Instituto Superior
de Engenharia
Politécnico de Coimbra

DEPARTMENT OF SYSTEMS AND COMPUTER ENGINEERING

Enhancing Breast Cancer Diagnosis: A Mammogram Retrieval System and Ground Truth Application

Dissertation to fultill the Master's degree in Informatics Engineering

Specialization Area of Intelligent Data Analysis

Author

Cátia Inês Melo Roriz

Supervisor

Inês Campos Monteiro Sabino Domingues

Coimbra, December 2023

INSTITUTO POLITÉCNICO
DE COIMBRA

INSTITUTO SUPERIOR
DE ENGENHARIA
DE COIMBRA

Summary

Breast cancer is a significant global health concern, affecting thousands of individuals, primarily women, with estimated cases expected to climb by 2040. Early-stage diagnosis is essential for effective treatment and better patient outcomes. This dissertation presents a mammogram retrieval system based on the aggregation of image classifiers to aid specialists in diagnosing breast cancer. The system uses a retrieval model that combines the output of multiple classifiers, each targeting different dimensions related to breast cancer diagnosis. These dimensions include breast density, asymmetries, BI-RADS classification, calcifications, distortions, laterality, masses, and image incidence. This dissertation also describes the creation of an application to collect ground truth data to aid engineers in the development of a mammography retrieval system. The application is built upon OutSystems, a low-code application platform. Key features of the application include allowing experts to view probe images and associate them with relevant images from the database. Additionally, the platform allows image filtering based on eight mammogram dimensions. While the ultimate goal is to create a system for medical specialists, the current platform represents a step in the process, facilitating the acquisition of ground truth. As for the results obtained from the individual models, in the training set, for the models of each dimension, they reach an average accuracy of around 99.3%, while in the test set, the average accuracy is around 78%. Four approaches were then developed for the final retrieval model, one assigning equal weights to every dimension, another with empirically defined weights, a third where the weights were defined according to the literature, and a final one where the values of the weights were defined by a specialist. The quantitative results of the final retrieval model according to the four approaches represent the similarity between the probe image and the most similar image (the first image in the top-5). The similarity results are the result of using the individual models in a weighted sum. The first approach scored 0.319, the second 0.191, the third 0.197 and finally the last 0.292.

Keywords: Mammogram retrieval, Breast cancer, Image classification, Deep Learning, Software

Dedication

I dedicate this work to my parents, brothers, sister-in-law and nephews who, with great affection and support, spared no effort to get me to this stage in my life. To my cousin Joel (in memorian), who is no longer with us, but who continues to be my greatest strength in life. His memory inspires me and makes me persist. To my friend Catarina Batista, a great supporter of my studies and a partner at all times. To the Master's degree in Informatics Engineering, to the people I've met over the years, especially my brilliant friends: Ana Alves, Renata Pereira and Rafael Abrantes for the exceptional support and encouragement they gave me throughout this dissertation. Finally, thank you to all the people who have contributed to my success and to my growth as a person.

Acknowledgements

I would like to thank my supervisor, Professor Inês Domingues, for agreeing to accompany me on this project. Her commitment was essential to my motivation as difficulties arose along the way. To Professors Verónica and Inês Moreira for all their support and help in carrying out this work, I thank you with deep admiration for your professionalism. Finally, I would like to express my gratitude to all the professionals in the computer engineering department at the Coimbra Institute of Engineering (ISEC).

Index

Summary . . . i

Dedication . . . ii

Acknowledgements . . . iii

Index . . . 1

Index of tables . . . 3

Index of figures . . . 4

List of acronyms and abbreviations . . . 6

1 Introduction . . . 8
- 1.1 Motivation . . . 8
- 1.2 Contributions . . . 9
- 1.3 Document structure . . . 9

2 Background Knowledge . . . 11
- 2.1 Application . . . 11
- 2.2 Breast cancer . . . 11
- 2.3 Models . . . 18
- 2.4 Summary . . . 19

