & Prevention,* Article vol. 16, no. 1, pp. 43-49, Jan 2007.

[20] C. Diorio *et al.*, "Insulin-like growth factor-I, IGF-binding protein-3, and mammographic breast density," (in English), *Cancer Epidemiology Biomarkers & Prevention,* Article vol. 14, no. 5, pp. 1065-1073, May 2005.

[21] V. A. McCormack and I. D. S. Silva, "Breast density and parenchymal patterns as markers of breast cancer risk: A meta-analysis," (in English), *Cancer Epidemiology Biomarkers & Prevention,* Article vol. 15, no. 6, pp. 1159-1169, Jun 2006.

[22] G. J. Wengert, T. H. Helbich, D. Leithner, E. A. Morris, P. A. T. Baltzer, and K. Pinker, "Multimodality Imaging of Breast Parenchymal Density and Correlation with Risk

Assessment," (in English), *Current Breast Cancer Reports,* Article vol. 11, no. 1, pp. 23-33, Mar 2019.

[23] R. J. Hooley, M. A. Durand, and L. E. Philpotts, "Advances in Digital Breast Tomosynthesis," (in English), *American Journal of Roentgenology,* Review vol. 208, no. 2, pp. 256-266, Feb 2017.

[24] H. R. Peppard, B. E. Nicholson, C. M. Rochman, J. K. Merchant, R. C. Mayo, and J. A. Harvey, "Digital Breast Tomosynthesis in the Diagnostic Setting: Indications and Clinical Applications," (in English), *Radiographics,* Article vol. 35, no. 4, pp. 975-990, Jul-Aug 2015.

[25] M. A. Orsi, M. Cellina, C. Rosti, D. Gibelli, E. Belloni, and G. Oliva, "Digital breast tomosynthesis: A state-of-the-art review," *Nuclear Medicine and Biomedical Imaging,* vol. 3, no. 4, pp. 1-9, 2018.

[26] M. Lyra and A. Ploussi, "Filtering in SPECT image reconstruction," *International journal of biomedical imaging,* vol. 2011, p. 693795, 06/23 2011.

[27] M. Al and M. Al-Hayani, "The Use of Filtered Back projection Algorithm for Reconstruction of tomographic Image," pp. 151-156, 01/01 2014.

[28] I. Sechopoulos, J. Teuwen, and R. Mann, "Artificial intelligence for breast cancer detection in mammography anddigital breast tomosynthesis: State of the art," *Seminars in Cancer Biology,* 2020.

[29] R. C. Chen *et al.*, "Measurement of the linear attenuation coefficients of breast tissues by synchrotron radiation computed tomography," (in English), *Physics in Medicine and Biology,* Article vol. 55, no. 17, pp. 4993-5005, Sep 2010.

[30] S. Azam *et al.*, "Mammographic density change and risk of breast cancer," (in English), *Annals of Oncology,* Meeting Abstract vol. 30, pp. 4-5, May 2019.

[31] M. J. Yaffe, "Mammographic density - Measurement of mammographic density," (in English), *Breast Cancer Research,* Review vol. 10, no. 3, p. 10, 2008, Art. no. 209.

[32] K. Holland, C. H. van Gils, R. M. Mann, and N. Karssemeijer, "Quantification of masking risk in screening mammography with volumetric breast density maps," (in English), *Breast Cancer Research and Treatment,* Article vol. 162, no. 3, pp. 541-548, Apr 2017.

[33] C. M. Vachon *et al.*, "Mammographic density, breast cancer risk and risk prediction," (in English), *Breast Cancer Research,* Review vol. 9, no. 6, p. 9, 2007, Art. no. 217.

[34] J. M. Park, E. A. Franken, M. Garg, L. L. Fajardo, and L. T. Niklason, "Breast tomosynthesis: Present considerations and future applications," (in English), *Radiographics,* Article; Proceedings Paper vol. 27, pp. S231-S240, Oct 2007.

[35] E. U. Ekpo and M. F. McEntee, "Measurement of breast density with digital breast tomosynthesis-a systematic review," (in English), *British Journal of Radiology,* Review vol. 87, no. 1043, p. 9, Nov 2014, Art. no. Unsp 20140460.

[36] J. N. Wolfe, "RISK FOR BREAST-CANCER DEVELOPMENT DETERMINED BY MAMMOGRAPHIC PARENCHYMAL PATTERN," (in English), *Cancer,* Article vol. 37, no. 5, pp. 2486-2492, 1976.

[37] I. Muhimmah, A. Oliver, E. R. E. Denton, J. Pont, E. Perez, and R. Zwiggelaar, "Comparison between Wolfe, Boyd, BI-RADS and Tabar based mammographic risk assessment," in *Digital Mammography, Proceedings*, vol. 4046, S. M. Astley, M. Brady, C. Rose, and R. Zwiggelaar, Eds. (Lecture Notes in Computer Science, Berlin: Springer-Verlag Berlin, 2006, pp. 407-415.

[38] N. F. Boyd *et al.*, "MAMMOGRAPHIC SIGNS AS RISK-FACTORS FOR BREAST-CANCER," (in English), *British Journal of Cancer,* Article vol. 45, no. 2, pp. 185-193, 1982.

