

Instituto Superior de Engenharia do Porto – Instituto Politécnico do Porto



Aplicação de técnicas de *data mining* para suporte ao diagnóstico de cancro da mama

Using data mining techniques to support breast cancer diagnosis

Departamento de Física

Joana Moreira Diz

Dissertação submetida no Instituto Superior de Engenharia do Porto para a obtenção
do grau de Mestre em Engenharia de Computação e Instrumentação Médica

Orientador Científico: Doutora Goreti Marreiros, ISEP

Coorientador Científico: Doutor Alberto Freitas, FMUP

2014

Acknowledgements

The preparation of a master's thesis can be extremely exasperating but with the help of many people and support for various purposes, at the end concludes that everything is possible. Thereby, there are some people whom I have to offer my sincere thanks, because it was with the help of each one that I could get here.

Firstly, I would like to thank Dr. Jaime Cardoso, representative of the Breast Research Group and that allowed me access to the INbreast[®] database.

I would like to thank to Prof. Dr. Goreti Marreiros for helping me throughout the process of improving the initial objective of this thesis and throughout the follow-up given along the way.

I also would like to thank Prof. Dr. Alberto Freitas for all the support given and all the hours of answering questions that were extremely useful.

To Prof. Dr. Inês Dutra for the meetings that gave me some new ideas and more knowledge in the breast cancer research field, and thanks for giving me confidence in the work developed through showing interest in my results.

I would also like to thank to Prof. Paulo Oliveira and Prof. Dr. Manuel Cruz for all the helpful hints about the software used.

To a great friend and reviewer of all the writing of this thesis, Sara Tavares, there are no words for how much you helped me, thank you for everything.


To Nuno, definitely would not be the same thesis without your support. Thanks for enduring all the bad times and to emphasize the good ones. You inspired me to always improve.

To Sofia and my other friends who helped me throughout the initial research, I am very thankful and also to Sara and Angela who helped me in the final phase.

Finally, I would like to thank to my family, especially my parents, for during all these years have never given up on me and for all the support and strength you have provided to me.


My sincere thanks to all,

Joana Moreira Diz



“Some people spend
the whole of their
lives sitting waiting for
one train, only to find
that they never even
made it to the
station.”

- Joanne Harris *in*
Peaches for
Monsieur le Curé



Abstract

More than ever, there is an increase of the number of decision support methods and computer aided diagnostic systems applied to various areas of medicine. In breast cancer research, many works have been done in order to reduce false-positives when used as a double reading method.

In this study, we aimed to present a set of data mining techniques that were applied to approach a decision support system in the area of breast cancer diagnosis. This method is geared to assist clinical practice in identifying mammographic findings such as microcalcifications, masses and even normal tissues, in order to avoid misdiagnosis.

In this work a reliable database was used, with 410 images from about 115 patients, containing previous reviews performed by radiologists as microcalcifications, masses and also normal tissue findings.

Throughout this work, two feature extraction techniques were used: the gray level co-occurrence matrix and the gray level run length matrix.

For classification purposes, we considered various scenarios according to different distinct patterns of injuries and several classifiers in order to distinguish the best performance in each case described. The many classifiers used were Naïve Bayes, Support Vector Machines, k-nearest Neighbors and Decision Trees (J48 and Random Forests).

The results in distinguishing mammographic findings revealed great percentages of PPV and very good accuracy values. Furthermore, it also presented other related results of classification of breast density and BI-RADS[®] scale.

The best predictive method found for all tested groups was the Random Forest classifier, and the best performance has been achieved through the distinction of microcalcifications.

The conclusions based on the several tested scenarios represent a new perspective in breast cancer diagnosis using data mining techniques.

Resumo

Cada vez mais assistimos a um aumento global do número de métodos de apoio a decisão e diagnóstico assistido por computador, aplicados a diversas áreas da medicina. Na área de investigação do cancro da mama muitos são os trabalhos que têm sido desenvolvidos como segunda leitura de modo a reduzir o número de falsos positivos no diagnóstico.

Neste estudo é apresentado um conjunto de técnicas de *data mining* que poderão ser aplicadas a um sistema de apoio à decisão na área do diagnóstico de cancro da mama. Esta abordagem tem por objetivo ajudar os clínicos na identificação de achados mamográficos como microcalcificações, massas e mesmo tecidos normais, de forma a evitar diagnósticos errados.

Para isso, neste trabalho é usada uma base de dados fidedigna, de 410 imagens correspondentes a 115 pacientes, contendo análises prévias, realizadas por radiologistas, de microcalcificações, massas e tecidos considerados normais.

Ao longo deste trabalho são utilizadas duas técnicas de extração de características, a matriz de coocorrência de níveis de cinza e a matriz de comprimento da linha de níveis de cinza.

Para a classificação foram considerados diferentes cenários de acordo com diferentes padrões de distinção de lesões e ainda vários classificadores de forma a distinguir as melhores performances em cada caso descrito. Os vários classificadores usados foram *Naïve Bayes*, *Support Vector Machines*, *k-nearest Neighbors* e *Decision Trees (J48 e Random Forests)*.

Os resultados obtidos na distinção dos achados mamográficos revelaram percentagens de valor preditivo positivo e de precisão bastante boas. São ainda apresentados outros resultados relacionados com sistemas de classificação de densidade mamária e escala BI-RADS®.

O melhor método de previsão encontrado, perante todos os grupos testados, foi o classificador *Random Forest* e o melhor desempenho foi conseguido através da distinção de microcalcificações.

As conclusões feitas ao longo dos vários cenários testados foram interessantes em termos que representam uma nova perspetiva no diagnóstico do cancro da mama, utilizando técnicas de *data mining*.

Contents

List of Figures	XIII
List of Tables.....	XV
Abbreviations	XVII
Chapter 1	- 1 -
Introduction	- 1 -
1.1. Framing the issue.....	- 1 -
1.2. Motivation.....	- 2 -
1.3. Specific objectives	- 3 -
1.4. Main structure of the document	- 3 -
Chapter 2	- 5 -
Breast Cancer	- 5 -
2.1 Mammography	- 7 -
2.2 Lexicons and Terminologies.....	- 8 -
2.3 Breast Cancer types	- 13 -
2.4 Evaluation Metrics.....	- 15 -
Chapter 3	- 17 -
State of Art	- 17 -
3.1 Classification Methods	- 17 -
3.2.1 Artificial neural network.....	- 18 -
3.2.2 K - nearest Neighbor.....	- 19 -
3.2.3 Support Vector Machine	- 19 -
3.2.4 Decision Trees	- 20 -
3.2.5 Random Forest.....	- 21 -
3.2.6 Naïve Bayes	- 22 -
3.2 Related Work.....	- 22 -
Chapter 4	- 27 -
Methodology	- 27 -
4.1. Initial Data Comprehension	- 28 -
4.2 Initial Data Preprocessing	- 29 -
4.3 Features Extraction.....	- 31 -
4.3.1 GLCM	- 31 -

4.3.2 GLRLM.....	- 34 -
4.4 Features Selection	- 35 -
4.5 Data Set Organization	- 37 -
4.6 Classification Methods.....	- 38 -
4.6.1 Random Forest	- 38 -
4.6.2 SVM.....	- 39 -
4.6.3 k-nearest Neighbor	- 40 -
4.6.4 Decision Tree	- 40 -
4.6.5 Naïve Bayes	- 40 -
4.6.6 Ensemble Methods.....	- 41 -
Chapter 5	- 43 -
Results and Discussion.....	- 43 -
5.1. SPSS® outcome	- 43 -
5.2. WEKA® outcome.....	- 44 -
Chapter 6	- 61 -
Conclusions and Further Work	- 61 -
References.....	- 65 -

List of Figures

Figure 2.1: Internal mammary constituents	5
Figure 2.2: Quadrants of the Breast	6
Figure 2.3: Example of mammographic images with BI-RADS 1 ranking in the left breast with no pathological changes and with low breast density	10
Figure 2.4: Example of mammographic images with BI-RADS 2 rating in the left breast with no pathological changes and with low breast density	10
Figure 2.5: Example of mammographic images with BI-RADS 3 ranking in the right breast and with low breast density.	11
Figure 2.6: Example of mammographic images with BI-RADS 4 classification in the right breast and with low breast density	11
Figure 2.7: Example of mammographic images with BI-RADS 5 rating in the left breast and with low breast density.	12
Figure 2.8: Example of mammographic images with BI-RADS 6 ranking in the left breast and with low breast density	12
Figure 2.9: Breast density Scale of ACR.....	13
Figure 2.10: Example of a cranium-caudal (CC) image	14
Figure 2.11: Mammogram example in oblique lateral median (MLO) with bilateral asymmetry.....	14
 Figure 3.1: Neural Network system.....	21
Figure 3.2: Separating hyperplanes in SVM method	23
Figure 3.3: A simple decision tree model.....	24
 Figure 4.1: General block diagram of the propose methodology.....	27
Figure 4.2: Initial data graphic with the percentage of findings reported for each breast density class.....	28
Figure 4.3: Initial data analysis graphic showing the percentage of findings reported for each BIRADS class.	29
Figure 4.4: Matlab selection of region of interest on a mammogram.....	30
Figure 4.5: GLCM construction model for 0°	32
Figure 4.6: Illustration of Random Forest nodes.....	39

Figure 6.1: Graphic of results extracted from scenario 1 for benign class	46
Figure 6.2: Graphic of results extracted from scenario 1 for malignant class	46
Figure 6.3: Graphic of results extracted from scenario 2 related to class 1 of BrD	50
Figure 6.4: Graphic of results extracted from scenario 2 related to class 2 of BrD	51
Figure 6.5: Graphic of results extracted from scenario 2 related to class 3 of BrD	51
Figure 6.6: Graphic of results extracted from scenario 2 related to class 4 of BrD	51
Figure 6.7: Graphic of results extracted from scenario 3a related to mass class	55
Figure 6.8: Graphic of results extracted from scenario 3a related to microcalcification class.....	55
Figure 6.9: Graphic of results extracted from scenario 3b related to microcalcification class.....	57
Figure 6.10: Graphic of results extracted from scenario 3b related to no micro class ..	58

List of Tables

Table 2.1: Breast Imaging Reporting and Database System (BIRADS®).....	9
Table 3.1: Summary of work developed in the texture-feature analysis	25
Table 3.2: Summary of work developed for breast cancer diagnostic	26
Table 4.1: Haralick features equations	34
Table 4.2: Characteristics extracted from GLRLM	35
Table 4.3: Significance level for each characteristic, obtained in SPSS t-test outputs ..	37
Table 5.1: Results for GLRLM with BI-RADS® subclass classification	44
Table 5.2: Results for GLRLM with BI-RADS® for ensemble methods	45
Table 5.3: Results of GLCM with BrD	48
Table 5.4: Results of GLCM with BrD for ensemble methods	49
Table 5.5: Results of 3a subgroup	53
Table 5.6: Results of 3a subgroup with ensemble methods	54
Table 5.7: Results of 3b subgroup	56
Table 5.8: Results of 3b subgroup with ensemble methods	57
Table 5.9: Results of 3c subgroup	59

Abbreviations

ACR: American College of Radiology
ANN: Artificial Neural Network
AUC: Area Under the ROC Curve
BCDR: Breast Cancer Digital Repository
BI-RADS: Breast Imaging Reporting and Database System
BrD: Breast Density Classification
CAD: Computer-Aided Diagnosis
CC: Cranium-Caudal
DCIS: Ductal Carcinoma in Situ
DES: Dietilestilbestrol
DICOM: Digital Imaging and Communications in Medicine
DNA: Deoxyribonucleic Acid
DT: Decision Trees
GLCM: Gray-Level Co-occurrence Matrix
GLRLM: Gray-Level Run Length Matrix
IARC: International Agency for Research on Cancer
k-NN: k- nearest Neighbor
LBP: Local Binary Pattern
LCIS: Lobular Carcinoma in Situ
MATLAB: MATrix LABoratory
MIAS: Mammographic Image Analysis Society
MLO: Median Lateral Oblique
PPV: Positive Predicted Value
QIE: Inferior External Quadrant of Breast
QII: Inferior Internal Quadrant of Breast
QSE: Superior External Quadrant of Breast
QSI: Superior Internal Quadrant of Breast
ROC: Receiver Operating Characteristic

ROI: Region of Interest

SPSS: Statistical Package for the Social Sciences

SVM: Support Vector Machine

WEKA: Waikato Environment for Knowledge Analysis

Chapter 1

Introduction

The main goal of this work is the development of an approach that could be implemented for a decision support system which might help oncologists to better classify and diagnose breast cancer with the use of data mining techniques.

For this purpose, throughout this work we will present some of the reasons why we chose breast cancer diagnosis as a subject, as well as some advantages of their application based on decision support methods.

The following chapters present the framework and motivation for the subject, as well as the specific objectives and structure of the document.

1.1. Framing the issue

Breast cancer is globally the most prevalent cancer among women over 40 years of age (1). Several studies have revealed that the early detection of breast cancer considerably reduces the mortality and morbidity caused by this disease (1-3).

Currently, there are several programs for screening and prevention of breast cancer based on the analysis of mammograms.

However this technique requires a specialized interpretation that takes considerable time. Also, it is an imperfect exam, with many agreeing that this test has a sensitivity of about 85 to 90% (4, 5). In order to improve the effectiveness of this method, systems for computer aided diagnosis (CAD) are arising and according to several authors this is leading to an incremental improvement of detection for this type of cancer by

coordinating the work of the radiologist with the implementation of the CAD system (1, 4, 6).

In the area of breast cancer research, various diagnostic support systems have been developed and discussed in terms of future potential (7-11), such as those enumerated later in this work (Chapter III). The same happened in the case of lung and colon cancer (12) .

In this context, the aim of this work was the development of a diagnostic system in clinical practice to identify breast cancer.

1.2. Motivation

According to data collected by the “Liga Portuguesa contra o cancro” (13), breast cancer is a public health problem that, despite not being the most lethal, has a high incidence and mortality rate, especially among women. Of these data, it is also estimated that in Portugal, with a female population of about 5 million people, about 4,500 new cases of breast cancer are detected annually, the equivalent to 11 new cases per day, and about 1,500 women die from this disease each year, or roughly 4 cases per day.

In a global way, and to the International Agency for Research on Cancer (IARC), in 2012 there were estimated close to 1.7 million breast cancer diagnoses, around 11.9% of all worldwide cancers diagnosed in the same period, and about 522,000 women worldwide have died from the disease in the same year (14).

In the same study, 19.3 million new cancer cases were expected by year 2025, in which the largest increase will be in breast cancer (1, 14).

Consequently, the worldwide incidence of breast cancer showed a continuous growth in the last decade, most probably related to the unstable and uneven socio-economic development, and also demographic factors affecting the provision and accessibility to health services (15) . The increased life expectancy also could be linked to the higher breast cancer incidence nowadays, since a greater number of people live longer than before, approaching ever closer to the maximum lifetime of about 120 years, and therefore increasing the risk of cancer.

In an approach to counteract this trend, there is a need for implementation of early diagnosis and patient monitoring systems, as well as better and more approachable

health care. The quicker advancement of technology and increasingly computerized means allows the realization of new platforms for medical care, which include electronic recording and decision support systems.

1.3. Specific objectives

The main objective of this project relies on the creation of an effective toolkit to detect breast cancer by the application of data mining techniques. By this we intend to create a decision support approach that can be used either to assist professionals or to help students in the correct identification using learning platforms.

The application of classifying methods was later evaluated with the aim to determine which features and classifiers have better performance in identifying features on mammograms.

The main contributions of this work are:

- The extraction of image features from a mammographic database using several approaches;
- A feature selection methodology based on statistical correlations between the values of each image processed;
- Evaluation of the contribution of classification methods in order to detect and classify mammograms findings.

1.4. Main structure of the document

Throughout this document different aspects will be addressed to form a better understanding of breast cancer, with special emphasis on types of associated breast pathologies and their classification in accordance with internationally recognized standards.

In the next chapter we will present some relevant facts concerning breast cancer disease, a brief explanation about the mammographic technique and also the criteria used for breast density and cancer stage classification. In chapter 3 we present some of the most important CAD systems developed and also some classification methods which were used in related works.

Our approach is later presented in chapter 4, where the methodology adopted is explained. In chapter 5 results as well as their discussion are revealed. At the end of the document, in chapter 6, conclusions reached throughout the work as well as some final thoughts are presented.

Chapter 2

Breast Cancer

Breast cancer is a disease that reflects an uncontrolled growth of breast cells in the event of an error in its DNA sequence. This disarray normally occurs in women, but can also manifest in men, with the probability of only about 1 in 100 developing cancer in men (16) .

The proliferation of the disease occurs through the lymphatic system of the breast and due to the extent of this system, it is often an obstacle for cancer detection and analysis of their level of development. Nevertheless, when cancer cells are found in the lymph nodes of the breast there is a high probability that these spread to the bloodstream and consequently create metastases elsewhere in the body (3, 5, 16).

In breast cancer, about 95% of these are carcinomas once they arise as breast epithelial elements (17). In turn, usually about 80% of carcinomas are originated in the mammary ducts (DCIS) and 20% lobes (LCIS), as shown in Figure 2.1 (18, 19).

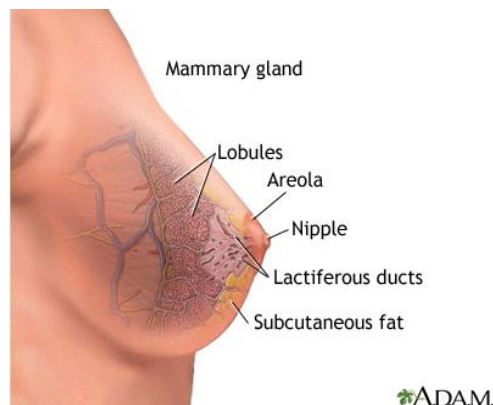


Figure 2.1: Internal mammary constituents (19).