3 State of the art . . . 20
- 3.1 Retrieval . . . 20
- 3.2 Available Mammogram Retrieval Software . . . 25
- 3.3 Summary . . . 25

4 Methods . . . 28
- 4.1 Database . . . 28
- 4.2 Models . . . 30
  - 4.2.1 Classifiers . . . 31
  - 4.2.2 Retrieval Model . . . 31
  - 4.2.3 Experimental Setting . . . 32

4.3 Summary . . . 32

5 Software developed . . . 34
5.1 Outsystems . . . 34
5.2 Requirements analysis . . . 37
5.3 Application database . . . 38
5.4 Implementation . . . 38
5.5 Ground truth input analysis . . . 41
5.5.1 Analysis . . . 41
5.5.2 Software improvements . . . 44
5.6 Summary . . . 44

6 Results . . . 46
6.1 Individual classifiers . . . 46
6.2 Division by patient . . . 48
6.3 Retrieval models . . . 49
6.4 Evaluation . . . 53
6.4.1 All images . . . 53
6.4.2 Evaluation of division by patient . . . 54
6.5 Summary . . . 54

7 Conclusion . . . 56

Bibliographical references . . . 58

Annexes . . . 62
Annex A - Selection of papers . . . 63
Annex B - CIARP paper . . . 64
Annex C - RecPad Extended Abstract . . . 80
Annex D - RecPad poster . . . 83
Annex E - Industrial Seminars paper . . . 85
Annex F - Software operating instructions . . . 93
Annex G - Retrieval extended results . . . 94
Annex H - Version 2 of the extended results retrieval . . . 98

Index of tables

3.1 Literature review of retrieval works in breast cancer 26
3.2 Competing Software 27

6.1 Summary of weights in different models 50
6.2 Retrieval results 50

Index of figures

2.1 Anatomy of a woman's breast . . . 12
2.2 Mass and Benign calcifications . . . 13
2.3 Malignant calcification and Architectural distortion . . . 13
2.4 Size asymmetry and Focal asymmetry . . . 14
2.5 Global asymmetry . . . 14
2.6 Mammogram examples with different densities . . . 14
2.7 Mammogram examples with and without asymmetry . . . 15
2.8 Mammogram examples with different BI-RADS classification . . . 15
2.9 Mammogram examples with and without calcifications . . . 15
2.10 Mammogram examples with and without distortions . . . 16
2.11 Mammogram examples of different lateralities . . . 16
2.12 Mammogram examples with and without masses . . . 16
2.13 Mammogram examples of different views . . . 17
2.14 Examples of mammograms . . . 17

4.1 Overview of the training methodology . . . 28
4.2 Findings from the INBreast database . . . 29
4.3 Annotated dimensions available in InBreast . . . 30
4.4 BI-RADS classification . . . 31

5.1 User interaction . . . 35
5.2 WEB_Response_Requests . . . 35
5.3 Server-side . . . 36
5.4 Client-side . . . 36
5.5 Use case diagram . . . 38
5.6 Database . . . 39
5.7 Static Entities Records . . . 39
5.8 Login page . . . 40
5.9 Mammogram retrieval page . . . 40
5.10 Registry example . . . 41
5.11 Ground Truth . . . 42
5.12 Number of equal values per line . . . 43
5.13 Average number of equal values per line . . . 43
5.14 Average Equal Values per dimensions . . . 44

6.1 BI-RADS and breast density results. 46
6.2 Mass and calcifications results. 47
6.3 View and laterality results. 47
6.4 Asymmetries and distortions results. 47
6.5 Results obtained in the test set. 48
6.6 Results of patient-wise division. 49
6.7 Results obtained in the test set for the division by patient. 49
6.8 Probe image along with the ground truth 51
6.9 Top-5 images retrieved by r_{const} 51
6.10 Top-5 images retrieved by r_{emp} 52
6.11 Top-5 images retrieved by r_{lit} 52
6.12 Top-5 images retrieved by r_{exp} 52
6.13 Retrieval evaluation 53
6.14 Retrieval evaluation - Version 2 54