[39] N. F. Boyd, G. A. Lockwood, J. W. Byng, D. L. Tritchler, and M. J. Yaffe, "Mammographic densities and breast cancer risk," (in English), *Cancer Epidemiology Biomarkers & Prevention,* Review vol. 7, no. 12, pp. 1133-1144, Dec 1998.

[40] L. Tabar and P. B. Dean, "MAMMOGRAPHIC PARENCHYMAL PATTERNS - RISK INDICATOR FOR BREAST-CANCER," (in English), *Jama-Journal of the American Medical Association,* Article vol. 247, no. 2, pp. 185-189, 1982.

[41] I. T. Gram, E. Funkhouser, and L. Tabar, "The Tabar classification of mammographic parenchymal patterns," (in English), *European Journal of Radiology,* Article vol. 24, no. 2, pp. 131-136, Feb 1997.

[42] U. Sovio *et al.*, "Comparison of fully and semi-automated area-based methods for measuring mammographic density and predicting breast cancer risk," (in English), *British Journal of Cancer,* Article vol. 110, no. 7, pp. 1908-1916, Apr 2014.

[43] S. Petroudi, K. Marias, R. English, and M. Brady, "Classification of mammogram patterns using area measurements and the standard mammogram form (SMF)," *International Workshop on Digital Mammography, Lecture Notes in Computer Science, Springer Verlag Berlin Heidelberg, 2002,* 01/17 2014.

[44] S. Pahwa, S. Hari, S. Thulkar, and S. Angraal, *Evaluation of breast parenchymal density with QUANTRA software*. 2014.

[45] A. Tagliafico, G. Tagliafico, D. Astengo, S. Airaldi, M. Calabrese, and N. Houssami, "Comparative estimation of percentage breast tissue density for digital mammography, digital breast tomosynthesis, and magnetic resonance imaging," (in English), *Breast Cancer Research and Treatment,* Article vol. 138, no. 1, pp. 311-317, Feb 2013.

[46] C. Nickson *et al.*, "AutoDensity: An automated method to measure mammographic breast density that predicts breast cancer risk and screening outcomes," *Breast cancer research : BCR,* vol. 15, p. R80, 09/11 2013.

[47] (Accessed on 10 April 2020). *Volpara Solutions*. Available: https://volparahealth.com/our-products/volparadensity/

[48] M. Nasor and W. Obaid, "Detection and Localization of Early-Stage Multiple Brain Tumors Using a Hybrid Technique of Patch-Based Processing, k-means Clustering and Object Counting," (in English), *International Journal of Biomedical Imaging,* Article vol. 2020, p. 9, Jan 2020, Art. no. 9035096.

[49] W. K. Moon *et al.*, "Quantitative breast density analysis using tomosynthesis and comparison with MRI and digital mammography," (in English), *Computer Methods and Programs in Biomedicine,* Article vol. 154, pp. 99-107, Feb 2018.

[50] R. Caldas, R. Sarai, F. B. D. Neto, and B. Markert, "Validation of two hybrid approaches for clustering age-related groups based on gait kinematics data," (in English), *Medical Engineering & Physics,* Article vol. 78, pp. 90-97, Apr 2020.

[51] L. Rundo *et al.*, "Tissue-specific and interpretable sub-segmentation of whole tumour burden on CT images by unsupervised fuzzy clustering," (in English), *Computers in Biology and Medicine,* Article vol. 120, p. 13, May 2020, Art. no. 103751.

[52] M. Y. Qiao *et al.*, "Breast DCE-MRI radiomics: a robust computer-aided system based on reproducible BI-RADS features across the influence of datasets bias and segmentation methods," (in English), *International Journal of Computer Assisted Radiology and Surgery,* Article; Early Access p. 10.

[53] M. Mustra, M. Grgic, and K. Delac, "Breast Density Classification Using Multiple Feature Selection," (in English), *Automatika,* Article vol. 53, no. 4, pp. 362-372, Oct-Dec 2012.

[54] J. C. Ang, A. Mirzal, H. Haron, and H. N. A. Hamed, "Supervised, Unsupervised, and Semi-Supervised Feature Selection: A Review on Gene Selection," (in English), *Ieee-Acm Transactions on Computational Biology and Bioinformatics,* Review vol. 13, no. 5, pp. 971-989, Sep-Oct 2016.

[55] F. Pereira, T. Mitchell, and M. Botvinick, "Machine learning classifiers and fMRI: A tutorial overview," (in English), *Neuroimage,* Article; Proceedings Paper vol. 45, no. 1, pp. S199-S209, Mar 2009.

[56] K. Clark *et al.*, "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository," (in English), *Journal of Digital Imaging,* Article vol. 26, no. 6, pp. 1045-1057, Dec 2013.

[57] A. Badano *et al.*, "Evaluation of Digital Breast Tomosynthesis as Replacement of Full-Field Digital Mammography Using an In Silico Imaging Trial," (in English), *Jama Network Open,* Article vol. 1, no. 7, p. 12, Nov 2018, Art. no. e185474.

[58] ABing. (Accessed on 10 January 2020). *Spatial Fuzzy Clustering and Level Set Segmentation*. Available: https://www.mathworks.com/matlabcentral/fileexchange/31068-spatial-fuzzy-clustering-and-level-set-segmentation

[59] X. Wei. (Accessed on 12 May 2020). *Histogram features of a gray level image*. Available: https://www.mathworks.com/matlabcentral/fileexchange/17537-histogram-features-of-a-gray-level-image

[60] R. M. Haralick, K. Shanmugam, and I. Dinstein, "TEXTURAL FEATURES FOR IMAGE CLASSIFICATION," (in English), *Ieee Transactions on Systems Man and Cybernetics,* Article vol. SMC3, no. 6, pp. 610-621, 1973.