In the clinical practice, some general methods are used for subdividing the breast into small areas in order to better describe the state and location of the lesions. The system of quadrants is the most used method, outlining four quadrants with reference to the position of the nipple, as shown in Figure 2.2 (a). Another system that has come into disuse but is still applied by some professionals is the system clock, also shown in Figure 2 (b).

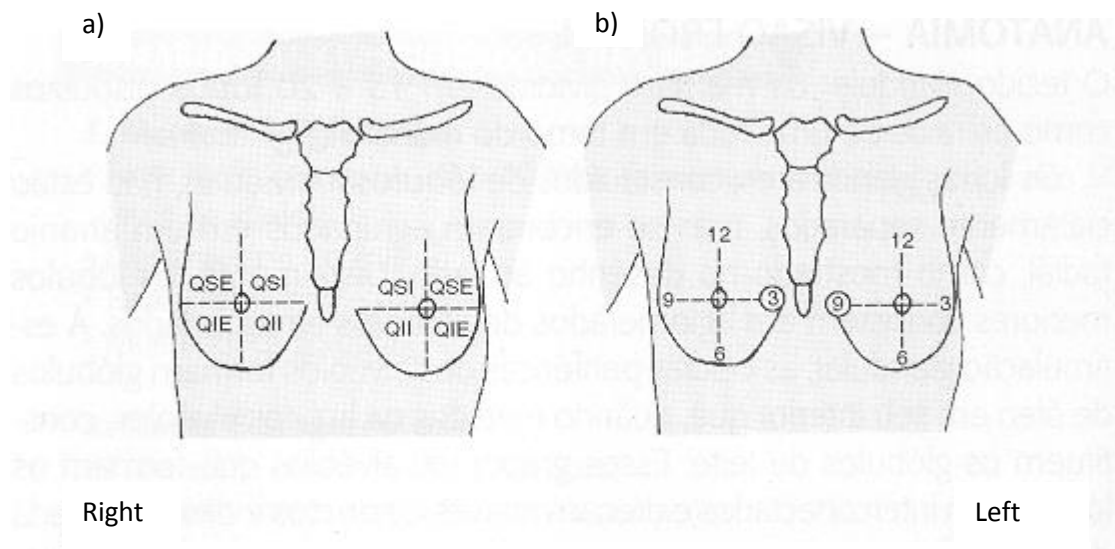


Figure 2.2: Quadrants of the Breast (20).

For breast cancer diagnosis some risk factors must be evaluated, although individually they cannot directly quantify the probability of risk, some of them that must be considered as a group of risk factors.

A risk factor is defined as any variable that alters the probability of having a particular disease (16). Given a set of risk factors for a cause, however, cannot indicate that a particular individual has the disease, but rather provides a marker and a better way of performing the clinical diagnosis and patient monitoring.

The following list presents the summary of some risk factors (16, 19, 21):

- **Gender:** About 99% of cases occur in women;
- **Age:** Increased risk of 1/8 for women under 45 years and 2/3 for women over 50;

- **Genetic Factor:** 5 to 10% of the cancers are hereditary cases, from which 80% are mutations in the BRCA1 gene and 45% represent mutations in the BRCA2 gene.
- **Menstrual cycle:** The risk increased in cases of menarche before age 12 and menopause after 55.
- **Family history:** Doubles the risk in case of family in first order with the disease. About 20-30% of women with breast cancer have a family history of the disease.
- **DES treatment** (diethylstilbestrol): Daughters born to the mother that in her pregnancy has taken DES as synthetic estrogen to reduce the risk of miscarriage, have an increased risk of breast cancer. (DES was a commonly prescribed drug between 1940 and 1960)
- **Previous treatment with radiation:** Significant increase in the risk of breast cancer varies with patient age at the radiation date. Children or young people who receive radiation treatment to the chest area have much higher risk. On the other hand in the case of chemotherapy, it can stop ovarian hormone production for some time which decreases the risk.
- **Hormone replacement therapy:** Increased risk of breast cancer if estrogen therapy is performed for more than 1 year.
- **Dense breast tissue:** Dense breasts have a higher risk of cancer. However dense breast tissue can also lead to less accurate mammography. This factor is considered in more detail later in the document (Chapter 2.4).

2.1 Mammography

The favorable survival rate in breast cancer is due to two factors. The first one relies on the detection of the disease at an early stage through mammograms. The second factor is due to advances in adjuvant systemic treatment such as chemotherapy and hormonal therapy for example.

Mammography is a technique which uses X-rays directed to breast tissue, and usually has two perspectives for the same breast, the cranium-caudal (CC) and median lateral oblique (MLO) already mentioned above in Figures 2.3 and 2.4. In recent years this technique has been the most effective method of early diagnosis, being the only area where radiology is possible to identify a cancer in the initial state (22) .

Breast X-rays were performed for over more than 70 years, however the most modern practices, and the ones used today, have been in practice since 1969. These devices use low levels of radiation (about 0.1 to 0.2 rads per image), which does not present risk to the affected cells (16).

The goal of mammography is to produce detailed images of high resolution and contrast, in order to have a probability of detecting cancer at about 85% before it is revealed, and about 95% of accuracy in detecting a negative case (22).

At the same level to the advancement of technology and interest in the area, mammography emerges as the most effective technique for detecting a clinical picture of cancer. However, their interpretation is difficult and requires considerable time due to variability of anomalies that may be presented with test results.

Due to the fact that mammography isn't the optimum solution, assisted diagnosis and methods of second reading for an intelligent system (CAD) emerged (1).

It's a well-known fact that a standard mammography image equals to 12 bits or the equivalent of 4096 gray levels (2^{12}) but it's also a fact that the perception of the human eye is restricted to 16 gray levels 2^4 (23). Therefore, using computing algorithms for detecting minimal lesions and extracting the maximum information possible from mammograms is a key to transcend human limitations of manual perception.

Thus we intend to take the knowledge of mammography images, which is already quite reliable in the diagnosis and adjust it so that clinical practice and medical diagnosis can be faster and more efficient in the detection of breast cancer.

2.2 Lexicons and Terminologies

In order to believe and ensure quality of collection and processing of data collected through mammography, the development of international guidelines has become essential to ensure the quality of data supplied by various sources (21) .

The American College of Radiology (ACR) (24) is the copyright owner of a work entitled "Breast Imaging Reporting and Database System (BI-RADS[®])" that contains a guide for the standardized mammographic opinion, including a lexicon of terminology, an organization of medical report as well as an evaluation framework and coding system (Table 2.1). The BI-RADS[®] then suggests a standardized method for reporting breast imaging that does not presume to dictate individual decisions of case management.

Table 2.1: Breast Imaging Reporting and Database System (BI-RADS®)

Category	Interpretation	Recommendations and Follow-Up	Risk of Malignancy
BI-RADS 0	Inconclusive Incomplete review.	Additional images and / or exams comparison with earlier ones to a final evaluation test.	
BI-RADS 1 (figure 2.3)	Normal Unchanged.	Annual routine mammogram (women over 40 years).	0.05%*
BI-RADS 2 (figure 2.4)	Benign findings.	Annual routine mammogram (women over 40 years).	0.05%*
BI-RADS 3 (figure 2.5)	Findings probably benign.	Initial follow-up shorter (examination every 6 months).	Up to 2%
BI-RADS 4 (figure 2.6)	Anomalies suspected consideration of biopsy. Subdivisions: 4A: Findings with low suspicion for malignancy. 4B: Lesion with intermediate malignancy 4C: Findings for moderate concern, but are not common malignancy.	Usually requires biopsy to confirm.	> 20%
BI-RADS 5 (figure 2.7)	Highly suggestive malignancy - Appropriate action should be taken.	Requires biopsy or surgical treatment.	> 75 %
BI-RADS 6 (figure 2.8)	Malignancy proven by biopsy did not undergo surgery / treatment.	Category reserved for lesions identified on the image with biopsy before starting definitive therapy for cancer removal.	100 %

* Probability of a woman over the age 40 to develop breast cancer

(table adapted in accordance with American College of Radiology and American Cancer Society) (16, 24)

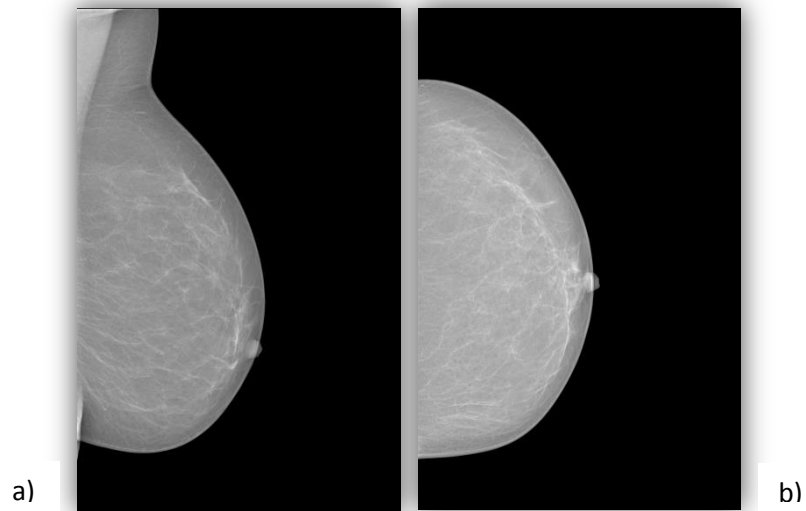


Figure 2.3: Example of mammographic images with BI-RADS[®] 1. **Source:** INbreast, 2011 (60).

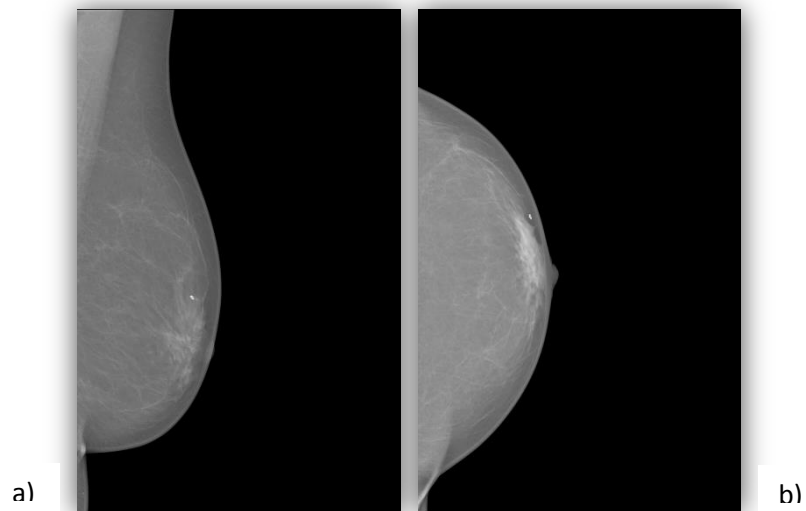


Figure 2.4: Example of mammographic images with BI-RADS[®] 2. **Source:** INbreast, 2011 (60).

Figures 2.3 and 2.4 present some examples taken from the INbreast, database content, which represent mammographic images with BI-RADS[®] 1 and 2, respectively. Both breast images, for each figure, are ranking in the left breast with no pathological changes and with low breast density. The presented incidences are marked as a) MLO and b) CC.

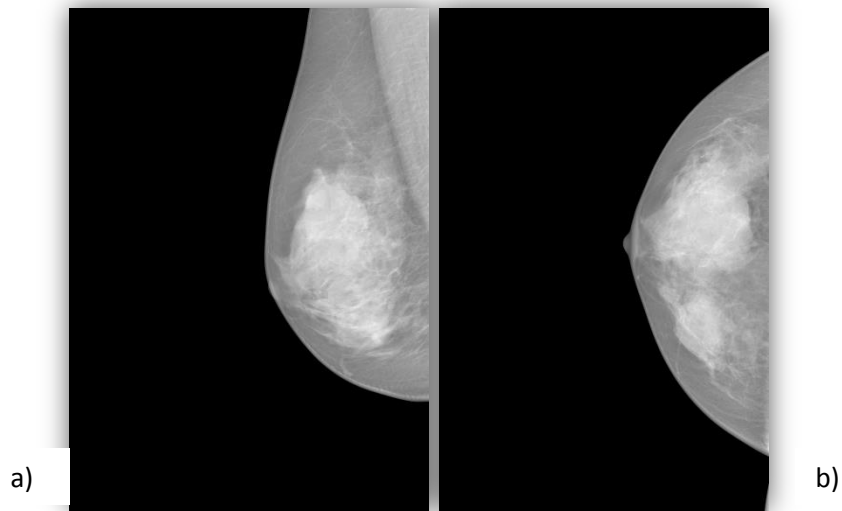


Figure 2.5: Example of mammographic images with BI-RADS® 3 **Source:** INbreast, 2011 (60).

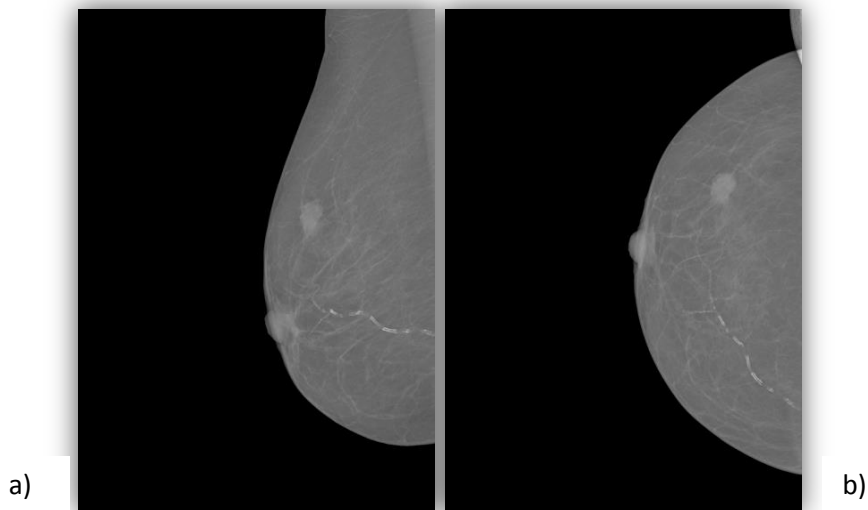


Figure 2.6: Example of mammographic images with BI-RADS® 4. **Source:** INbreast, 2011 (60).

Figures 2.5 and 2.6 are also taken from the INbreast, database content and according to their legend, represent mammograms with BI-RADS® classifications as 3 and 4. All these images are showing the right breast of the patient and have low breast density. Once more the incidences are marked as MLO and CC as a) and b), respectively.

In turn, the next figures (2.7 and 2.8) both represent a left breast with low breast density and the same incidences as referenced above. These last four examples of mammograms represent the last two categories of BI-RADS® classification as described at the legend of each figure.

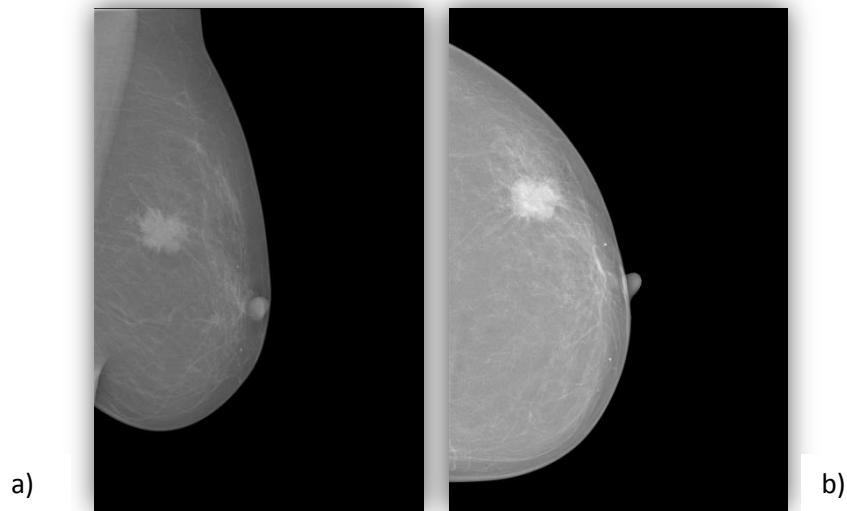


Figure 2.7: Example of mammographic images with BI-RADS® 5. **Source:** INbreast, 2011 (60).

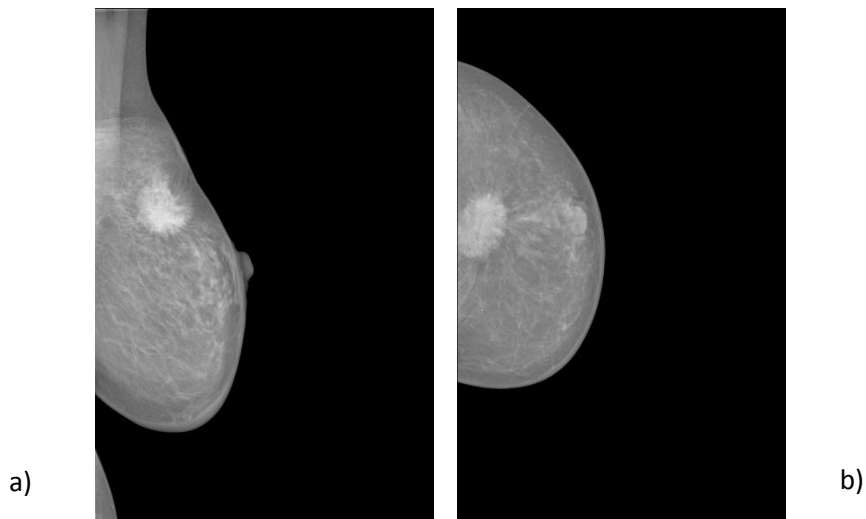


Figure 2.8: Example of mammographic images with BI-RADS® 6. **Source:** INbreast, 2011 (60).