7.1 Collection of papers. 63
7.2 Choice of papers with citations. 63
7.3 Choice of papers without citations. 63

List of acronyms and abbreviations

ACR	*American College of Radiology*
ANN	*Artificial Neural Network*
BI-RADS	*Breast Imaging-Reporting and Data System*
CC	*Cranio-Caudal*
CAD	*Computer-Aided Diagnostics*
CNN	*Convolutional Neural Networks*
DEMS	*Dokuz Eylul University Mammogram Set*
DANN	*Discriminant Adaptive Nearest Neighbours*
DBMS	*Database Management System*
DDSM	*Digital Database for Screening Mammography*
DICOM	*Digital Imaging and Communications in Medicine*
FFDM	*Full-field digital mammography*
GRNN	*General Regression Neural Network*
HTML	*HyperText Markup Language*
HTTP	*HyperText Transfer Protocol*
ISEC	*Institute of Engineering of Coimbra*
JS	*JavaScript*
KNN	*K-nearest neighbors algorithm*
LLNL	*Lawrence Livermore National Laboratory*
LDA	*Linear Discriminant Analysis*
MIA	*Mammographic Image Analysis Society*
MAP	*Mean Average Precision*
MCs	*Microcalcifications*
MART	*Ontology-based Mammogram Annotation and Retrieval Tool*
NB	*Naive Bayes*
PACS	*Picture Archiving and Communication System*
PCA	*Principle Component Analysis*
REST	*Representational State Transfer*
RF	*Random Forest*
ROI	*Regions of Interested*
RWTH	*Rheinisch-Westfälische Technische Hochschule*

SIFT	*Scale Invariant Feature Transform*
SOA	*Semantic Query-Enhanced Web Rule Language*
SQWRL	*Semantic Query-Enhanced Web Rule Language*
SVM	*Support Vector Machine*
SVD	*Singular Value Decomposition*
URL	*Uniform Resource Locator*

1 Introduction

This document describes the work carried out within the scope of the dissertation, included in the Master's Degree in Computer Engineering in the area of specialisation of intelligent data analysis of the Coimbra Institute of Engineering (ISEC), in the academic year 2022/2023.

1.1 Motivation

Breast cancer is a malignant disease that affects thousands of individuals worldwide, primarily women. Although it can also occur in men, they represent a minority of cases. By the year 2040 [1], a significant increase in the number of breast cancer cases is estimated, with more than 3 million new cases annually, representing a 40% increase. In addition, the number of deaths related to this disease is expected to increase by 50%, exceeding 1 million deaths per year. Breast cancer [1] is currently the most common type of cancer diagnosed, accounting for about 1 in 8 cancer cases worldwide. In 2020, approximately 2.3 million new cases of breast cancer were reported globally, resulting in about 685,000 deaths. However, it is important to note that these numbers vary significantly between countries and regions [1]. Although various approaches have been developed, much more effort has to be done to successfully fight this disease.

In Portugal [2], approximately 7,000 new cases of breast cancer are diagnosed annually, resulting in around 1,800 deaths among women. About 1% of breast cancer cases in Portugal are in men [2]. Given the recommended two-year screening interval for women aged 45 to 69, taking Portugal as an example, a substantial number of mammograms require analysis each day [3]. The objective of this work is to develop a framework that helps the specialists to perform an early-stage diagnosis of breast cancer, facilitating prompt treatment for patients. Timely treatment significantly improves the chances of a swift recovery and prevents disease progression.

The main goal of the present work is to develop a medical aid tool that, when loading an image to analyse, the system returns a set of relevant images with known diagnoses. To achieve this main goal, a retrieval model that combines the outputs of several classifiers was developed.

An application for collecting ground-truth data was created to help engineers develop a mammogram retrieval system. The application is built upon OutSystems, a low-code application platform. Key features of the application include allowing experts to view

probe images and associate them with relevant images from the database. Additionally, the platform allows image filtering based on eight (8) mammogram dimensions.