[61] (Accessed on 1 May 2020). *Graycomatrix*. Available: https://www.mathworks.com/help/images/ref/graycomatrix.html

[62] R. Monzel. (Accessed on 5 May 2020). *haralickTextureFeatures*. Available: https://www.mathworks.com/matlabcentral/fileexchange/58769-haralicktexturefeatures

[63] F. Albregtsen and B. Nielsen, "Texture Classification based on Cooccurrence of Gray Level Run Length Matrices," *Australian Journal of Intelligent Information Processing Systems,* vol. 6, no. 1, pp. 38-45.

[64] X. O. Tang, "Texture information in run-length matrices," (in English), *Ieee Transactions on Image Processing,* Article vol. 7, no. 11, pp. 1602-1609, Nov 1998.

[65] M. M. Galloway, *Texture analysis using gray level run lengths*. University of Maryland, Computer Science Center, 1974.

[66] W. Elferink. (Accessed on 12 May 2020). *Gray Level Run Length Image Statistics*. Available: https://www.mathworks.com/matlabcentral/fileexchange/52640-gray-level-run-length-image-statistics

[67] A. Chu, C. M. Sehgal, and J. F. Greenleaf, "USE OF GRAY VALUE DISTRIBUTION OF RUN LENGTHS FOR TEXTURE ANALYSIS," (in English), *Pattern Recognition Letters,* Article vol. 11, no. 6, pp. 415-419, Jun 1990.

[68] (Accessed on 20 April 2020). *ANACONDA*. Available: https://www.anaconda.com/products/individual

[69] C. X. Ling, J. Huang, and H. Zhang, "AUC: A better measure than accuracy in comparing learning algorithms," in *Advances in Artificial Intelligence, Proceedings*, vol. 2671, Y. Xiang and B. ChaibDraa, Eds. (Lecture Notes in Artificial Intelligence, Berlin: Springer-Verlag Berlin, 2003, pp. 329-341.

[70] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," (in eng), *Biometrics,* vol. 33, no. 1, pp. 159-74, Mar 1977.

[71] K. S. Beevi, M. S. Nair, and G. R. Bindu, "A Multi-Classifier System for Automatic Mitosis Detection in Breast Histopathology Images Using Deep Belief Networks," (in English), *Ieee Journal of Translational Engineering in Health and Medicine-Jtehm,* Article vol. 5, p. 11, 2017, Art. no. 4300211.

[72] Esener, II, S. Ergin, and T. Yuksel, "A New Feature Ensemble with a Multistage Classification Scheme for Breast Cancer Diagnosis," (in English), *Journal of Healthcare Engineering,* Article p. 15, 2017, Art. no. 3895164.

[73] S. Alaliyat, "Video - based Fall Detection in Elderly's Houses," Master of Science in Media Technology, Department of Computer Science and Media Technology, Gjøvik University College, 2008.

[74] A. Tharwat, "Linear vs. quadratic discriminant analysis classifier: a tutorial," (in English), *International Journal of Applied Pattern Recognition,* Article vol. 3, no. 2, pp. 145-180, 2016.

[75] R. Gandhi, "Naive Bayes Classifier," vol. 2020, ed. Towards Data Science, 2018.

[76] E. C. Merkle and V. A. Shaffer, "Binary recursive partitioning: Background, methods, and application to psychology," (in English), *British Journal of Mathematical & Statistical Psychology,* Article vol. 64, no. 1, pp. 161-181, Feb 2011.

[77] J. A. S. Sá, A. d. C. Almeida, B. R. P. d. Rocha, M. A. Mota, J. R. Souza, and L. J. C. Dentel, "Lightning Forecast Using Data Mining Techniques On Hourly Evolution Of The Convective Available Potential Energy," pp. 1-5, 2016.

[78] A. Yadav, "SUPPORT VECTOR MACHINES (SVM)," ed. Towards Data Science, 2018.

[79] A. Chakure, "Random Forest Classification and its implementation in Python," vol. 2020, ed. Towards Data Science, 2019.

[80] M. S. Choudhry and R. Kapoor, "Performance Analysis of Fuzzy C-Means Clustering Methods for MRI Image Segmentation," in *Twelfth International Conference on Communication Networks, Iccn 2016 / Twelfth International Conference on Data Mining and Warehousing, Icdmw 2016 / Twelfth International Conference on Image and Signal Processing, Icisp 2016*, vol. 89, K. R. Venugopal, R. Buyya, L. M. Patnaik, P. D. Shenoy, S. S. Iyengar, and K. B. Raja, Eds. (Procedia Computer Science, Amsterdam: Elsevier Science Bv, 2016, pp. 749-758.

6

Appendix A

1. Instructions Manual

This is a detailed manual about the functioning and tools featured in the ABC application. This application is run through an exe. file and it doesn't require a MATLAB license and therefore the user doesn't need to install MATLAB in the device.