According to a study by Boyd (25) and later confirmed by several other studies (26, 27), it was found that breast density is the most important factor that influences the mammographic sensitivity. These final studies were based on the concept of exposure of breast tissue as a relevant measure for cancer incidence.

Thus, and with other studies that demonstrate the importance of breast density to identify breast cancer (5, 25, 28) arises the need for standardization of the same. Various classification methods of breast density have emerged over the years (29), of which the most widely used worldwide is the ACR, developed in 2003, which identifies four types of mammographic density (Figure 2.9) (24) .

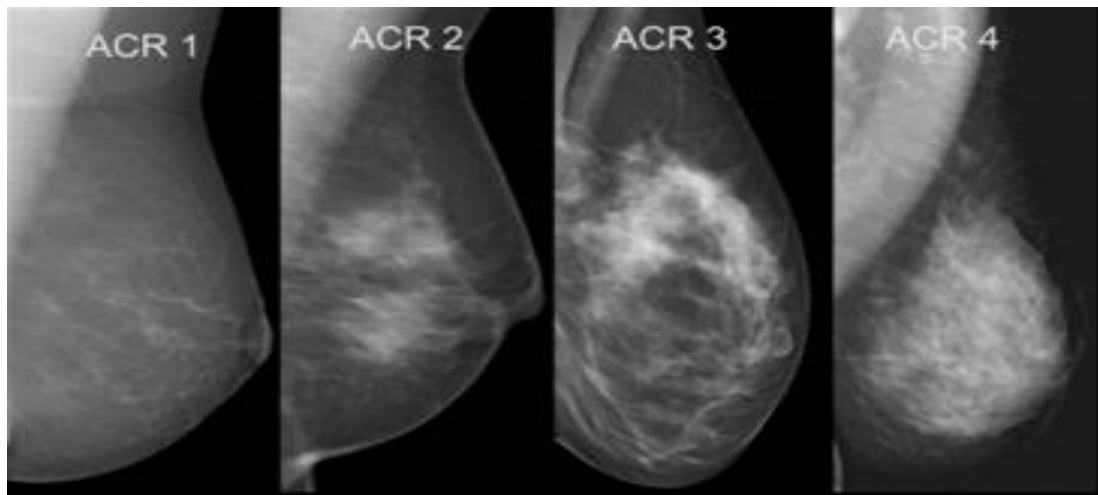


Figure 2.9: Breast density Scale of ACR (25).

The model of classification and standardization presented meets the following criteria:

- (1) Predominantly lipomatous (<25% glandular tissue);
- (2) Density fibrogranular dispersed (25-50%);
- (3) Heterogeneously dense breast (51-75%);
- (4) Extremely dense breast (> 75% glandular tissue) (24).

The main purpose of the creation of lexicons is to avoid confusion around mammogram reports which contain data and recommendations that are sometimes not clear for all physicians. In this sense, throughout this work we will use the terminology developed by ACR (BI-RADS[®]).

2.3 Breast Cancer types

Not all single breast masses are synonymous with cancer. These are distinguished according to frequency of appearance, as in the case of cysts and fibroids, which appear and disappear in a given period, or as fibro adenomas and intraductal papilloma, which are abnormal growths that may indicate a risk factor. The malignant masses clearly indicate that cancer is a carcinoma; a term used to describe a cancer that starts in the coating layer of organs (epithelial cells) such as the breast (16).

Among the identifiable mammography types that may indicate cancer and are assessed in examinations, are the masses, calcifications, bilateral breast asymmetry and architectural distortion (Figures 2.10 and 2.11).

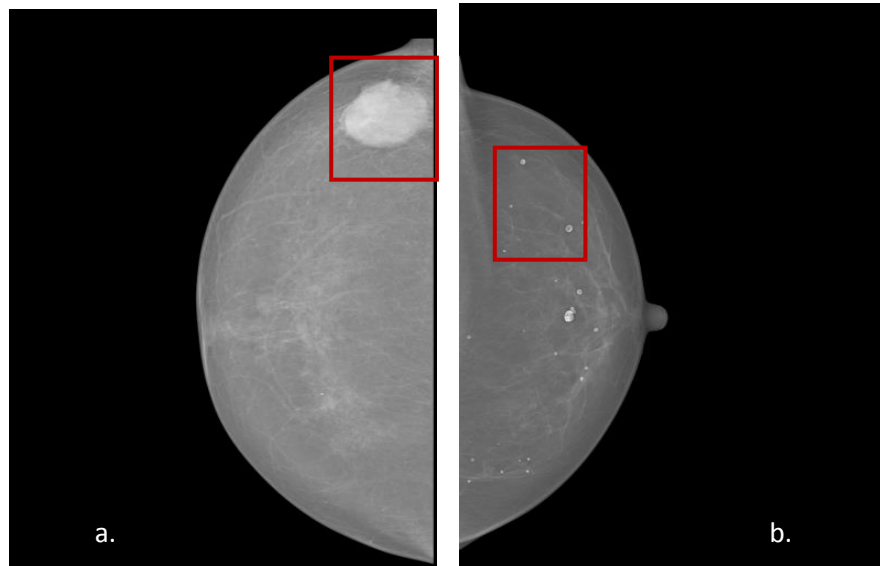


Figure 2.10: Example of a cranium-caudal (CC) image. Image from INbreast (60) and processed by the author.

On the figure 2.11 are presented some examples of a mass identification (a) and calcification (b), both marked on the image, respectively.

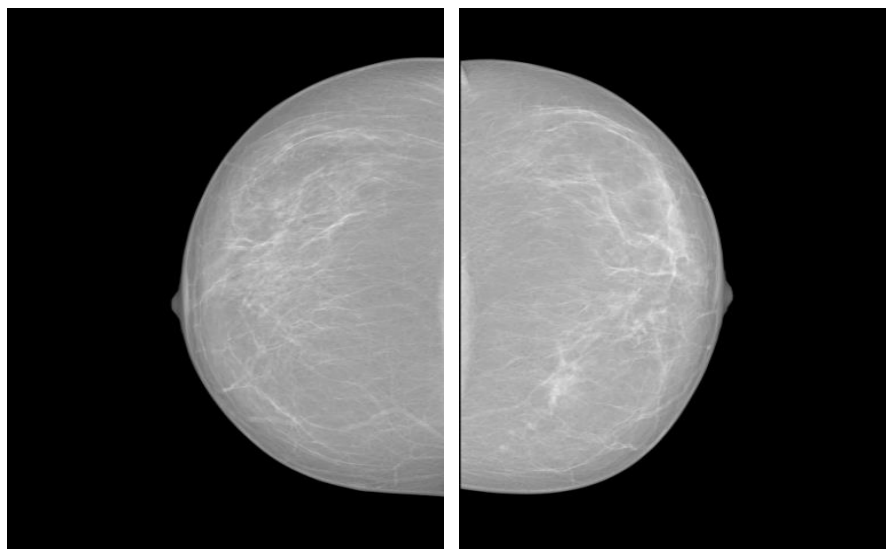


Figure 2.11: Mammogram example in oblique lateral median (MLO) with bilateral asymmetry. Image from INbreast (60) and processed by the author.

In breast tissue, a mass is an important change seen on a mammogram and can be cysts (non-cancerous fluid-filled sacs) and fibro adenomas (non-cancerous solid tumors) but they always require biopsy to identify whether or not they are malignant (Figure 2.10a). On the other hand, in the case of the presence of a calcification (Figure 2.10b), it consists of small deposits of minerals calcium, which can appear singly or in clusters (16).

If there are differences in the overall appearance of the breast images compared in reference to one another, we are in the presence of a case of bilateral asymmetry, as shown in Figure 2.11 (30). An architectural distortion is defined as a focal interruption of the normal mammographic pattern of lines, usually presenting as a star-shaped distortion, with no definite mass visible (31). In this project we will not include architectural distortion or bilateral asymmetry as there are cases with treatment and removal of tissue also presented.

2.4 Evaluation Metrics

In order to analyze the results, some metrics were applied to compare values of the area under the ROC curve, sensitivity, specificity and positive predictive value. These metrics are based on the evaluation criteria that are relevant in terms of the values of clinical interest.

The area under the ROC curve represents the way to select optimal models and to discard the suboptimal ones, independently of the cost or the class distribution context. Using ROC curves, researchers can trace the ideal profile and if the value of the area was close to one, greater efficiency is represented method.

Sensitivity and specificity are statistical measures of the performance. The first refers to the test's ability to identify a condition correctly and also can be called the true positive rate, or the recall rate. The second metric, specificity, represents the proportion of negatives which are correctly identified as such, and also known as the true negative rate.

The positive predicted value (PPV) was also analyzed, in order to better adapt the results with clinical practice. This result can tell us the probability that a patient has a relapse for the specific case analyzed, and represents the most interesting value for clinical practice recognition (32, 33). The PPV measures the precision of the test, revealing the quality measure in the positive instances classification.

Chapter 3

State of Art

Through this chapter we will present some bibliography related to classification models that were being applied on the aided diagnosis for cancer. In the last sub chapter is also presented some related work conducted at the level of the methods and techniques most promising in identifying lesions in x-ray images, always focusing on the field of mammography.

3.1 Classification Methods

The main advantage of data mining techniques is the ability to provide a set of useful rules capable of discriminating between a series of supposed risks (34).

Classification is a fundamental task in data mining techniques and relies on a process of differentiating two or more classes by labelling each similar set of data in a single class. Currently, there are several methods of classification for predicting breast cancer used in CAD systems such as Artificial Neural Network, K-nearest Neighbour, Support Vector Machine and Decision Trees.

The classifier construction has two different phases: the training and test groups. In the training phase, the training set is used to determine how different classes must be separated and how the features should be weighted and combined. On the other hand, the testing phase does not have to know the classes, so the weights determined on the previous phase are applied in order to identify the classes that this data set should belong to.

3.2.1 Artificial neural network

This classification model has the groundwork of a nervous system (particularly the brain). Artificial Neural Networks (ANN or NN) are generally presented as systems of interconnected "neurons" which can compute values from inputs.

Basically, a NN is an interconnected group of nodes, similar to the network of neurons in a brain. Here each circular node represents an artificial neuron and an arrow represents a connection from the output of one neuron to the input of another (Figure 3.1). On breast cancer detection, each node could include quality of the breast lesion, type lesion or even the type of breast tissue.

However, interest in traditional neural networks appears to have declined since the arrival of support vector machines, perhaps because the latter generally require fewer parameters to be tuned to achieve the same accuracy (35).

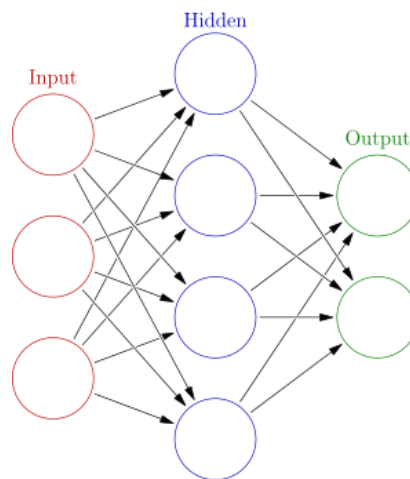


Figure 3.1 – Neural Network system (48).

In Mazurowski's work (12), related to creating an effective technique to reduce false positives in mammogram screening, six principal components were calculated for the extracted features using an artificial neural network and 10-fold cross-validation, and an average recognition rate of 77% was achieved. Using the receiver operating characteristic analysis, the overall sensitivity of the technique measured by the value of Area Under the ROC Curve (AUC) was found to be 0.74.

The CALMA project developed by Amendolia et al (36) and previously described, shows a sensitivity of 93% and specificity of 79% by using Feed-Forward Neural Networks in microcalcification classification.

3.2.2 K - nearest Neighbor

By adopting the K-nearest neighbor classification method we are able to suppress the problem of database corruption by noisy examples. Here some fixed and small k-numbers of nearest neighbour are located and used together to determine the class of the test instance through a simple majority vote (35). For the k-NN classifier, the membership value of a class is proportional to the number of neighbors belonging to this class and it is calculated by Euclidean distance metric (37).

Oliver et al. (37) by means of comparison of techniques for breast tissue classification, obtained 82% correct classifications by using this methodology ($k = 0.75$). A different work and recently developed by Bueno et al. (38), and also related to breast density classification, revealed 79% of agreement using hold-out and re-substitution, to test and train the classifiers with k-NN ($k=1$).

In 2012 a study made by Lesniak et al. (39) considered k-NN as the best classification method as far as mass detection in mammograms was concerned.

In the work done by Fonseca for her MSc Dissertation (40), and previously described in 3.1 of this chapter, the best results were acquired by using k-NN classifier.

3.2.3 Support Vector Machine

Support Vector Machine (SVM) is a method to estimate the function that classifies data into two classes. SVM uses a linear separating hyperplane to create a classifier with a maximal margin to discriminate between two classes, although when the classes cannot be linearly separated in the original input space, the classifier transforms it into a higher dimensional feature space where the classes might be separated (Figure 3.2) (41).

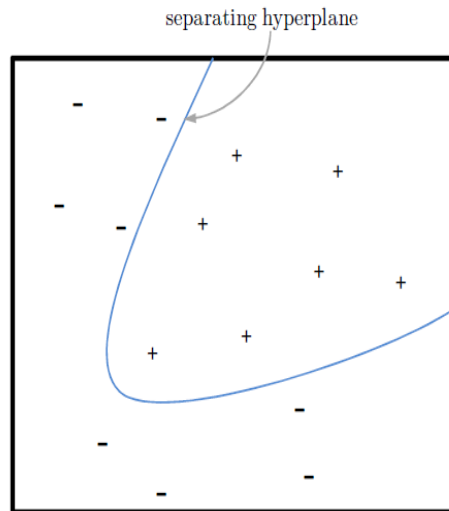


Figure 3.2 – Separating hyperplanes in SVM method (52).

Martins et al. (42) considered that among the classifiers available in the literature, in terms of lesion detection on mammograms, SVM comes off since it deals with sets of high dimensionality non-linear data with much accuracy and computational efficiency. Through his work, Martins (42) achieved a global accuracy for masses and micro of 93.11% using SVM.

By means of classifying breast tissue as dense or fatty, Bueno et al (38) acquired the best results by using SVM with 83% and 64% of accuracy for benign and malignant lesions, respectively.

Recently, Pollán et al. (43, 44) obtained 96.21% accuracy and 0.924 AUC with SVM, characterizing breast lesions according to BI-RADS classes grouped as benign or malignant.

On her work, Fonseca et al. (40) using the INbreast database, achieved 67.10% and 64.82% of accuracy using GLCM features for classifying dense and fatty tissue, respectively.

3.2.4 Decision Trees

A Decision Tree (DT) is perhaps the easiest method to understand and also the most widely used method that falls into the category of supervised learning (Figure 3.3) (35). In DT the test of a node, which involves testing a particular attribute, compares the attribute value with a constant. However, some trees compare two attributes with each

other, or use some function of one or more attributes. A classification was given by leaf nodes, and is applied to all instances that reach the leaf (35).

An unknown instance to be classified is routed down the tree according to the values of the attributes tested in successive nodes, and when a leaf is reached the instance is classified according to the class assigned to the leaf.

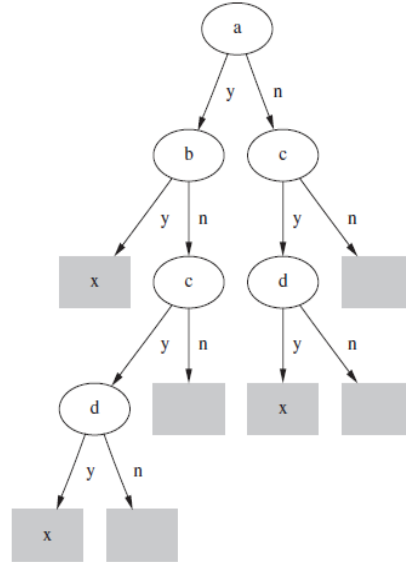


Figure 3.3 – A simple decision tree model (48).

There are some related works that employed a decision tree using C 4.5 as the decision algorithm (37, 38, 45).

3.2.5 Random Forest

Random Forest (RF) is a bootstrapping method applied to classify trees. This is an ensemble classifier consisting of many decision trees, where the final predicted class for a test example is obtained by combining the predictions of all individual trees, as described further in this work (chapter 4.6) (46).

This method can provide measures of the similarity between pairs of examples in the dataset and are often applied to high-dimensional data sets (47).

Some works have been done by using this method in the breast cancer field. Some of them are presented by Fonseca in 2013 (40), who revealed accuracies of 52.73% for dense and 51.60% for fatty tissue; and Melendez (48) who increased the mean sensitivity from 0.926 to 0.948 on detecting breast lesions by using RF.

3.2.6 Naïve Bayes

Naïve Bayes (NB) is the simplest form of Bayesian network, in which all attributes are considered independent given the value of the class – the reason why it is called “naïve” (35). It is obvious that the conditional independence assumption is rarely true in most real-world applications (49).

As Zhang explains on his work (49), many empirical comparisons between NB and DT such as C 4.5 showed that NB predicts equally well as C 4.5.

Some works have used this classification method in the breast cancer area, such as the work developed by Bueno et al. (38) and the one made by Castella et al.(50). By one hand, the first author just uses the method without any conclusive value of efficacy reported; on the other hand, Castela (50) proved that NB performance is excellent for low-dimensional features spaces, where the independent assumption can still be considered as valid. He obtained 86% accuracy by relating mammogram density with breast cancer risk.

With the bibliographical study here presented we can identify the most common methods used in the detection of lesions in mammography as well as the techniques for extracting the most promising features for diagnostic support. By this it was possible to identify the techniques to use as well as the potential classifier methods.