It was therefore motivating to start this dissertation and find a solution that would make it possible to diagnose breast cancer at an early stage, so that patients could receive treatment as soon as possible. By receiving treatment early, the patient has a better chance of recovering quickly and not letting the disease progress.

1.2 Contributions

During the preparation of this dissertation, the following contributions were produced:

- The development of individual classifiers according to eight dimensions of the mammograms;
- The proposal of a framework for mammogram image retrieval;
- The instantiation of the above framework, producing four different concrete retrieval models that combine the above mentioned eight classifiers;
- Development of software to collect ground truth.

Besides, the writing of the present work, the following scientific efforts were made:

- Writing a paper for the 26th Iberoamerican Congress on Pattern Recognition (see Annex B - CIARP paper). This paper was rejected. Suggestions from the reviewers were taken into consideration when writing the present document.
- Writing a paper for the 29th Portuguese Conference on Pattern Recognition (see Annex C - RecPad Extended Abstract). This paper was accepted.
- Poster preparation for the 29th Portuguese Conference on Pattern Recognition (see Annex D - RecPad poster). The poster was presented on the 27th of October, 2023.

1.3 Document structure

To make this document easier to understand, it is divided into seven chapters:

- Introduction - The first chapter introduces the dissertation, the motivations behind its development, the respective contributions and finally the structure of this document;
- Background Knowledge - The second chapter provides a detailed description of previous knowledge about the area in question;
- State of the art - The third chapter studies works with similar functionalities and objectives to the one we intend to develop in this dissertation;

- Methods - The fourth chapter describes all the methods adopted during the development of this dissertation, from the database used, the classifiers produced for each dimensions, the retrieval models that resulted from combining the eight classifiers and finally the experimental setting;
- Software developed - The fifth chapter describes the entire environment of the software developed, from the general description of the platform used to develop the software, to the analysis of the necessary requirements, the creation of the software database and its implementation. Finally, an analysis is made of the inputs (ground truth) provided by an specialist who tested the software;
- Results - The sixth chapter presents and analyses the results obtained in relation to the individual classifiers, retrieval models and the evaluation of the retrieval models;
- Conclusions - The seventh chapter presents the final conclusions of the dissertation, including the objectives achieved and possible future work.

2 Background Knowledge

This section covers all the knowledge necessary to understand the subject of this dissertation.

2.1 Application

Breast cancer and breast cancer awareness have led to more and more advances in the treatment of this disease. Today it is possible for patients to lead a relatively normal life. However, due to the high number of deaths, more and more researchers need to study this disease. Mammography [4] is the tool that best enables medical specialists to discover the disease at an early stage. Mammograms can show various characteristics such as breast density, asymmetries, masses, calcifications and distortions. However, mammograms [4] sometimes fail to detect some cancers and in other cases they find that there is cancer even though there is not. When an abnormal area is identified [4], it is necessary to carry out a biopsy because this is the only way to be sure whether there is cancer or not.

2.2 Breast cancer

Breast cancer is a pathology characterised by the abnormal, autonomous and uncontrolled development of breast tissue, usually in the ducts (channels that lead milk to the nipple) or lobules (glands responsible for milk production) (see Figure 2.1) [5]. Although it is a more common condition in women, it is important to note that it can also occur in men, although it is extremely rare [5]. The process of breast cancer begins when healthy cells in the breast undergo changes, resulting in disordered and uncontrollable growth, forming a mass known as a tumour [5]. This tumour can be classified as invasive or non-invasive, invasive being when it has the ability to spread to other parts of the body [5].