However, the user needs to install MATLAB Compiler Runtime (MCR), which can be found in the following URL: https://www.mathworks.com/products/compiler/mcr.html available for Windows, Linux and Mac.

1.1. Patient Information

On the left-hand side of the interface (in Figure 6.1), the user can find the Patient Information Panel. Here the user can access and edit patient information, such as its name, gender, whether it is the patient's first exam and additional comments that are relevant for the patient's chart like, for example, if the patient is a smoker or, in case of a female patient, whether she has given birth before. Some features in this panel are automatically filled out according to the meta information present in the DICOM images loaded, such as patient's name, gender, thickness compression and *kVp* used in the exam.

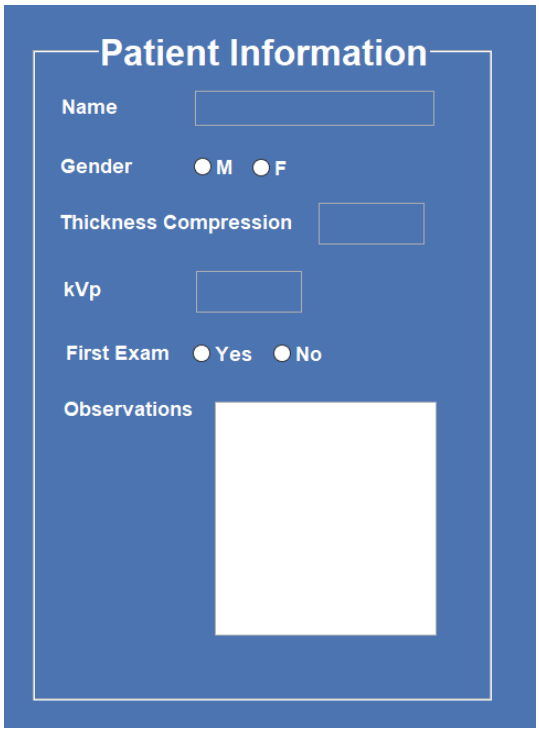

Figure 6.1: Patient Information panel.

1.2. Control Panel

On the right-hand side is where the Control Panel is located. Here is where the user can find the commands to determine the patient's BD upon loading the exam onto the interface and it is illustrated in Figure 6.2.

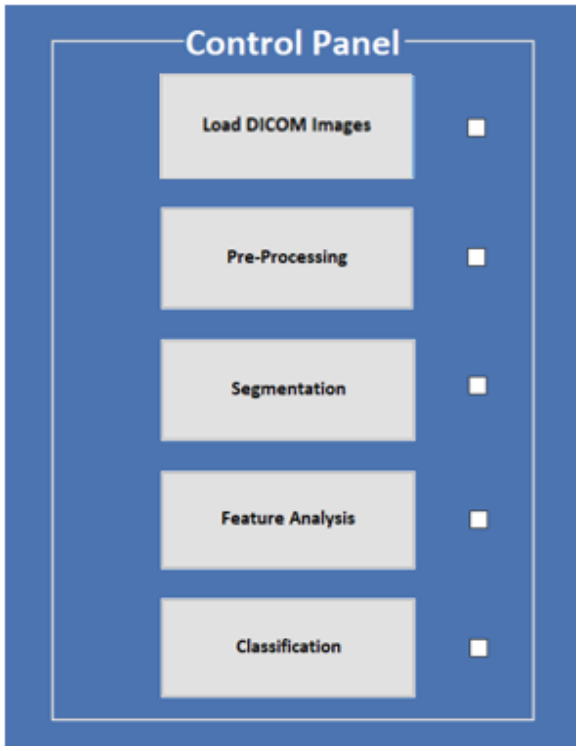

Figure 6.2: Control Panel.

1.3. Commands

The first command is "Load DICOM Images" which, when pressed, asks the user to choose the directory where the DICOM images are located so that they can be loaded and analysed.

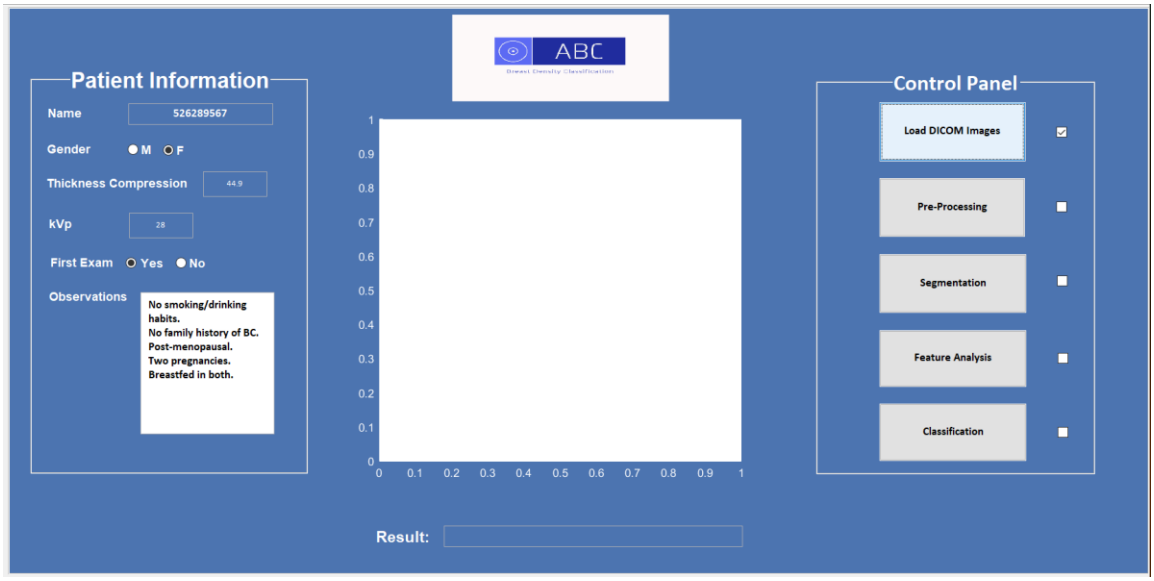

Figure 6.3: Environment after command "Load DICOM Images" is executed.