3.2 Related Work

Computer-aided diagnosis (CAD) in mammography is a topic that many research groups have already addressed. The first presentation of a system for computer-aided interpretation appeared in 1966, developed by the American Medical Association and represents a system of standardized terms for medical procedures used to facilitate documentation, called Current Procedural Terminology (CPT®). Today this is the most widely accepted medical nomenclature used to report medical procedures and services under public and private health insurance programs (51, 52).

Considering the history of computer science, this is a remarkable fact. So far it is generally believed that CAD systems can provide a valuable second look and improve the accuracy over optical methods.

One of the important events in the history of CAD is that, in 1998, the R2 Technology has succeeded in commercialization of the first CAD system for detection of breast lesions in mammography. This was based on licensing of CAD technologies from The University of Chicago, and that it obtained US Food and Drug Administration (FDA) approval for the clinical use of their system in 1998 (4, 11). This device was designed to search for signs that may be associated with breast cancer, marking masses and calcification clusters. The accuracy of detection was reported as 98.5% sensitivity for calcifications and 85.7% for masses, with 0.74 and 1.32 false-positives per case respectively (10). Since then, it was found that although the improvements on mammography are double read, there was room for improvement in cancer detection by implementing CAD.

Given these considerations, many qualitative researchers considered the idea of computer-aided diagnosis with data analyses for breast cancer detection. Since the 21st century, several studies have been made in order to create better methods of automatic detection in digitized mammograms (53) .

For research scientists there are several interesting topics in cancer detection and diagnosis, such as high-efficiency and high-accuracy lesion detection algorithms. On the other hand, for radiologists the importance of these methods relies on the effectiveness of clinical applications of the CAD system (51) .

Amendolia et al. (36) proposed and presented the Computer-aided Library for Mammography (CALMA). This is an automated search for mammogram's texture (essentially in masses and microcalcification lesions). CALMA's main purpose was to collect a database of mammographic images, developing CAD tools to be used as a second radiologist in the classification of breast cancer diseases. GPCalma, the grid platform from CALMA project, has been under development since 1999 and the sensitivity in the detection of lesions obtained for microcalcifications and masses was 96% and 80%, respectively (7).

Similar projects were developed, as the Magic-5 Project proposed by Bellotti et al. (54). This project offered algorithms for mammographic database analysis by means of grid services for breast cancer detection, with a sensitivity of 80% for massive lesions and 96% for microcalcification clusters.

More recently in 2011, Pollán et al. (43) created a software framework called BiomedTK. Through this project, the authors collected performance data of many trained classifiers with Artificial Neural Network (ANN) and Support Vector Machine

(SVM), using features on the region of interest and comparing them in CC and MLO views. Their results showed accuracy values of 97.14% and 96.91% and values of area under ROC between 0.933 and 0.924, for SVM and ANN respectively (43, 44).

Valiente et al. (55) in 2013, reported a novel software suite for processing, analyzing and diagnosing mammography images, entitled Mammography Image Workstation for Analysis and Diagnosis (MIWAD). This work is a result of the on-going collaboration between CETA-CIEMAT (Spain) and INEGI - Faculty of Engineering and the Hospital of São João - Faculty of Medicine, at the University of Porto (Portugal) in the research of breast cancer diagnosis support (55). MIWAD software offers a user-friendly desktop application which feeds the first Portuguese Breast Cancer Digital Repository (BCDR) with data reviewed by specialists (56) .

In a large measure, the BreastIQ framework from Biomediq (8) offers automatic, computer-based analysis of mammograms for analysis of future risk of breast cancer. The BreastIQ framework is founded on a segmentation framework that locates the breast region in MLO and CC views. Based on localizing the chest-breast interface and the nipple location, a breast coordinate system is placed on the mammogram. This commercial CAD automatically estimates the breast density from mammograms and offers computer support for radiologists scoring of density percentage in a smooth workflow.

The main goal of these projects, based on a grid infrastructure or software, is to be able to identify lesions on mammograms and compare them with a new one. The algorithms used on their CAD systems are constructed on SVM or ANN, and generally the features extracted to compare lesions are based on mass shape, mass margins, depth and size, among others. There are a few studies training texture analysis as quantitative measures, and those only use small measures such as correlation, contrast and entropy.

CAD systems have been highly developed, although the algorithms used are practically the same as well as the features extracted from the mammography. Texture features have been proven to be useful in differentiating masses and normal breast tissues and have also been referenced as having a promising future for CAD methods, being that individual experiments have been already done (45, 57-61).

Concerning extraction of texture features to detected lesion on mammograms, individual studies have been made (62, 63) and the majority of them apply grey level matrices using co-occurrence measures as the best approach. Some of the most important studies in this field are presented in Table 3.1.

Ferreira et al. (64) in 2011 create an online application (called Mamoclass) that allows anyone, with access to mammogram information, predict mass density and/or classify a mammography as benign or malignant according to a reduce set of input findings.

Nevertheless, in a recent MSc Dissertation by Fonseca, in 2013 (40), the same database presented in our work was used, and the main goal was to distinguish abnormal mammograms in accordance to breast density types. For detection she used GLCM and LBP (Local Binary Pattern) features and some classifiers to identify classes, as k-NN, SVM and Random Forests, applied on MATLAB[®]. The results showed that GLCM features have the best outcomes for dense breasts represented on the mammograms, and LBP obtained better results for all classifiers. It is also reported that k-NN classifier achieved better results in dense tissue and the SVM classifier obtained higher accuracy for fatty tissues. The major conclusion was that separate mammograms by density do not reveal major advantages to this database.

Table 3.1: Summary of work developed in the texture-feature analysis

Texture-feature analysis (authors of individual studies)	Study intent
• Karahaliou, 2006 (61)	→ To characterize microcalcifications on mammograms
• Manavalan, 2012 (59)	→ Applied to prostate cancer diagnosis
• Meselhy, 2012 (57)	→ To distinguish normal from abnormal tissue and the abnormality type on mammograms
• Zulpe, 2012 (60)	→ For brain tumor classification
• Chikamai, 2013 (58)	→ For classification of mammogram images
• Mohanty, 2013 (45)	→ Applied to evaluate masses on mammograms
• Nanni, 2013 (63)	→ Different approaches of extracting information
• Gebejes, 2013 (62)	→ Texture characterization
• Fonseca, 2013 (40)	→ Abnormal mammograms classification

Table 3.2, includes a summary of the more recent work done in the area of CAD systems development.

Table 3.2: Summary of work developed for breast cancer diagnostic

CAD System		Author
• CALMA project	→	Amendoilia et al, 2001 (36)
• Magic-5 project	→	Bellotti, 2004 (54)
• BiomedTK	→	Raul Póllan, 2010 (43, 44)
• Biomediq	→	Mads Nielsen, 2011(8)
• MIWAD	→	José Valiente, 2012 (55)

Chapter 4

Methodology

In order to be able to extract, select and classify textural features, to extract information about the breast lesion type, their malignancy and the breast density, we here present the methods and techniques applied to achieve that.

The methodology was based in seven major steps: Initial Data Comprehension and Preprocessing, Features Extraction and Selection, Data Set Organization, Classification Methods and finally the Evaluation Criteria. The block diagram with the approach used is presented in Figure 4.1.

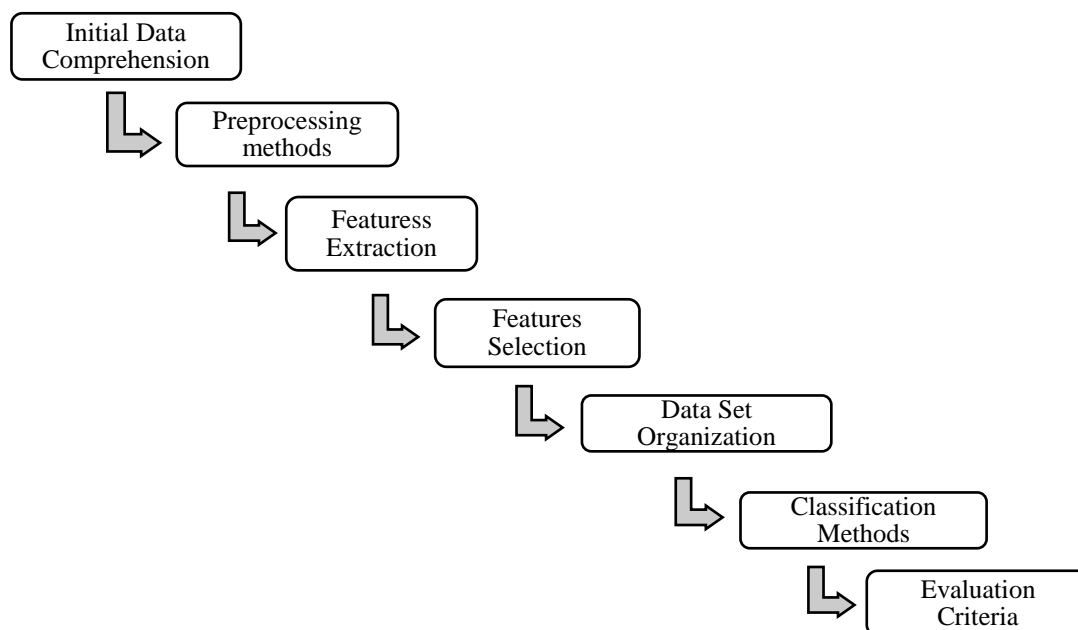


Figure 4.1: General block diagram of the proposed methodology

In order to perform the steps proposed in the course of this work, the software used was MATLAB[®], SPSS[®] and WEKA[®] which only the last one is an open source license.

4.1. Initial Data Comprehension

The images used in this work were taken from the repository INbreast (31), developed by multiple institutions of the University of Porto and available to the public with authors' consent. The INbreast includes a total of 410 images (115 patients) from which 90 patients are women whose both breasts are affected (four images per case) and 25 of the cases are mastectomy patients (two images per case). The sample also includes some types of findings (masses, calcifications, distortions and asymmetries) which were annotated by specialists.

Medical reports associated with each case using BI-RADS classification and the levels of breast density for each image are also included. Due to the small number of breast distortions present in the database, this kind of lesion was not considered in the present study.

The images are in DICOM format, and specify which breast and projection stands, for example, the case of an R_MLO image which indicates the right breast and a median lateral oblique projection, or an L_CC, representing the left breast and a cranium caudal projection.

In Figures 4.2 and 4.3, some graphics alluding to the initial data are revealed, and some initial relations between the radiologist's manual classifications are presented.

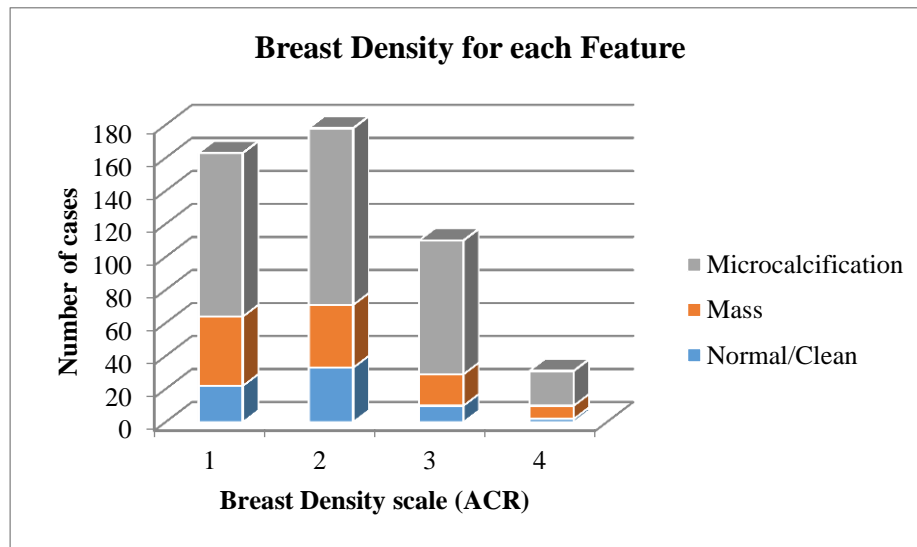


Figure 4.2: Initial data graphic with the percentage of findings reported for each breast density class.

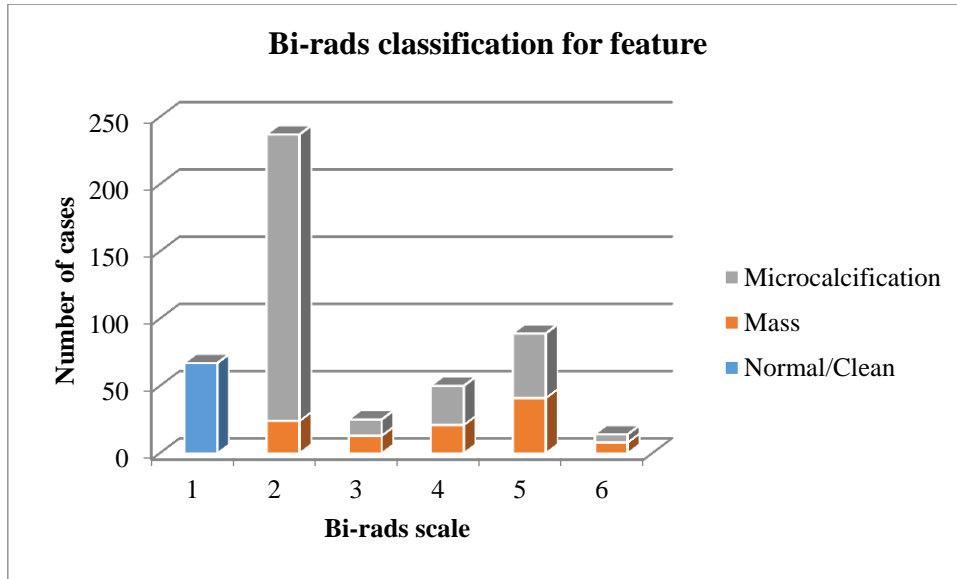


Figure 4.3: Initial data analysis graphic showing the percentage of findings reported for each BIRADS class.

Through the quick exploration of the graphs above it is possible to observe that the number of cases for each type of finding is not equal as well as the number of images with microcalcification and mass discrepancy. It is also important to note that among the classes of each group studied (ACR and BI-RADS) the levels are not linear with cases that have few occurrences for a particular lesion.

4.2 Initial Data Preprocessing

Mammograms are medical images frequently difficult to interpret because the pre-processing is a necessary step to improve image quality and to ensure greater accuracy of result (23, 45, 65).

Therefore, for a complete and effective analysis of the images' characteristics, at first it was necessary to improve them by using Matlab®, because to process an image in "DICOM" format requires more time and becomes more exhaustive. Due to the total number of images, they were converted to Portable Network Graphics (png) format and inserted into Matlab® (66).

Using the described software, and existing functions (fspecial, imtophat, imadjust) the images were smoothed by applying a Gaussian filter to eliminate noise, and the discrete wavelet transform was then implemented, both with proven efficacy (45, 63, 67). Wavelet transform is a sparse and efficient way to represent an image by improving

the image quality multiresolution representation. The Wavelet transform is similar to filters that use a high pass filter and a low pass filter to decompose an image into sub frequency bands along the image's rows and columns (57, 60) .

After the previously described steps, the initial image with 5000 x 5000 pixels size, was converted into a 4-bit image (128 x 128 pixels corresponding to 16 gray levels) with the purpose of increasing the effectiveness of the feature extraction process and reduce computational time and cost, since this process relies on the analysis of the image gray levels (23).

Each image was individually processed, so that the selected region of interest (ROI) with the lesion turned out to be as reliable as possible to the associated medical report and the data is incrementally directed to Microsoft Excel® until all the images are processed (Figure 4.4). In case of normal images, without ROI indication, this selection was not taken in consideration.

In this way, a total of 410 images were processed, and in some cases several lesions were observed on a single mammogram, giving a total of 439 images and respective characteristics, with respect to the multiple lesions found in a single image.

The functions applied to the image quantified to 4 bits and to computing the gray-level run length matrix were obtained from a program developed by Karunanithi Rajamanickam (68) and is available to the public. On the other hand, the gray-level co-occurrence matrix was obtained from Matlab® functions (graycomatrix and graycoprops).

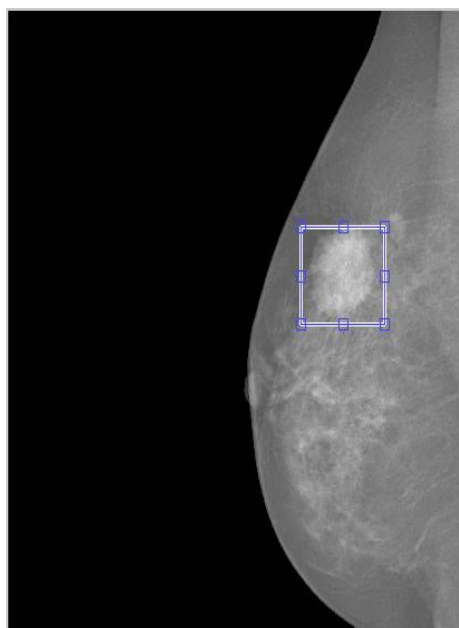


Figure 4.4: Matlab selection of region of interest on a mammogram.

4.3 Features Extraction

Feature selection is a key point that should be taken into account in the implementation of a decision support system, in recognizing breast tissue by selecting the most important features, and also due to the ability to describe and maximize differences in tissues and/or lesions (57) .

Texture is an important image characteristic that has been widely used in medical image analysis especially in their automatic classification (45, 69).

According to Haralick (70) texture is an innate property of virtual surfaces and contains relevant information on their structural arrangement and their relationship with the surrounding environment. Haralick highlighted texture independence relative to the image tones, as far as different shades may have the same texture and vice versa.

Texture features can be extracted by various methods using models based on statistical information or structure models based on image processing. It has been proven by several authors that texture-based features are useful in the differentiation of lesions and normal tissue on mammograms (45, 65, 71-74).