Secondly, the command to be pressed corresponds to the initial pre-processing part that corresponds to the choice of the amount of breast region to be analysed and it is named "Pre-Processing". Here, the user has to make two decisions: firstly, one has to choose whether one wants to use the whole image without cropping it, i.e., "Use full image" option (that means not choosing a ROI, which is inherently not advised in MLO views for the reasons described in section 2.2.2. of Chapter 2) or if, on the contrary, one wants to choose a ROI to be analysed that will then be applied to all the slices in the patient's exam, "Use cropped image". After the user decides this first step and if the decision is to crop the image, he is asked to choose an option regarding another feature: whether he intends the ROI delimitation to be manual or automatic. If the user chooses the first option, he will be asked to draw a rectangular ROI which he can redo the number of times he likes until he is satisfied with the result. On the other hand, if he chooses "Automatic Mode", a pre-defined rectangular ROI is computed and used to apply to all the slices and posteriorly be evaluated. The two are depicted below in Figures 6.4 and 6.5.

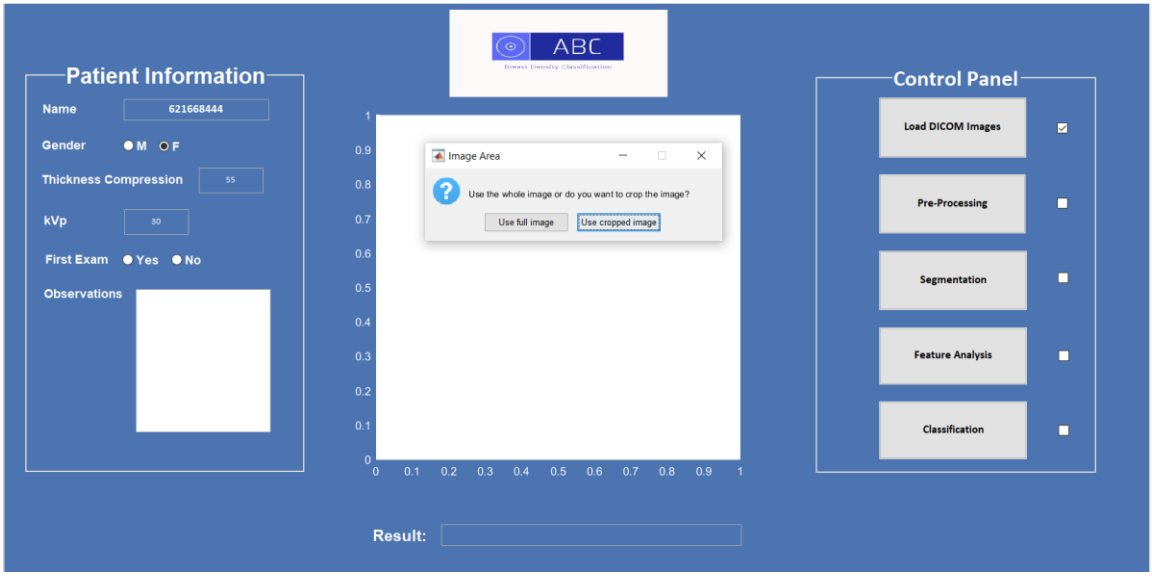

Figure 6.4: Question after command "Pre-Processing" is executed.