In this work, two of these feature-based matrix were used: GLCM (Gray-level co-occurrence matrix), applied before ROI selection and GLRLM (Gray-level run length matrix) applied after ROI selection and both later described in detail.

4.3.1 GLCM

The GLCM is a statistical method of texture examination that considers the spatial relationship between the image pixels. This matrix is created by the observation of image cell pairs at a distance, d , from each other and with the increase in the position corresponding to the gray level of both cells (Figure 4.5) (45, 63, 75).

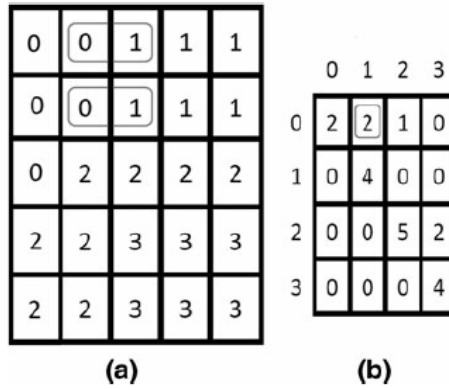


Figure 4.5: GLCM construction, a) Example of a 4-gray level image; b) GLCM distance 1 and direction 0° (44).

For this work, in order to obtain the GLCM matrices, the default values of spatial relationship horizontal adjacent were maintained and performed with Matlab® functions, being a distance of 1 and direction of 0°.

The *graycomatrix* function creates the GLCM by calculating how often a pixel with a certain grey value i occurs adjacently to pixels with a value j in each one of previously mentioned directions. In the resultant matrix, each element (i, j) is the sum of the number of times that the value i has occurred in the spatial direction with the other pixel value j , as showed in Figure 4.5.

In the GLCM the number of rows and columns equals to the number of grey tones in the image (G). The element $P(i, j | \Delta x, \Delta y)$ is the relative frequency between pixels separated by distance $(\Delta x, \Delta y)$.

The matrix element is also represented by $P(i, j | d, \theta)$, which contains the second-order probability of change between grey levels in a range of i and j at a distance d and a θ angle.

$$p_x(i) = \sum_{j=0}^{G-1} P(i, j) \text{ e } P_y(j) = \sum_{i=0}^{G-1} P(i, j) \quad (1)$$

As a result, Equation 1 shows how the array is constructed, in which the value of $p_x(i)$ is the input value i in the matrix, obtained by the rows of $P(i, j)$ summation (60)

Another used function was *graycoprops* which allowed us to extract the features originally calculated by Haralick (70) which were described in the present work, as follows:

- **Contrast**

The measure of intensity between a pixel and its border pixel. The contrast is higher in strong and colorful images (approximately 1), than in white or very constant images, which have lower value (approximately 0).

- **Correlation**

The measure of linear dependency between gray levels. In horizontal readings of 0 degrees, the correlation values will be the highest. In a clear image, the gray values are more constant and less correlated. These values range from $[-1, 1]$, with 1 being a perfectly positive correlation.

- **Energy**

Also referred as angular second momentum (ASM) or uniformity, this feature refers to the proximity of the distribution of elements in the array (65). There are uniform images where few transitions of gray levels exist. Thus, in a clearer image and with small gray value, this feature will be higher; on the other hand, in a sanded image with many gray values to oscillate, the value will be smaller. The value of 1 in this field represents a steady state.

- **Homogeneity**

Calculated from the IDM (Inverse Difference Moment), this feature measures the distribution of gray levels in a diagonal image. If the values decrease along the diagonal, the result will be higher, and the image will reveal little contrast. In return, lower values show very intense images, somewhat homogeneous and therefore with greater contrast. A value of 1 is equivalent to a completely homogeneous image on its diagonal (66).

Each feature is taken from the entire image and the equations are described in the Table 4.1.

Table 4.1: Haralick features equations (66)

Feature	Equation
Contrast	$\sum_{n=0}^{G-1} n^2 \left\{ \sum_{i=1}^G \sum_{j=1}^G P(i,j) \right\}, \quad i-j = n$
Correlation	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{\{i \times j\} \times P(i,j) - \{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y}$
Energy	$ASM = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{P(i,j)\}^2$
Homogeneity	$IDM = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1 + (i-j)^2} P(i,j)$

4.3.2 GLRLM

Another set of features based on the texture are GLRLM which are based on computerizing the number of lines of grey levels at various angles (57, 59, 63). Chu (76) and Dasarathy (77) introduced different features representatives of the lengths of arrays of gray lines.

A grey line is a set of collinear and consecutive points with the same pixel grayscale value for a specific linear orientation (0 degrees). For a given image, the run length of the line $P(i, j)$ is defined as a number of lines with pixels of grey-level value i and line length value j (57, 59, 63)

Being n as the total number of lines in the image, M as the number of grey levels and N the maximum line length, the authors (76) and (77) defined the following characteristics (Table 4.2).

Table 4.2: Characteristics extracted from GLRLM

Feature	Expected values	Equation
Short run emphasis (SRE)	Higher values for soft textures.	$\frac{1}{n} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i, j)}{j^2}$
Long run emphasis (LRE)	Higher values for higher and denser textures.	$\frac{1}{n} \sum_{i=1}^M \sum_{j=1}^N j^2 p(i, j)$
Gray level non- uniformity (GLN)	Lower values for homogeneous shades	$\frac{1}{n} \sum_{i=1}^M \left(\sum_{j=1}^N p(i, j) \right)^2$
Run length non- uniformity (RLN)	Lower values for lines with homogeneous sizes	$\frac{1}{n} \sum_{j=1}^N \left(\sum_{i=1}^M p(i, j) \right)^2$
Run percentage (RP)	Value nearest to 1 indicates short lines	$\sum_{i,j} \frac{n}{p(i, j)j}$
Low Gray Level Run Emphasis (LGRE)	High values show ROI with lower gray values	$\frac{1}{n} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i, j)}{i^2}$
High Gray Level Run Emphasis (HGRE)	High values show ROI with higher gray values	$\frac{1}{n} \sum_{i=1}^M \sum_{j=1}^N i^2 p(i, j)$

The values of each feature mentioned above are extracted twice, the first for a ROI with a lesion finding, SRE_L, for example, and the second for a clean/normal ROI, such as SRE_N, at the same image. It is important to notice that, for normal mammograms, the ROI selection was also made, in order to extract texture features for further comparison.

The total number of eleven features, referring to both matrices, was saved in a file for later feature selection.

4.4 Features Selection

With the purpose of selecting the most relevant characteristics for the study, after collecting the eleven values presented above, they were inserted in statistical analysis software by IBM ®, SPSS version 20® for Microsoft Windows®. With the support of

the tools available a statistical analysis of what values most significantly distinguish masses of microcalcifications was performed.

The dataset introduced corresponding to each feature (GLRLM features) refers both to ROI and to the area without lesion on the image, being called the lesion and the normal, respectively. In SPSS® the ratio of these two values (lesion/normal as represented in table 4.3) was calculated in order to reveal the relative value of the difference caused by the injury on the tissue.

On this dataset a new section was also added containing the found characteristic for each case described. Using this field we incremented the value 1, 2 or 3 representing masses, microcalcifications and normal/clean findings, respectively.

After the normal distribution was confirmed, in the majority of cases, we performed an Independent-Sample t-test that allowed us to test whether two samples mean values were statistically different from each other (mass and microcalcification). Through the test output we evaluated the sample sizes (N), means, standard deviations, and the standard error of the mean (the standard deviation divided by the square root of N) for the continuous variable, separate for each group (masses and microcalcifications).

For a significance level of 5%, if p-value is greater than 0.05, then we can assume that group variances are equal and if the p-value is 0.05 or less, then we should assume that the group variances are not equal (78). Our case study revealed that the majority of variances are equal, showing a significant level (2-tailed) approximately to 0.001, which reveals that in addition to having significant correlation ($p < 0.05$) they also revealed significant interest ($p \approx 0.00$).

The complete list of SPSS outputs with 95% confidence intervals, are present on the Table 4.3, with reference to each p-value. Though SPSS refers to a significance level of 0.000, it is generally inappropriate to report it as such, and reporting 0.001 is the preferred method (78).

Table 4.3: Significance level for each characteristic, obtained in SPSS t-test outputs

Characteristic		p-value (t-test)
SRE	Lesion	0,001
	Lesion/Normal ratio	0,001
LRE	Lesion	0,001
	Lesion/Normal ratio	0,001
GLN	Lesion	0,733
	Lesion/Normal ratio	0,397
RP	Lesion	0,001
	Lesion/Normal ratio	0,001
RLN	Lesion	0,001
	Lesion/Normal ratio	0,001
LGRE	Lesion	0,001
	Lesion/Normal ratio	0,001
HGRE	Lesion	0,001
	Lesion/Normal ratio	0,001

4.5 Data Set Organization

In a following step, still using the SPSS[®] software, we added two more classes representing BI-RADS classification. At this point, and confirming all the process with medical reports' analysis, for each case presented at SPSS[®] file, we incremented the respective value of BI-RADS classification and the value of breast density (BrD), both in accordance with chapter II.

In order to create separate files that could be used as an input for another program, the SPSS[®] output was organized into three initial records. This procedure is extremely important since by doing this we were able to evaluate the influence of different classifications at the same dataset.

The first file has the BI-RADS classification with GLRLM outputs for each case described, and was named BI-RADS file. The second comprehends the breast density classification made by radiologists and the GLCM matrix, due to the fact that it was the only matrix related to the complete image analysis (without ROI selection) that we called BrD file. Finally, the last one was created in order to relate the finding type (mass, microcalcification or normal) with GLRLM, for every image finding described

(Characteristic file). By doing this we were able to create three major scenarios that will be the input files on WEKA[®] (Bi-rads.arff; BrD.arff and Characteristic.arff).

4.6 Classification Methods

As mentioned in Chapter 3.2, classification is a fundamental task in data mining techniques and is one of the most important steps for constructing a system for computer aided detection.

In this work, for each of the three different files previously constructed in Chapter 4.5, several classification methods were used, such as: k-nearest Neighbor Support Vector Machine, Decision Tree (J48), Random Forest and Naive Bayes. The application of these different scenarios was prepared in order to better understand which the most efficient classifier for each case was. To apply them we use WEKA[®] software version 3.6.11, Explorer module, developed from Waikato University, and the cross-validation *k-folds* test option with ten-folds. The main principles related to this validation method are based on a data subdivision into *k* identical subsets (10 for this specific case), that are trained for each subset *k-1* (training) and the prediction is verified at the correspondent *k* subset (test).

4.6.1 Random Forest

Random Forest (RF) was one of the chosen classifiers due to its suboptimal performance in cases of strongly unbalanced data that is, being almost an optimal level when class sizes differ considerably, as shown in recent studies (79, 80).

In the RF algorithm several individual decision trees are combined to make a final prediction. Each tree represents a random sample of observations from the original sample. The ones that are not used to construct the tree are called as out-of-bag (OOB) observations (79). By aggregating the predictions of the OOB data across all trees it is possible to determine an internal estimate of the general error of the method (47). For each node or split in a tree, features are randomly selected from the entire dataset. The selection of the best node in a tree is acquired by a conditional inference test in an unbiased way. For each node, each candidate predictor from the random subset is globally tested for its association with the response, yielding a global p-value. The selected one has the smallest p-value and the best node is finally chosen (Figure 4.6).

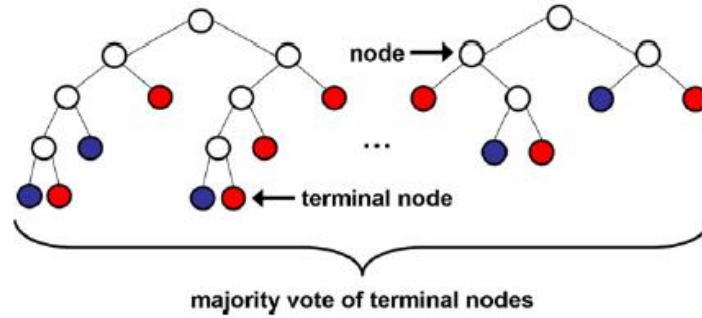


Figure 4.6 – Illustration of Random Forest nodes (56).

For the present work, the arguments chosen to be altered from the default values were the number of attributes used in random selection (K) and the number of trees (I) to be generated. For each case studied, those parameters can be different, due to the variability of the class.

4.6.2 SVM

As explained in Chapter 3.2.3, SVM classifier is based on the idea of minimizing the generalization error when the classifier is applied to test samples that do not exactly match any training sample used to train (42).

In this case, SVM was used because it provides a good generalization, contributing to the effectiveness of the methodology during the classification of segmented structures. In WEKA[®] a library was available for SVM, which was LibSVM, and it was used for training and testing each group file described in Chapter 4.5. The parameters altered on SVM default settings were: the seed to be used, the degree of the kernel and the kernel type. This last parameter specifies the type of the kernel function used by the function to map the training data into the kernel space and it can be: linear, quadratic, polynomial, Gaussian radial basis function or sigmoid. On this work each one was tried, in order to find which performed the best results.

The SMO function that implements a sequential optimization algorithm for training SVM was used, in order that with this function we could specify the method used to separate the hyperplane. In SMO the kernel chosen was the polykernel with default values. Throughout this task SVM or SVM-SMO was used, depending on which of them revealed a better result, for each case studied.

4.6.3 k-nearest Neighbor

This classifier was chosen because it is a very simple algorithm to understand and implement. Due to this, it should always be considered in classification tasks, despite the fact that it can take longer to make a prediction since it requires a large amount of memory. In spite of these disadvantages, this was the first classifying method implemented in this study being compared to the others afterwards.

This method relies on two basic steps: first, to find the k instances in the training set that are similar to the instances on the testing set, according to some metric, and afterwards there is a majority vote of the k nearest instances to the test instance in order to find the test class of that test instance.

K-NN that can be found on WEKA® as IBk on lazy classifier group was used, and the algorithm chosen for the nearest neighbour search was linear, using the Euclidean distance formula (35) to calculate the distances between the train and test instances. The k value was tested from 1 to 10, but above $k=4$ the TN rate started to get lower.

4.6.4 Decision Tree

As mentioned in Section 3.2.4, a DT has the advantage of visualizing the data classification more easily. In the same section some studies using C4.5 DT as the classifier were also presented. Both C4.5 and the improved versions use information gained as a measure of purity, which is based on the notion of entropy (45).

For this reason, the DT chosen for the present work relies on C4.5 model. However, this model has been eliminated from WEKA®, where J48 became the substitute method as an open source Java® implementation of the C4.5 decision tree algorithm. All parameters used were the default ones.

4.6.5 Naïve Bayes

This method was mainly used for comparison with the other classifiers, given that there were few studies in the area that justify its use. Nevertheless, as we previously mentioned NB guarantees an excellent performance for low-dimensional features spaces.

Even though, this method has fewer parameters to be chosen, we preferred to keep them on the default values, once it was decided only to compare it with the other classifier's results.

4.6.6 Ensemble Methods

Ensemble methods in machine learning aim to induce a collection of diverse predictors which are both accurate and complementary, so that, better prediction accuracy on previously unseen data is obtained when the decisions of different learners are combined (81).

For this purpose three ensemble methods were built using DT's as a base classifier. The chosen DT's were already used in individual tests, in order to be possible to analyze the ensemble methods' efficiency.

Three different ensemble classifiers were used:

- **Bagging**

Also known as bootstrap aggregating, it is usually applied to decision trees classifiers, and the improvement of generalization errors is made by reducing the variance of the base classifier (81). Bagging applies the learning scheme to each one of artificially derived dataset, and the classifiers generated from them vote for the class to be predicted (35).

- **AdaBoost**

Adaptive Boosting is a technique that, to simplify matters, assumes that the learning algorithm can handle weighted instances, where the weight of an instance is a positive number. The presence of instance weights changes the way in which a classifier's error is calculated. This error is the sum of the weights of the misclassified instances divided by the total weight of all, instead of the fraction of misclassifiers. Throughout this task, the algorithm can be forced to concentrate on a particular set of instances, those with high weight (35).

- **MultiBoosting**

This is a technique that is an extension to the highly successful Adaptive Boosting for forming decision committees. In general terms it combines Bagging and AdaBoost methods (81).

More detailed information about these methods can be found at (35) .

Chapter 5

Results and Discussion

In this section the results of our work are presented including both data provided by SPSS[®] and WEKA[®] software's, as well as the different scenarios that were considered. We will also analyze the results described by making comparisons between all the methods applied and taking in consideration all the scenarios tested.

In order to distinguish the best performance obtained in each scenario, we first analyzed each of them separately and then compared the global results, in order to find both the best predictor and the best method for breast cancer diagnosis.

5.1. SPSS[®] outcome

As presented in Table 4.3 (Chapter 4.4), 0.001 was the value considered as a significant interest level, allowing us to recognize that all correlations are statistically significant, in order that the increase or decrease of one variable will affect the others. This fact is extremely important for this work since we aimed to find dependencies related to each lesion and normal tissue for each texture characteristic extracted from the two matrices.

At this point, any finding characterized as normal or clean was dismissed since the matrix obtained was obviously different because they did not have a lesion report field.

Besides that, some assumptions can be made, such as the identification of the most relevant variables, and the removal of the GLN variable due to the fact that it is the only one that does not have a significant correlation with the lesion discrimination ($p > 0.05$).

5.2. WEKA[®] outcome

In order to better analyze the performance of the tested methods applied to different classification standards, WEKA[®] results were divided into three major scenarios, related to the classification attributed to each one.

- **Scenario 1**

In the first scenario we related the BI-RADS[®] classification with features extracted from GLRLM. At first, and considering all five classification groups mentioned in Chapter 2, the results showed that some of the groups were not correctly classified. To control this fact, the classification was made considering two major groups. The data was rearranged into two subclasses (benign and malignant) according to the risk reported by each on the BI-RADS[®] lexicon. For this purpose, the benign group has the 1, 2 and 3 categories (with a total of 239 cases) and the malignant group includes the 4, 5 and 6 categories from BI-RADS[®] (with 258 cases considered).