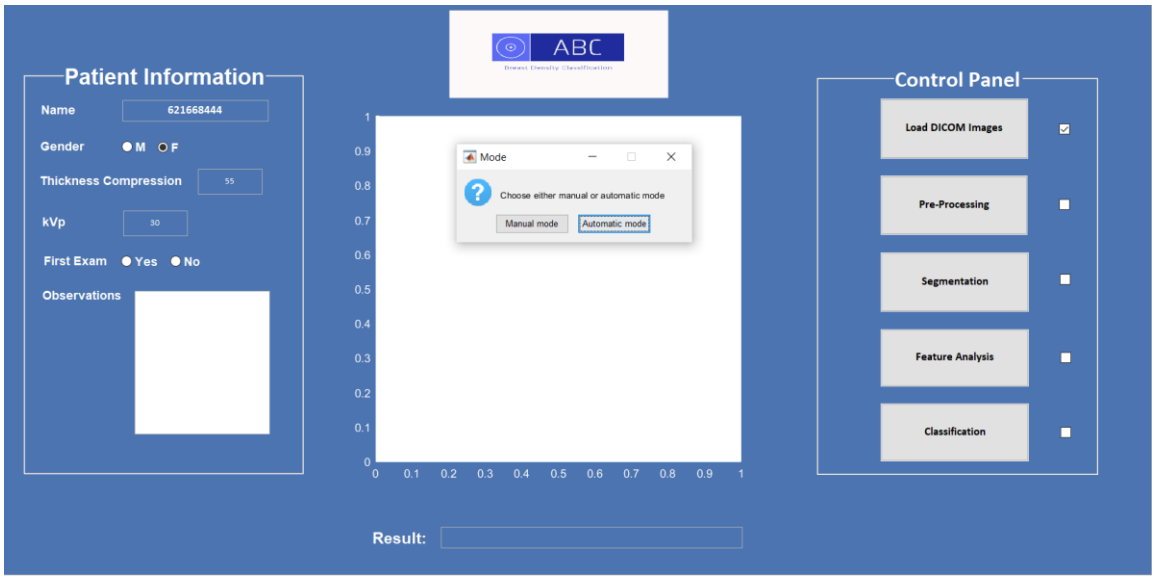

Figure 6.5: Question after "Use cropped image" option is selected.

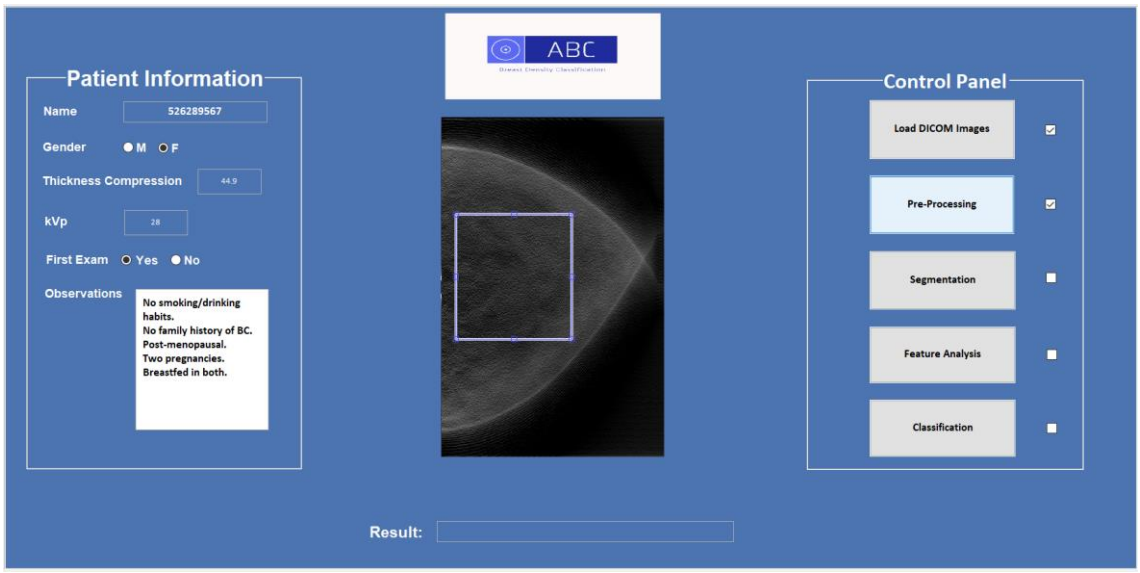

Figure 6.6: Environment after all the command "Pre-Processing" is executed.

The following button is "Segmentation" and it executes the programmed FCM method on all the images, as well as computing the features for each clustered image when pressed. This is the command that is more time-consuming, depending on the number of images in the exam being assessed, and it can take from 3 minutes up to 5 minutes to execute (also depending on the computer's RAM). When segmentation is

complete, the final clustered ROI is displayed in the centre of the interface and this moment is illustrated below in Figure 6.7.

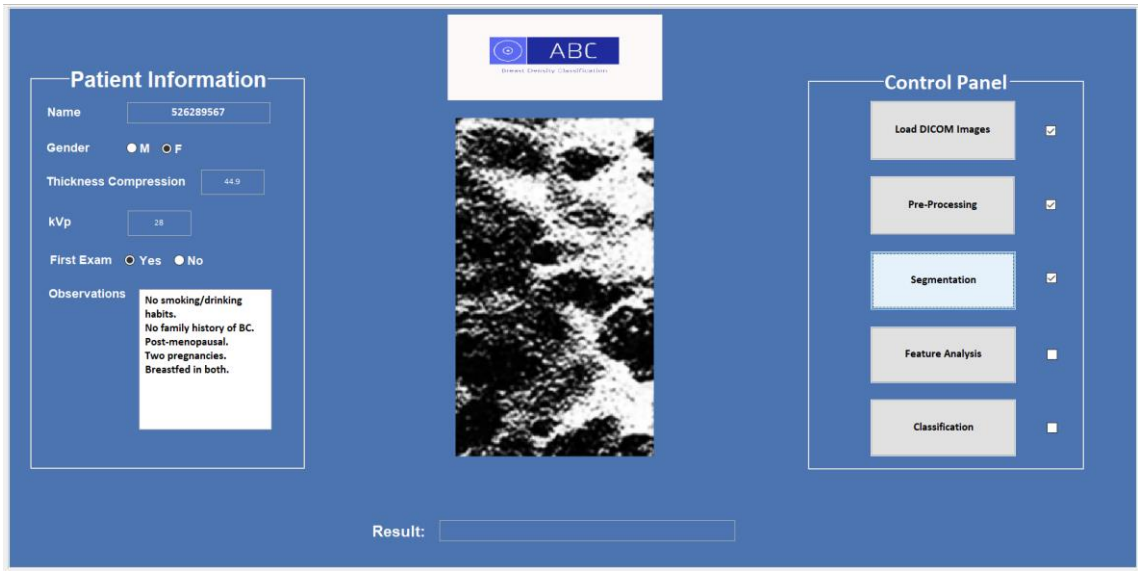

Figure 6.7: Environment after command "Segmentation" is executed.