In Table 5.1 the results from each classifying method used for this dataset are presented, showing their PPV, AUC, sensitivity, specificity and global accuracy.

Table 5.1: Results for GLRLM with BI-RADS[®] subclass classification.

Classifier	BI-RADS [®] subclass	PPV	AUC	Sensitivity	Specificity	Global Accuracy
Naïve Bayes	Benign	0.205	0.486	0.538	0.532	53.3 %
	Malignant	0.837	0.486	0.532	0.538	
SVM-SMO	Benign	0.050	0.508	0.571	0.523	52.5 %
	Malignant	0.965	0.508	0.523	0.571	
k-NN	Benign	0.657	0.789	0.724	0.707	71.4 %
	Malignant	0.767	0.789	0.707	0.723	
J48	Benign	0.427	0.604	0.634	0.592	60.6 %
	Malignant	0.771	0.604	0.592	0.633	
Random Forest	Benign	0.757	0.831	0.757	0.775	76.7 %
	Malignant	0.775	0.831	0.775	0.757	

Using the classifiers mentioned above, some parameters were changed in order to find the best performance for each case. On the one hand, the first two methods, as well as the J48, were used without any change to their default values. On the other hand, to use the k-NN we choose a k value of 1 and for Random Forest classifier, the k and I

parameters were changed into 2 and 15, respectively. Besides that, the cross-validation option was changed and used with fifteen-folds instead of the ten-folds that were used for all the other cases.

To ensure the best results ensemble methods were also tested on the same dataset. The methods used with the ensemble classifier (Random Forest) have the same parameters previously stipulated (k and I). The results are showed in Table 5.2.

Table 5.2: Results of Ensemble Methods to GLCM with BrD.

Method		PPV	AUC	Sensitivity	Specificity	Global Accuracy
Bagging						
J48	Benign	0.665	0.791	0.697	0.703	70.0 %
	Malignant	0.733	0.791	0.703	0.697	
Random Forest	Benign	0.724	0.809	0.709	0.739	72.4 %
	Malignant	0.725	0.81	0.739	0.709	
AdaBoost						
J48	Benign	0.565	0.693	0.652	0.641	64.6 %
	Malignant	0.721	0.693	0.641	0.652	
Random Forest	Benign	0.728	0.798	0.719	0.745	73.3 %
	Malignant	0.736	0.798	0.745	0.719	
MultiBoosting						
J48	Benign	0.582	0.730	0.685	0.660	67.0 %
	Malignant	0.752	0.730	0.660	0.685	
Random Forest	Benign	0.724	0.814	0.712	0.740	72.6 %
	Malignant	0.729	0.814	0.740	0.712	

On the subject of predicting if a mammographic finding was benign or malignant and considering the global percentage of correctly classified instances, the best results were achieved by Random Forest classifier, with 76.7% accuracy, followed by k-NN with 71.4% accuracy. These two classifiers also presented the highest values for sensitivity and specificity.

In spite of this, and regarding PPV, which is the value with the most interest in clinical practice recognition, for malignant lesions identification, the SVM-SMO and Naïve Bayes classifiers showed the higher values, 96.5% and 83.7%, respectively compared with 77.5% for Random Forest. However, when we took into consideration

the mean value of PPV, Random Forest had the best result having less false-positive for both classes.

Moreover, when comparing the results of both decision trees (J48 and Random Forest) with their results for ensemble methods, we were able to determine that these methods showed improvements on the J48 classifier with an increase of 9.0% global accuracy for bagging method however, this was not observed for Random Forest classifier. Even with this factor Random Forest was still the best predictor.

The results of this scenario also show that the Naïve Bayes and SVM-SMO revealed the worse percentages in terms of mean values for all measures.

On the figures (Figure 6.1 and Figure 6.2) it is possible to observe and compare the best results in relation to other methods for benign and malignant classes separately.

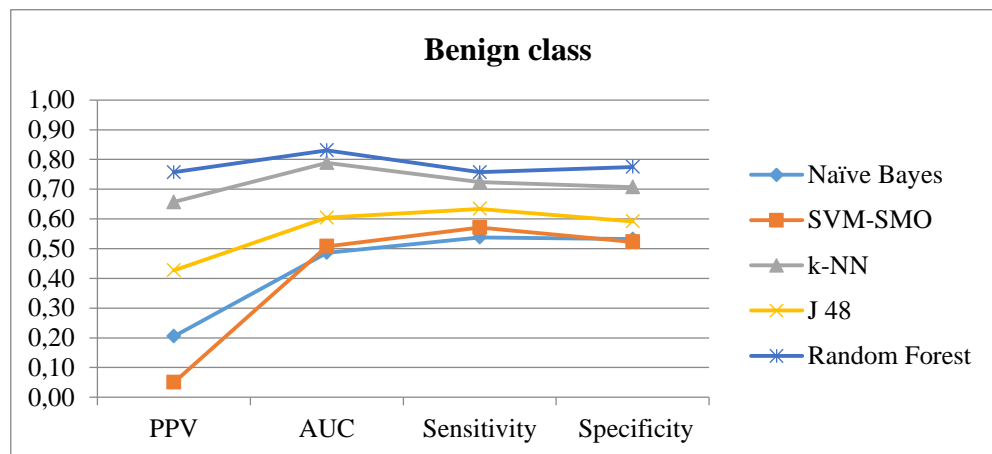


Figure 6.1: Graphic of results extracted from scenario 1 for benign class.

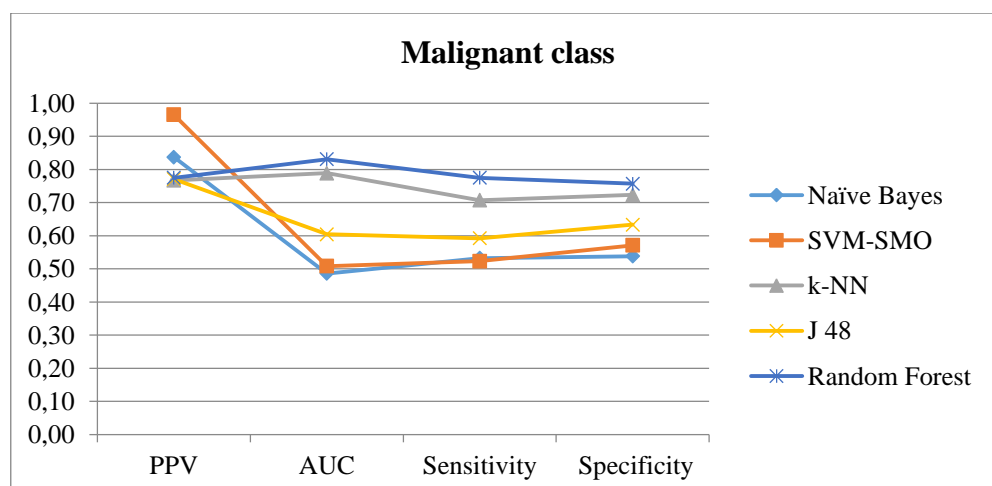


Figure 6.2: Graphic of results extracted from scenario 1 for malignant class.

For this scenario, grouping the classes became an effective technique and the use of GLRLM, even though only a few works use them, and this was a good model for this class' distinction.

While conducting our study we concluded that the results were good for the chosen parameters to each classifier, but the best predictor method was the Random Forest. With this classifier we were able to achieve values around 76.0% PPV, both for benign and malignant findings, which led us to conclude that this method has a good success rate as well as AUC (0.831), sensitivity and specificity (approximately 0.700 each). We also realized that ensemble methods at this scenario did not represent any improvement.

- **Scenario 2**

The second scenario was a multiclass problem that was tested considering the GLCM features' extraction related to the four levels of BrD, in accordance to ACR. On a previous step the complete set of features was tested and analyzed however, the BrD values only showed positive results when related to the GLCM, as discussed later.

The complete list of results from this scenario is presented in Table 5.3. For each of the classification methods, PPV, AUC, sensitivity and specificity values for each class are presented, as well as the classifier global accuracy (Table 5.3).

Table 5.3: Results of GLCM with BrD.

Classifier	BrD by ACR	PPV	AUC	Sensitivity	Specificity	Global Accuracy
Naïve Bayes	1	0.592	0.725	0.534	0.791	34.5 %
	2	0.206	0.480	0.276	0.615	
	3	0.165	0.596	0.250	0.753	
	4	0.538	0.750	0.182	0.959	
SVM-SMO	1	0.633	0.782	0.679	0.828	49.7 %
	2	0.817	0.594	0.418	0.786	
	3	0.000	0.666	0.000	0.753	
	4	0.000	0.500	0.000	0.929	
k-NN	1	0.792	0.820	0.638	0.886	54.9 %
	2	0.557	0.647	0.514	0.743	
	3	0.363	0.712	0.500	0.808	
	4	0.038	0.651	0.091	0.930	
J48	1	0.625	0.784	0.658	0.823	51.6 %
	2	0.557	0.655	0.507	0.741	
	3	0.429	0.722	0.429	0.812	
	4	0.115	0.612	0.158	0.934	
Random Forest	1	0.725	0.829	0.690	0.864	58.7 %
	2	0.603	0.715	0.556	0.770	
	3	0.516	0.785	0.553	0.844	
	4	0.115	0.704	0.200	0.935	

Despite the fact that the majority of the methods used to classify this scenario have the default parameters established as the best ones, for the k-NN case, we have changed the k value to 3 in order to improve the results obtained.

In Table 5.4 the results from the ensemble methods tested for this scenario are presented.

Table 5.4: Results of GLCM with BrD.

Method		PPV	AUC	Sensitivity	Specificity	Global Accuracy
Bagging						
J48	1	0.717	0.830	0.623	0.857	53.5 %
	2	0.527	0.663	0.519	0.712	
	3	0.429	0.748	0.470	0.819	
	4	0.115	0.720	0.214	0.936	
Random Forest	1	0.717	0.838	0.637	0.854	55.7 %
	2	0.513	0.696	0.543	0.756	
	3	0.462	0.762	0.488	0.826	
	4	0.077	0.686	0.222	0.933	
AdaBoost						
J48	1	0.633	0.799	0.613	0.820	52.9 %
	2	0.519	0.674	0.511	0.732	
	3	0.538	0.736	0.516	0.846	
	4	0.077	0.579	0.125	0.932	
Random Forest	1	0.642	0.812	0.636	0.826	54.3 %
	2	0.588	0.684	0.524	0.756	
	3	0.473	0.735	0.231	0.829	
	4	0.115	0.618	0.532	0.935	
MultiBoosting						
J48	1	0.650	0.788	0.650	0.831	55.4 %
	2	0.603	0.717	0.534	0.764	
	3	0.484	0.763	0.524	0.834	
	4	0.115	0.637	0.188	0.935	
Random Forest	1	0.692	0.818	0.659	0.847	54.6 %
	2	0.55	0.702	0.503	0.738	
	3	0.505	0.725	0.511	0.838	
	4	0.000	0.740	0.000	0.928	

The results obtained for GLCM features related to BrD show that for each BrD class there was a completely different value of prediction, and because of that the global accuracy values are not as promising as we were expecting. Despite this, we were able to verify that the SVM-SMO method did not consider the last two classes (3 and 4), as shown by PPV value. By the global accuracy percentage, we also distinguished Naïve

Bayes and SVM-SMO as the worst classifiers (34.5% and 49.7%, respectively) and k-NN and Random Forest as the best ones (54.9% and 58.7%).

The best AUC represented were between 0.704 and 0.829 and were achieved with Random Forest that has also shown the best sensitivity and specificity values.

In terms of PPV percentage, for this scenario the best results were found in classes 1 and 2 however, considering the mean values of all four classes, the Random Forest classifier achieved the best performance, being the only one with the value of 50.0% for PPV.

Comparing the classifiers' performance, when represented with ensemble methods, we were able to determine those that had better results when combined. The Random Forest classifier did not show any improvement however, the J48 method have slightly increased the global accuracy for all the ensemble methods, having the best performance by applying MultiBoosting method.

In the figures presented next (Figure 6.3 to Figure 6.6) the values of PPV, AUC, sensitivity and specificity for each class are represented.

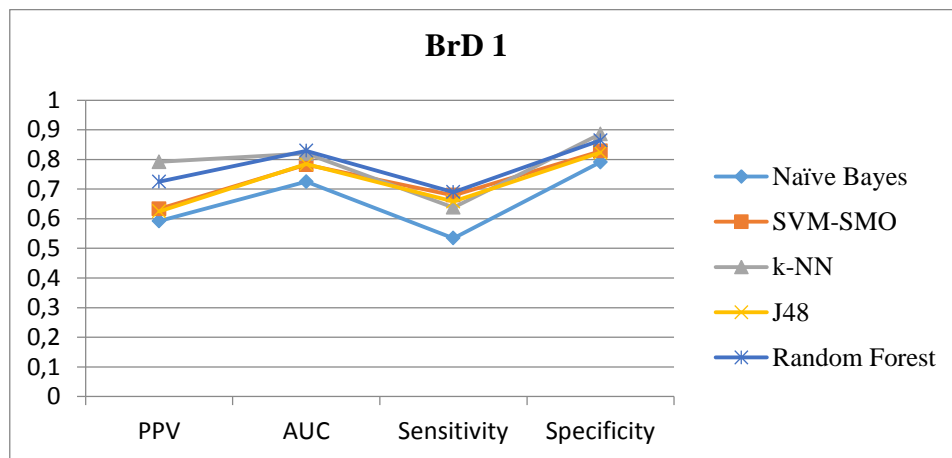


Figure 6.3: Graphic of results extracted from scenario 2 related to class 1 of BrD.

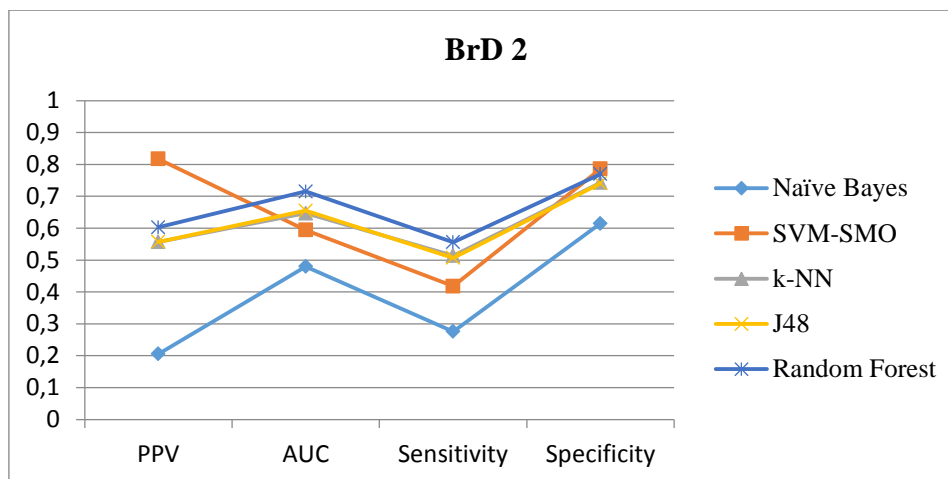


Figure 6.4: Graphic of results extracted from scenario 2 related to class 2 of BrD.

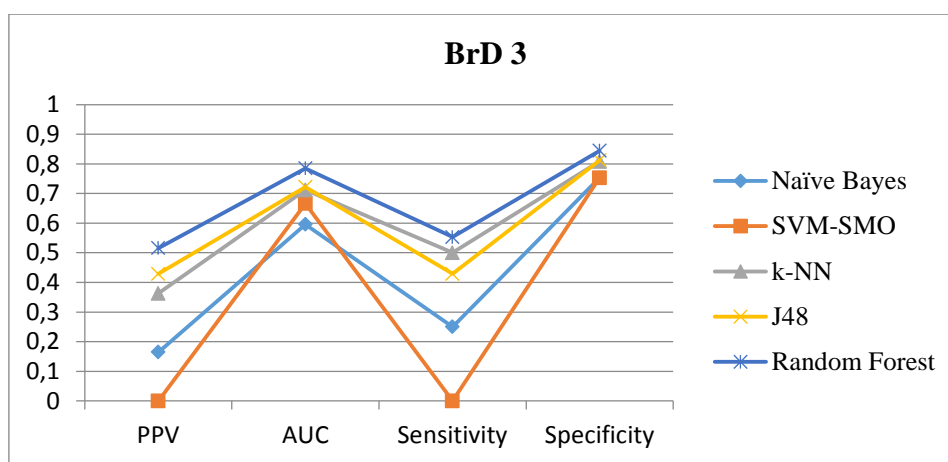


Figure 6.5: Graphic of results extracted from scenario 2 related to class 3 of BrD.

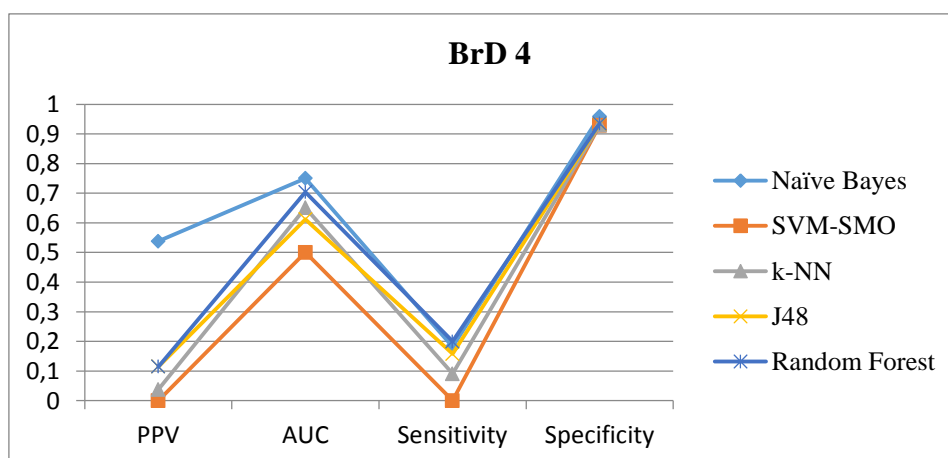


Figure 6.6: Graphic of results extracted from scenario 2 related to class 4 of BrD.