The last to final step corresponds to the final calculations of the features computed in the previous step and it is titled "Feature Analysis", illustrated in Figure 6.8..

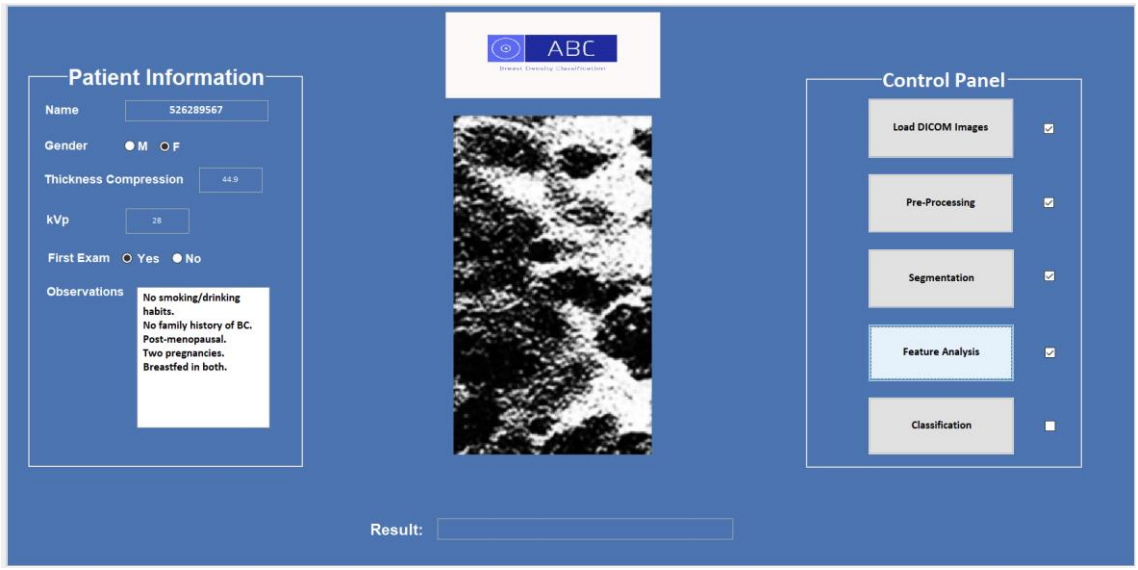

Figure 6.8: Environment after command "Feature Analysis" is executed.

Finally, the last command "Classification" computes and displays the exam's BD classification in the space below after "Result" as one of the four possible classification degrees: ACR1, ACR2, ACR3 or ACR4.

One relevant feature of the interface is the fact that when a command is complete the tick box's value next to it turns to one, so that the user knows when to move on to pressing the next command. In Figure 6.9, the final frame of the application is outlined, depicting the classification result and the tick boxes' values all turned to one.

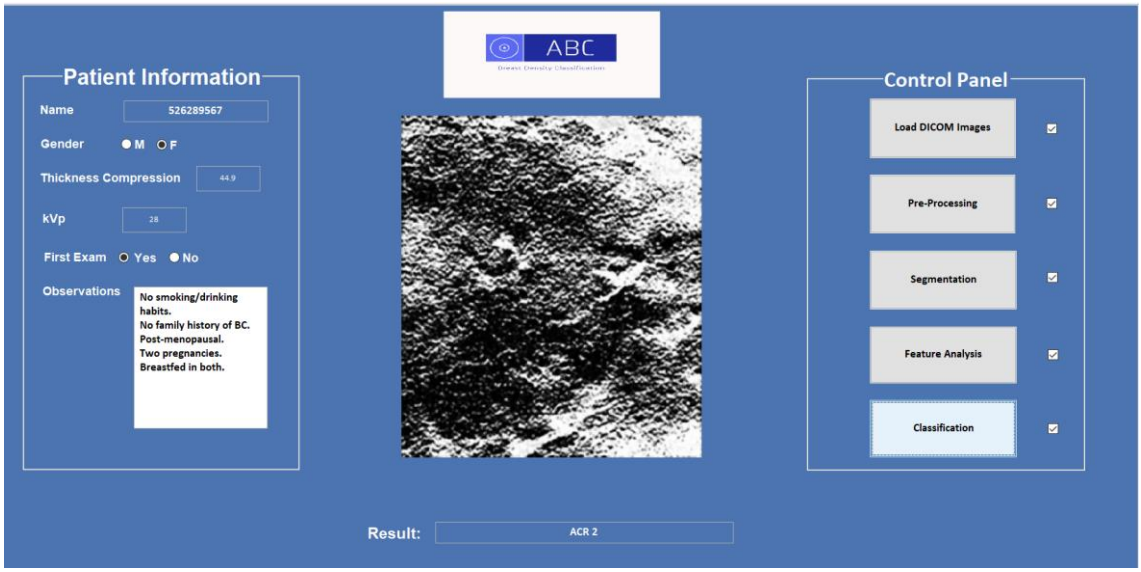

Figure 6.9: Environment after command "Classification" is executed, displaying the final BD result.