Concerning classification methods we found that global accuracy values were not as good as we were expecting for all classifiers, although the best result was achieved, once again, by Random Forest method. Using the ensemble method, the accuracy reached by J48 was increased, although it did not pass the Random Forest value which was 58.7%.

Although, comparing our results with the ones obtained in (40) using the same database, we realized that for k-NN and SVM methods the results are similar for correctly classified instances (approximately 55%) according to fatty tissue, that corresponds to our 1 and 2 classes which are the ones with highest instances overall. In terms of Random Forest we obtained the best results, comparing to (40), since in this work, the accuracy for fatty and dense tissue was around 52.0% while in our study we obtained an improvement, with 59.0% accuracy.

- **Scenario 3**

The last scenario analyzed was related to the type of finding for each case. In the chosen database, there was a conclusion made by a clinical expert who classifies the mammogram according to mass (101 cases), microcalcification (267 cases) or normal findings (67 cases). In order to find the best predicting method, those findings were grouped into three subsets (3a, 3b and 3c further described).

- **Subset 3a**

The first subset represents the previously given classification in the database.

The results obtained by combining this information with the GLRLM features extracted are present in Table 5.5. The only parameter that changed from the results presented was the k value at the k-NN classifier that we considered equal to 3 in order to achieve better results.

Table 5.5: Results of 3a subgroup.

Classifier	Subclass	PPV	AUC	Sensitivity	Specificity	Global Accuracy
Naïve Bayes	Mass	0.911	0.700	0.348	0.949	58.7 %
	Micro	0.356	0.774	0.913	0.487	
	Normal	1.000	1.000	1.000	1.000	
SVM-SMO	Mass	0.000	0.605	0.000	0.770	76.9 %
	Micro	1.000	0.705	0.726	1.000	
	Normal	1.000	1.000	1.000	1.000	
k-NN	Mass	0.228	0.636	0.315	0.787	70.8 %
	Micro	0.813	0.737	0.736	0.653	
	Normal	1.000	1.000	1.000	1.000	
J48	Mass	0.000	0.600	0.000	0.770	76.9 %
	Micro	1.000	0.703	0.726	1.000	
	Normal	1.000	1.000	1.000	1.000	
Random Forest	Mass	0.257	0.628	0.280	0.783	67.6 %
	Micro	0.749	0.729	0.727	0.591	
	Normal	1.000	1.000	1.000	1.000	

In the Table 5.6 the results for ensemble methods, using J48 and Random Forest, for subset 3a are presented.

Table 5.6: Results of 3a subgroup with ensemble methods.

Method		PPV	AUC	Sensitivity	Specificity	Global Accuracy
Bagging						
J48	Mass	0.010	0.663	0.333	0.771	76.7 %
	Micro	0.993	0.749	0.726	0.973	
	Normal	1.000	1.000	1.000	1.000	
Random Forest	Mass	0.050	0.630	0.128	0.760	70.4 %
	Micro	0.873	0.724	0.708	0.691	
	Normal	1.000	1.000	1.000	1.000	
AdaBoost						
J48	Mass	0.000	0.677	0.000	0.770	76.9 %
	Micro	1.000	0.762	0.726	1.000	
	Normal	1.000	1.000	1.000	1.000	
Random Forest	Mass	0.158	0.607	0.203	0.764	66.3 %
	Micro	0.764	0.713	0.706	0.580	
	Normal	1.000	1.000	1.000	1.000	
MultiBoosting						
J48	Mass	0.030	0.632	0.214	0.769	75.2 %
	Micro	0.959	0.731	0.723	0.871	
	Normal	1.000	1.000	1.000	1.000	
Random Forest	Mass	0.188	0.617	0.268	0.777	69.5 %
	Micro	0.805	0.728	0.724	0.634	
	Normal	1.000	1.000	1.000	1.000	

By analyzing the results obtained from subgroup 3a, in terms of global percentage of correctly classified instances in distinguishing between mass, microcalcification and normal tissue, our tests revealed that for all classifiers, normal tissues show a percentage of 100.0% correctly classified instances. Despite this fact, classifiers such as Naïve Bayes or Random Forest revealed low values at the level of overall effectiveness and the best results were obtained by using SVM-SMO and J48 both representing 76.9% accuracy. Even though these last two methods have shown better global performance, by analyzing their results for each class individually, we found that none of them have correctly classified a single mass, and achieved 100.0% of PPV for microcalcifications. In turn, k-NN with an accuracy of 70.8% has classified every class, even though for masses it did not reveal promising results (PPV = 0.228 and sensitivity = 0.315).

Another interesting fact regarding this scenario was that Naïve Bayes was the only classifier that achieved a better performance in classifying masses rather than microcalcifications. With this method and only concerning masses classification, the percentages of PPV obtained were 99.1% and AUC, sensitivity and specificity also revealed the best results with values of 0.700, 0.348 and 0.949, respectively. These results can be observed in the Figure 6.7 and Figure 6.8, where values of PPV, AUC, sensitivity and specificity are compared for mass and microcalcification classification. In these figures the normal class was not considered due to the invariance of the results.

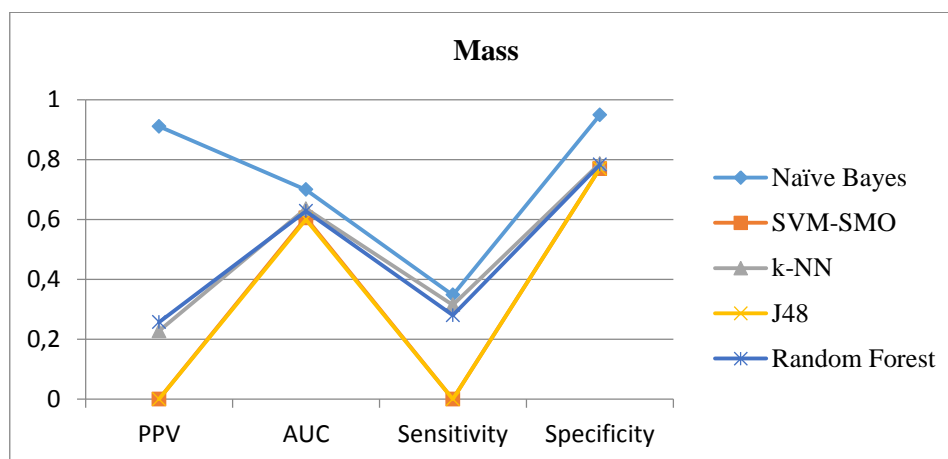


Figure 6.7: Graphic of results extracted from scenario 3a related to mass class.

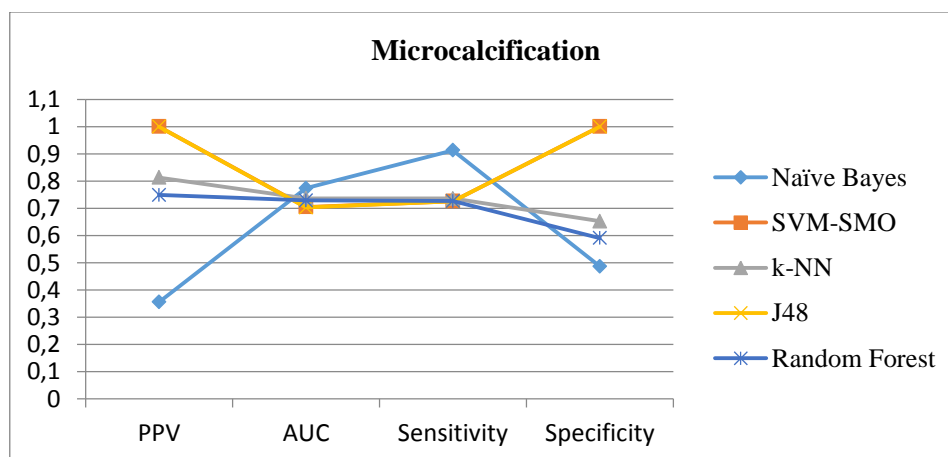


Figure 6.8: Graphic of results extracted from scenario 3a related to microcalcification class.

When comparing the results with the ensemble ones, we were able to identify that best performances were obtained using Random Forest methods and the best result in terms of global accuracy was obtained using bagging methods. In spite of this, the mean percentages for PPV, AUC, sensitivity and specificity did not present better results.

→ Subset 3b

The second subset created is only related to one type of reported lesion – microcalcification. At this point, we categorized the classes previously referred as microcalcification or no microcalcification, excluding all the normal findings. The results for each classifier method and for the ensemble ones are presented at Table 5.7 and 5.8, respectively.

Table 5.7: Results of 3b subgroup.

Classifier	Subclass	PPV	AUC	Sensitivity	Specificity	Global Accuracy
Naïve Bayes	Micro	0.352	0.630	0.879	0.337	49.5 %
	No micro	0.871	0.630	0.337	0.879	
SVM-SMO	Micro	1.000	0.500	0.000	---	72.5 %
	No micro	0.000	0.500	0.726	0.725	
k-NN	Micro	0.798	0.573	0.727	0.280	63.8 %
	No micro	0.208	0.573	0.280	0.727	
J48	Micro	1.000	0.492	0.726	---	72.5 %
	No micro	0.000	0.492	0.000	0.671	
Random Forest	Micro	0.798	0.577	0.753	0.752	66.3 %
	No micro	0.307	0.577	0.365	0.365	

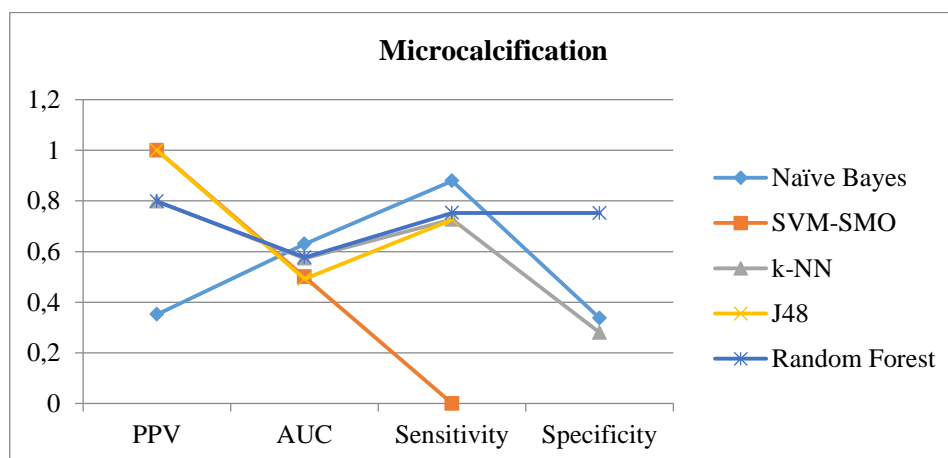
In terms of acquiring better results, for k-NN the preferred k was 3 and for Random Forest classifier we used k=2 and cross-validation with fifteen-folds.

Table 5.8: Results of 3b subgroup with ensemble methods.

Method		PPV	AUC	Sensitivity	Specificity	Global Accuracy
Bagging						
J48	Micro	1.000	0.534	0.726	---	72.5 %
	No micro	0.000	0.534	0.000	0.725	
Random Forest	Micro	0.880	0.618	0.737	0.347	68.5 %
	No micro	0.168	0.618	0.347	0.737	
AdaBoost						
J48	Micro	1.000	0.559	0.726	---	72.5 %
	No micro	0.000	0.559	0.000	0.725	
Random Forest	Micro	0.764	0.539	0.731	0.292	62.5 %
	No micro	0.257	0.539	0.292	0.731	
MultiBoosting						
J48	Micro	0.993	0.599	0.724	0.000	72.0 %
	No micro	0.000	0.599	0.000	0.724	
Random Forest	Micro	0.787	0.610	0.739	0.322	64.4 %
	No micro	0.267	0.609	0.321	0.739	

The best global accuracy in this subset was found using the SVM-SMO and J48 classifiers with 72.5% accuracy for both cases. Likewise previously presented in subset 3a, these two methods have only correctly classified microcalcification with 100.0% of PPV.

In order to better understand the analysis of these results, in Figure 6.9 and Figure 6.10, two graphic related to each subclass (micro, no micro) and their independent results are presented.

**Figure 6.9:** Graphic of results extracted from scenario 3b related to microcalcification class.

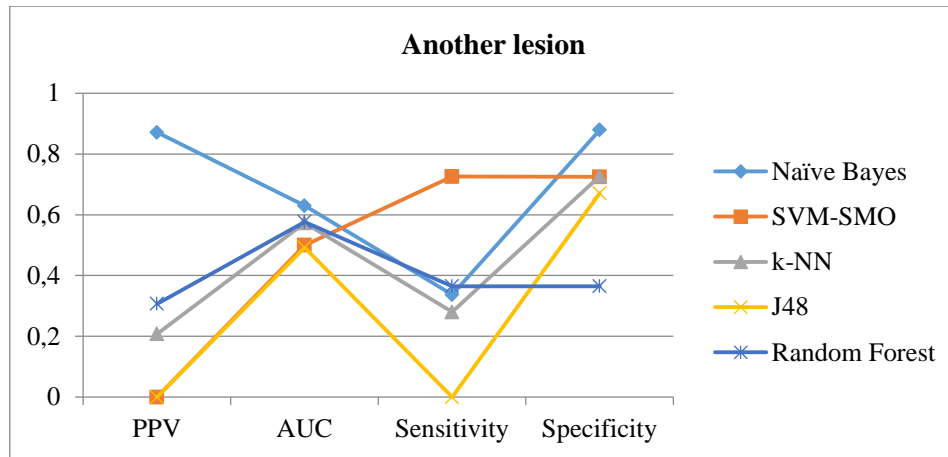


Figure 6.10: Graphic of results extracted from scenario 3b related to no micro class.

Through analysis of Figure 6.9 and Figure 6.10, we were able to recognize that there is no classifier that seems to be better at all evaluated levels. Regardless of this result and considering the PPV, the best performance achieved for micros' identification was made by J48 and SVM-SMO but for other lesions, in this case, mass identification, the results were the worst, Naïve Bayes being the best test.

By improving the results with applying ensemble methods, we have concluded that every result was improved by J48 classifier, but by using Random Forest we found an increase in the value of AUC from 0.577 to 0.618, as well as a slight increase of the global accuracy from 66.3% to 68.5%.

→ Subset 3c

The final subset that was studied concerns a general approach that reveals the presence of a lesion (mass or microcalcification) on the tissue, no matter its type. This subset contains two classes, lesion and normal, and the results for individual classifiers are presented in Table 5.9.

Table 5.9: Results of 3c subgroup.

Classifier	Subclass	PPV	AUC	Sensitivity	Specificity	Global Accuracy
Naïve Bayes	Normal	1.000	1.000	1.000	1.000	100%
	Lesion	1.000	1.000	1.000	1.000	
SVM-SMO	Normal	1.000	0.993	1.000	1.000	100%
	Lesion	1.000	1.000	1.000	1.000	
k-NN	Normal	1.000	0.987	1.000	1.000	100%
	Lesion	1.000	1.000	1.000	1.000	
J48	Normal	1.000	1.000	1.000	1.000	100%
	Lesion	1.000	1.000	1.000	1.000	
Random Forest	Normal	1.000	1.000	1.000	1.000	100%
	Lesion	1.000	1.000	1.000	1.000	

As observed, there was no need to apply the ensemble techniques on the 3c subset since all the individual methods accomplished the maximum value for prediction.

According to our global study on scenario 3 we found that to the identification of masses, even between masses, microcalcifications and normal tissue (subset 3a), or between masses and microcalcifications (subset 3b), the best performance was achieved by Naïve Bayes with PPV values of 91.1% and 87.1%, respectively. In all subsets using normal findings for distinguishing them from lesion tissues we concluded that normal ones have a percentage of 100.0% correctly classified instances.

Taking into consideration all of these conclusions, we can state that the best performance in distinguishing the type of mammographic lesion was obtained at 3a scenario using k-NN classifier (77.0% global accuracy) with PPV percentages of 81.3% for microcalcification and 22.8% for masses (also having the normal class with 100.0% PPV), and with Random Forest classifier at 3b scenario with a global accuracy of 66.0% and percentages of 79.8% and 30.7% for micro and masses classification, respectively. Concerning to AUC values, the best results were achieved by means of k-NN (considering mass, microcalcification and normal findings), and also showed the best specificity. In terms of sensitivity standards the best value was for microcalcifications (approximately 0.70) comparing to the masses results of approximately 0.30.

Chapter 6

Conclusions and Further Work

During the development of this project and based on the study conducted in the state of the art (Chapter 3) it was possible to understand that several studies have recently been done in order to improve breast cancer diagnose. In this sense, the idea behind this thesis was based on finding the best data mining method which allows the linking the different scenarios experienced with matrices considered.

By using the INbreast[®], a recent database but with the disadvantage of not being used in many works yet, making it more difficult to compare among different algorithms, we aim to extrapolate the knowledge and create new conclusions which might be useful for a more accurate diagnosis. Thereby, we intended to contradict the repetitiveness of tests done in other databases, in order that all the previous projects described in this document are based on public databases as MIAS[®] or BCDR[®]. It is also important to underline that the results obtained by those works cannot be fully compared to ours, due to details previously mentioned.

The aim is to compare the PPV, AUC, sensitivity and specificity values obtained, since these are the values which health professionals will be aware of when choosing a system for decision support.

As we have described throughout this document, three different scenarios were selected for the purpose of identifying which is the best predictive method for each scenario and also which is the best predictor overall.

From scenario 1 analysis, we can conclude that using BI-RADS[®] classification for prediction has a large disadvantage because of their 5 classes. Due to this and to avoid inconsistencies derived from the different weight for each class we adopted the method of dividing BI-RADS[®] classes into benign and malignant, according to some studies

conclusions (32, 33). On this test group, the best method was the Random Forest with PPV values around 76.0% and a AUC value greater than 0.8.

Through scenario 2 we observed that in terms of density distinction, using GLCM the results from classes' prediction presented a huge range, maybe due to the different number of initial data in each class and also related to the similarities of features extracted to neighboring classes as 1-2, 2-3 and so far. For Random Forest classifier, as the better one, the PPV and sensitivity values were not invariants according to each class, ranging from 72.5% to BrD 1 and 11.5% to BrD 4. Once again the reason for these results could be related to the features' similarities or to the size of each class represented in the database. An effective approach to improve these results could equal the number of instances in each class or by grouping those four classes into half by way of creating a cutoff point at 50.0% of the glandular tissue. Therefore, we can conclude that even though the results obtained for this scenario were not the best ones, we were able to guarantee a superior performance than a previous work that used the same database.

The conclusions made over scenario 3 were interesting for the reason that masses are in less number (267 micros and 101 masses), a fact emphasized in global accuracy percentages, which are 59.0% on subset 3a and 50.0% on subset 3b, approximately. Even though this global percentage was not optimal, it represents an interesting point for future investigation, since it contradicts other studies which concluded that microcalcification are a more predictable lesion than masses (36, 54). We can conclude that, for the scenario 3, in distinguishing the type of mammographic lesion the best performance was obtained for scenario 3a by the use of k-NN classifier and for Random Forest classifier in scenario 3b.

Unlike other studies previously presented on Chapter 3, by doing this work we obtained results for a range of methods taking into account different types of data organization and providing a wide perception using different lexicons as BI-RADS® and BrD by ACR.

The major conclusions that were made by this work are summarized as:

- GLRLM is an excellent matrix to extract lesion features, and reveals great results when combined with Random Forest classifiers;
- Grouping BI-RADS® class into benign and malignant is an effective way to improve results on predicting breast cancer;

- GLCM without ROI selection can achieve results on classifying breast density types.

Further work can be done to improve the results presented here.

In order to better understand if the global accuracy values are reliable, other databases should be used and the models here applied to compare results. The techniques used on image pre-processing step can also be improved and ROI selection for GLRLM extraction features could be automatically selected in order to avoid misclassifications. To implement these improvements some other software must be considered, because there was a need for faster and more efficient processing using language such as C++. Other findings should also be considered, such as bilateral asymmetry and architectural distortion.

In terms of the methods used to classify, other tests must be done to improve results using breast density discrimination, possibly by grouping classes. To improve the prediction of the tested models, a combination of clinical data (e.g. age and radiologist's observations) could also be implemented, as described in Moura et al. work (82).

Towards this project, we thought that the implementation of a similar system based on the one presented, would be an asset to health professionals in the area of breast cancer diagnosis, especially in Hospital São João (since the images are taken from there). Through the use of sensitivity and specificity values, professional could ensure better diagnoses and essentially faster diagnostic methods. The consolidation of models like the ones here described with the online application work developed by Ferreira et al. (64) could be an important step for the creation of an effective toolkit that could help on reducing misdiagnosis and shortening the detailed and handmade study of each case, guaranteeing more diagnoses made in the same space of time and essentially more efficient.

References

1. Wang X, Li L, Xu W, Liu W, Lederman D, Zheng B. Improving Performance of Computer-aided Detection of Subtle Breast Masses Using an Adaptive Cueing Method. *Phys Med Biol*. 2012;57(2):561-75.
2. Thompson A, Brennan K, Cox A, Gee J, Harcourt D, Harris A, et al. Evaluation of the current knowledge limitations in breast cancer research: a gap analysis. *Breast Cancer Res*. 2008;10(2):R26.
3. Cheng YC, Ueno NT. Improvement of survival and prospect of cure in patients with metastatic breast cancer. *Breast cancer*. 2012;19(3):191-9.
4. Dromain C, Boyer B, Ferre R, Canale S, Delaloge S, Balleyguier C. Computed-aided diagnosis (CAD) in the detection of breast cancer. *European journal of radiology*. 2013;82(3):417-23.
5. Brem RF, Hoffmeister JW, Rapelyea JA, Zisman G, Mohtashemi K, Jindal G, et al. Impact of breast density on computer-aided detection for breast cancer. *American Journal of Roentgenology*. 2005;184(2):439-44.
6. Brem RF. Clinical versus research approach to breast cancer detection with CAD: where are we now? *American Journal of Roentgenology*. 2007;188(1):234-5.
7. Lauria A. GPCALMA: implementation in Italian hospitals of a computer aided detection system for breast lesions by mammography examination. *Phys Med*. 2009;25(2):58-72.
8. Mads Nielsen P. Biomediq Department of Computed Sciences at the University of Copenhagen 2011 [updated 201112 / 02/ 2014]. Available from: biomediq.com/projBC.shtml.
9. Cheng HD, Cai X, Chen X, Hu L, Lou X. Computer-aided detection and classification of microcalcifications in mammograms: a survey. *Pattern Recognition*. 2003;36(12):2967-91.
10. Bovik AC. *Handbook of image and video processing*: Academic Press; 2010. 1195 p.
11. Doi K. *Current status and future potential of computer-aided diagnosis in medical imaging*. 2014.
12. Mazurowski MA, Zurada JM, Tourassi GD. Selection of examples in case-based computer-aided decision systems. *Physics in medicine and biology*. 2008;53(21):6079.

13. LPCC. Liga Portuguesa Contra o Cancro 2009 [13 / 02 / 2014]. Available from: <http://www.ligacontracancro.pt/gca/index.php?id=14>.
14. Bray F, Ren JS, Masuyer E, Ferlay J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *International Journal of Cancer*. 2013;132(5):1133-45.
15. Farmer P, Frenk J, Knaul FM, Shulman LN, Alleyne G, Armstrong L, et al. Expansion of cancer care and control in countries of low and middle income: a call to action. *The Lancet*. 2010;376(9747):1186-93.
16. Society AC. American Cancer Society: Breast Cancer detailed guide 2013. Available from: <http://www.cancer.org/cancer/breastcancer/>.
17. Kumar V, Abbas A, Fausto N, Aster JC. Robbins and Cotran Pathologic Basis of Disease, Professional Edition: Expert Consult-Online: Elsevier Health Sciences; 2009.
18. Khatib O, Modjtabai A. Guidelines for the early detection and screening of breast cancer.
19. Health NIO. Breast Cancer. ADAM Medical Encyclopedia: PubMed Health; 2012.
20. Bontrager K. Tratado de técnica radiológica e base anatômica: Guanabara Koogan; 2003.
21. Force UPST. Screening for breast cancer: US Preventive Services Task Force recommendation statement. *Annals of internal medicine*. 2009;151(10):716.
22. Caldas F, Isa H, Trippia A, Bísaro A, Souza E, Tajara L. Controle de qualidade e artefatos em mamografia. *Radiol Bras*. 2005;38(4):295-300.
23. Perre A, Freire L. Avaliação da caracterização de lesões em mamografia com recurso a sistemas CAD (Computer-Aided Diagnosis). 2014.
24. Radiology ACo. ACR BI-RADS® Atlas: American College of Radiology; [02/02/2014]. Available from: <http://www.acr.org/Quality-Safety/Resources/BIRADS/About-BIRADS>.
25. Boyd N, Martin LJ, Bronskill M, Yaffe MJ, Duric N, Minkin S. Breast tissue composition and susceptibility to breast cancer. *Journal of the National Cancer Institute*. 2010.
26. Boyd NF, Martin LJ, Yaffe MJ, Minkin S. Mammographic density and breast cancer risk: current understanding and future prospects. *Breast Cancer Research*. 2011;13(6):223.

27. Wang AT, Vachon CM, Brandt KR, Ghosh K, editors. Breast Density and Breast Cancer Risk: A Practical Review. Mayo Clinic Proceedings; 2014: Elsevier.
28. Zheng B, Wang X, Lederman D, Tan J, Gur D. Computer-aided detection: the effect of training databases on detection of subtle breast masses. Academic radiology. 2010;17(11):1401-8.
29. Gierach GL, Ichikawa L, Kerlikowske K, Brinton LA, Farhat GN, Vacek PM, et al. Relationship between mammographic density and breast cancer death in the Breast Cancer Surveillance Consortium. Journal of the National Cancer Institute. 2012.
30. Ayres FJ, Rangayyan R. Characterization of architectural distortion in mammograms. Engineering in Medicine and Biology Magazine, IEEE. 2005;24(1):59-67.
31. Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS. INbreast: toward a full-field digital mammographic database. Academic radiology 2012; 2012 1076-6332.
32. Lacquement MA, Mitchell D, Hollingsworth AB. Positive predictive value of the breast imaging reporting and data system. Journal of the American College of Surgeons. 1999;189(1):34-40.
33. Obenauer S, Hermann K, Grabbe E. Applications and literature review of the BI-RADS classification. European radiology. 2005;15(5):1027-36.
34. Malucelli A, Stein Junior Av, Bastos L, Carvalho D, Cubas MR, Paraíso EC. Classification of risk micro-areas using data mining. Revista de saude publica. 2010;44(2):292-300.
35. Witten IH, Frank E. Data Mining: Practical machine learning tools and techniques: Morgan Kaufmann; 2005.
36. Amendolia SR, Bisogni MG, Bottigli U, Ceccopieri A, Delogu P, Dipasquale G, et al. The CALMA project. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. 2001;461(1):428-9.
37. Oliver A, Freixenet J, Martí R, Zwiggelaar R. A comparison of breast tissue classification techniques. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006: Springer; 2006. p. 872-9.
38. Bueno G, Vállez N, Déniz O, Esteve P, Rienda MA, Arias M, et al. Automatic breast parenchymal density classification integrated into a CADe system. International journal of computer assisted radiology and surgery. 2011;6(3):309-18.

39. Lesniak J, Hupse R, Blanc R, Karssemeijer N, Székely G. Comparative evaluation of support vector machine classification for computer aided detection of breast masses in mammography. *Physics in medicine and biology*. 2012;57(16):5295.
40. da Fonseca JL, Cardoso JS, Domingues I. Pre-CADs in Breast Cancer. *idea*. 2013;2:3.
41. Pollán RR. Improving multilayer perceptron classifiers AUC performance - An approach in biomedical image analysis for breast cancer CAD supported by eInfrastructures: University of Porto; 2012.
42. de Oliveira Martins L, Junior GB, Silva AC, de Paiva AC, Gattass M. Detection of masses in digital mammograms using K-means and support vector machine. *Electronic Letters on Computer Vision and Image Analysis*. 2009;8(2):39-50.
43. Ramos-Pollán R, Guevara-López M, Oliveira E. A software framework for building biomedical machine learning classifiers through grid computing resources. *Journal of medical systems*. 2012;36(4):2245-57.
44. Ramos-Pollán R, Guevara-López M, Suárez-Ortega C, Díaz-Herrero G, Franco-Valiente JM, Rubio-del-Solar M, et al. Discovering mammography-based machine learning classifiers for breast cancer diagnosis. *Journal of medical systems*. 2012;36(4):2259-69.
45. Mohanty AK, Senapati MR, Beberta S, Lenka SK. Texture-based features for classification of mammograms using decision tree. *Neural Computing and Applications*. 2013;23(3-4):1011-7.
46. Breiman L. Random forests. *Machine learning*. 2001;45(1):5-32.
47. Gray KR, Aljabar P, Heckemann RA, Hammers A, Rueckert D. Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *NeuroImage*. 2013;65:167-75.
48. Melendez J, Sánchez CI, van Ginneken B, Karssemeijer N. Improving mass candidate detection in mammograms via feature maxima propagation and local feature selection. *Medical Physics*. 2014;41(8):081904.
49. Zhang H. The optimality of naive Bayes. *AA*. 2004;1(2):3.
50. Castella C, Kinkel K, Eckstein MP, Sottas P-E, Verdun FR, Bochud FO. Semiautomatic mammographic parenchymal patterns classification using multiple statistical features. *Academic radiology*. 2007;14(12):1486-99.
51. Berner ES. *Clinical Decision Support Systems*: Springer; 2007.

52. American Medical Association AMA. 2014. Available from: <http://www.ama-assn.org/ama>.
53. Thangavel K, Karnan M, Sivakumar R, Mohideen AK. Automatic detection of microcalcification in mammograms-a review. *International Journal on Graphics Vision and Image Processing*. 2005;5(5):31-61.
54. Bellotti R, Bagnasco S, Bottigli U, Castellano M, Cataldo R, Catanzariti E, et al., editors. The MAGIC-5 Project: medical applications on a GRID infrastructure connection. *Nuclear Science Symposium Conference Record, 2004 IEEE*; 2004: IEEE.
55. Valiente JMF, Ortega CS, del Solar MR, Herrero GD, Pollán RR, López MAG, et al., editors. MIWAD: A Software Suite For Building CAD Methods. 15th International Conference on Experimental Mechanics (22-17 July 2012), Porto, Portugal.
56. del Solar MR, Herrero GD, Pollan RR, Lopez MAG, de Posada NG, Moura DC, et al. Improving the breast cancer diagnosis using digital repositories.
57. Meselhy Eltoukhy M, Faye I, Belhaouari Samir B. A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation. *Computers in biology and medicine*. 2012;42(1):123-8.
58. Chikamai K, Viriri S, Tapamo JR. Combining Feature Methods for Content-Based Classification of Mammogram Images. *International Journal of Computers Communications & Control*. 2013;8(4):499-513.
59. Manavalan R, Thangavel K. Evaluation of textural feature extraction from Grfm for prostate cancer trus medical images. *International Journal of Advances in Engineering & Technology*. 2012;1(6).
60. Zulpe N, Pawar V. GLCM textural features for brain tumor classification. *IJCSI International Journal of Computer Science Issues*. 2012;9(3).
61. Karahaliou AN, Boniatis IS, Skiadopoulos SG, Sakellaropoulos FN, Likaki E, Panayiotakis GS, et al., editors. A texture analysis approach for characterizing microcalcifications on mammograms. *Proc Int Special Topic Conf on Inf Technol Biomed*; 2006: Citeseer.
62. Gebejes A, Huertas R. Texture Characterization based on Grey-Level Co-occurrence Matrix. 2013.
63. Nanni L, Brahnam S, Ghidoni S, Menegatti E, Barrier T. Different approaches for extracting information from the co-occurrence matrix. *PloS one*. 2013;8(12):e83554.

64. Ferreira P, Fonseca NA, Dutra I, Woods R, Burnside E, editors. Predicting malignancy from mammography findings and surgical biopsies. *Bioinformatics and Biomedicine (BIBM)*, 2011 IEEE International Conference on; 2011: IEEE.
65. Khuzi AM, Besar R, Zaki WW, Ahmad N. Identification of masses in digital mammogram using gray level co-occurrence matrices. *Biomedical Imaging and Intervention Journal*. 2009;5(3):e17.
66. Tai S, Chen Z, Tsai W. An Automatic Mass Detection System in Mammograms based on Complex Texture Features. 2014.
67. Eddaoudi F, Regragui F. Microcalcifications detection in mammographic images using texture coding. *Applied Mathematical Sciences*. 2011;5(8):381-93.
68. Rajamanickam K. Gray level run length matrix 2010. Available from: <http://www.mathworks.com/matlabcentral/fileexchange/26694-gray-level-run-length-matrix/content/GLRLM.m>.
69. Chen C-h, Pau L-F, Wang PS-p. *Handbook of pattern recognition and computer vision*: World Scientific; 2010.
70. Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*. 1973(6):610-21.
71. Pham TA. *Optimization of Texture Feature Extration Algorithm*: Delft University of Technology, Netherlands; 2010.
72. Bovis K, Singh S, editors. Detection of masses in mammograms using texture features. *Pattern Recognition, 2000 Proceedings 15th International Conference on*; 2000: IEEE.
73. Yang S-C, Lin Y-J, Chung P-C, Hsu G-C, Lo C-S, editors. Mass screening and feature reserved compression in a computer-aided system for mammograms. *Neural Networks, 2008 IJCNN 2008(IEEE World Congress on Computational Intelligence) IEEE International Joint Conference on*; 2008: IEEE.
74. Ball JE, Bruce LM, editors. Digital mammogram spiculated mass detection and spicule segmentation using level sets. *Engineering in Medicine and Biology Society, 2007 EMBS 2007 29th Annual International Conference of the IEEE*; 2007: IEEE.
75. Elshinawy MY, Adviser-Chouikha MF. *Pre-CAD normal mammogram detection algorithm based on tissue type*: Howard University; 2010.
76. Chu A, Sehgal C, Greenleaf JF. Use of gray value distribution of run lengths for texture analysis. *Pattern Recognition Letters*. 1990;11(6):415-9.

77. Dasarathy BV, Holder EB. Image characterizations based on joint gray level—run length distributions. *Pattern Recognition Letters*. 1991;12(8):497-502.
78. Pallant J. *SPSS survival manual: A step by step guide to data analysis using SPSS*: McGraw-Hill International; 2010.
79. Janitza S, Strobl C, Boulesteix A-L. An AUC-based permutation variable importance measure for random forests. *BMC bioinformatics*. 2013;14(1):119.
80. Lin W-J, Chen JJ. Class-imbalanced classifiers for high-dimensional data. *Briefings in bioinformatics*. 2012;bbs006.
81. Luo S-T, Cheng B-W. Diagnosing breast masses in digital mammography using feature selection and ensemble methods. *Journal of medical systems*. 2012;36(2):569-77.
82. Moura DC, López MAG. An evaluation of image descriptors combined with clinical data for breast cancer diagnosis. *International journal of computer assisted radiology and surgery*. 2013;8(4):561-74.