Considering the final result, the health professional may refer the patient to a specialized and appropriate treatment according to the BD degree.

7

Appendix B

Table 7.1: Histogram features.

Histogram Feature	Mathematical Expression
Mean	$\frac{1}{N}\sum_{i=1}^{N} X(i)$
Variance	$\frac{1}{N-1}\sum_{i=1}^{N}(X(i) - Mean)^2^{\frac{1}{2}}$
Skewness	$\frac{\frac{1}{N}\sum_{i=1}^{N}(X(i) - Mean)^3}{Variance^3}$
Kurtosis	$\frac{\frac{1}{N}\sum_{i=1}^{N}(X(i) - Mean)^4}{Variance^2}$
Energy	$\sum_{i=1}^{N} X(i)^2$
Entropy	$\sum_{i=1}^{N_l} P(i)\log_2 P(i)$

For the calculations of Haralick features (explicit in Table 7.2), the following notation and mathematical expressions should be considered, as such, by order of appearance [53]:

- *p(i,j)*, *(i,j)*th entry in a normalized GLCM
- *Ng*, number of distinct grey levels in the quantized image
- μ_x, means of the partial probability density function p_x
- μ_y, means of the partial probability density function p_y
- σ_x, standard deviation of the partial probability density function p_x
- σ_y, standard deviation of the partial probability density function p_y
- $p_{x+y}(i)$, probability of co-occurrence matrix coordinates summing to $x+y$, where x and y are the coordinates of an entry in the GLCM (row and column, respectively)
- $p_{x-y}(i)$, probability of co-occurrence matrix coordinates summing to $x-y$, where x and y are the coordinates of an entry in the GLCM (row and column, respectively)
- HXY, entropy of *p(i,j)*
- HXY1 = $-\sum_i \sum_j p(i,j) \log\{p_x(i)p_y(j)\}$
- HX, entropy of p_x
- HY, entropy of p_y
- HXY2 = $-\sum_i \sum_j p_x(i)p_y(j) \log\{p_x(i)p_y(j)\}$

Table 7.2: Haralick features.

Haralick Feature	Mathematical Expression
Angular Second Moment	$\sum_i \sum_j p(i,j)^2$
Contrast	$\sum_{n=0}^{Ng-1} n^2 \{\sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p(i,j)\}, \lvert i-j \rvert = n$
Correlation	$\frac{\sum_i \sum_j (ij) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
Variance	$\sum_i \sum_j (i-\mu)^2 p(i,j)$
Inverse Difference Moment	$\sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j)$
Sum Average	$\sum_{i=2}^{2Ng} i p_{x+y}(i)$
Sum Variance	$\sum_{i=2}^{2Ng} (i-f_s)^2 p_{x+y}(i)$
Sum Entropy	$-\sum_{i=2}^{2Ng} p_{x+y}(i) \log\{p_{x+y}(i)\} = f_s$
Entropy	$-\sum_i \sum_j p(i,j) \log(p(i,j))$
Difference Variance	$\sum_{i=0}^{Ng-1} i^2 p_{x-y}(i)$
Difference Entropy	$-\sum_{i=0}^{Ng-1} p_{x-y}(i) \log\{p_{x-y}(i)\}$
Information Measure of Correlation I	$\frac{HXY - HXY1}{max\{HX, HY\}}$
Information Measure of Correlation II	$(1 - exp[-2(HXY2 - HXY)])^{\frac{1}{2}}$
Maximal Correlation Coefficient	Square root of the second largest eigenvalue of Q, where $Q(i,j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$
Autocorrelation	$\sum_i \sum_j (i \cdot j)\, p(i,j)$
Cluster Prominence	$\sum_i \sum_j (i + j - \mu_x - \mu_y)^4 p(i,j)$
Cluster Shade	$\sum_i \sum_j (i + j - \mu_x - \mu_y)^3 p(i,j)$

Dissimilarity	$\sum_i \sum_j \lvert i-j \rvert \cdot p(i,j)$
Maximum Probability	$max\ \{p(i,j)\}$

The run-length features computed in the present work are explicit in Table 7.3, where:

- *p(i,j)*, run-length matrix
- *i*, grey level
- *j*, run-length
- *n*, total number of runs

Table 7.3: Run-length features.

Run-length Feature	Mathematical Expression
Short Run Emphasis	$\frac{1}{n}\sum_{i,j}\frac{p(i,j)}{j^2}$
Long Run Emphasis	$\frac{1}{n}\sum_{i,j} p(i,j) \cdot j^2$
Grey Level Non-Uniformity	$\frac{1}{n}\sum_{i}\left[\sum_{j} p(i,j)\right]^2$
Run Percentage	$\sum_{i,j}\frac{n}{p(i,j)\cdot j}$
Low Grey Level Run Emphasis	$\frac{1}{n}\sum_{i,j}\frac{p(i,j)}{i^2}$
High Grey Level Run Emphasis	$\frac{1}{n}\sum_{i,j} p(i,j) \cdot i^2$