The American College of Radiology (ACR) has developed the Breast Imaging Reporting and Data System (BI-RADS) scale. When mammograms are examined, specific abnormalities are sought [6]. The most common findings on mammograms are masses, calcifications, architectural distortion of the breast tissue and asymmetries when comparing the two breasts and the two views. According to the BI-RADS scale, it is possible to classify the aforementioned lesions into six BI-RADS categories [6]:

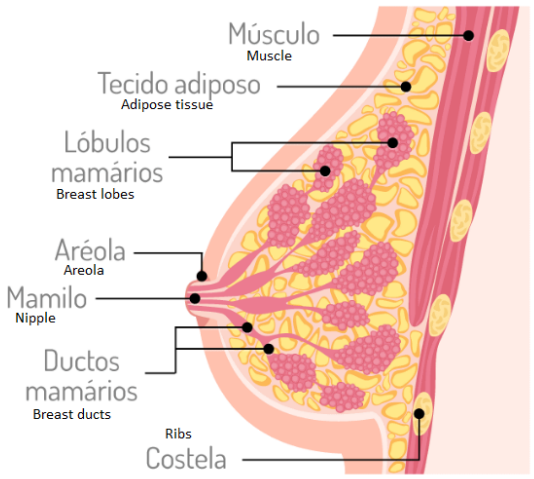
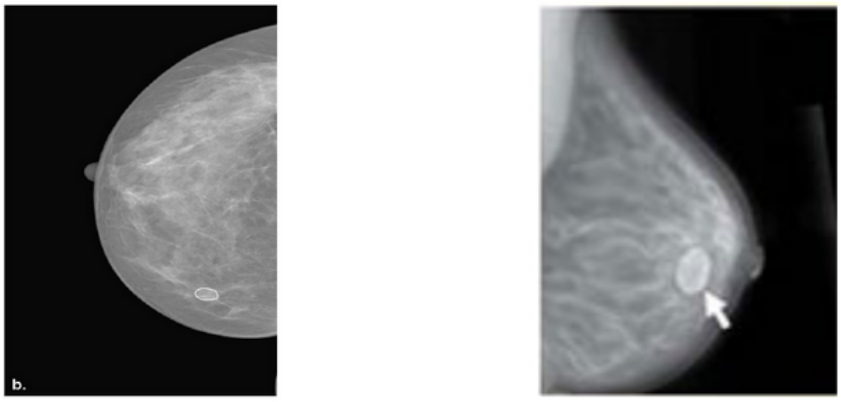


Figure. 2.1: Anatomy of a woman's breast [5]

- Category 0: The examination is inconclusive;
- Category 1: No findings;
- Category 2: Benign findings;
- Category 3: Probably benign findings;
- Category 4: Suspicious findings;
- Category 5: High probability of malignancy;
- Category 6: Proven cancer.

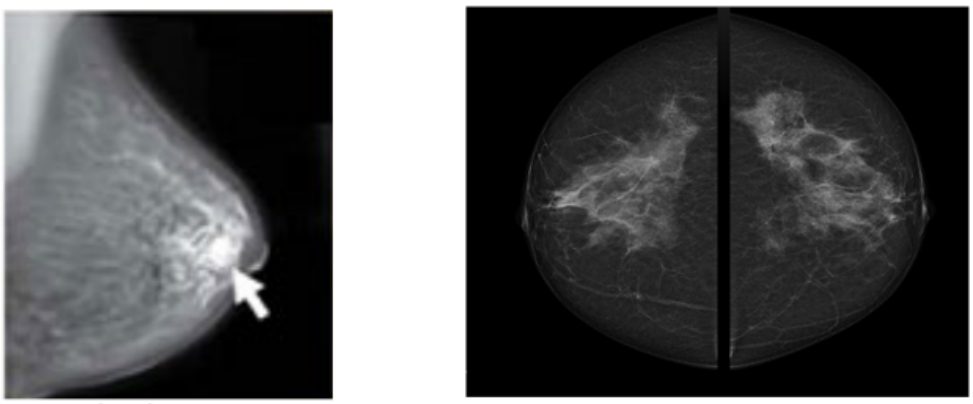
In the case of categories 4 and 5, a biopsy is required to exclude or confirm malignancy [6]. An important characteristic referred to by the ACR is the composition of the breast tissue, related to the breast density shown on the X-rays [6]. There are four categories ranging from 1 for low density (fatty tissue) to 4 for very high density (dense tissue). According to BI-RADS [6]:

- Mass: defined as a three-dimensional structure that has convex outer edges, usually evident in two orthogonal views (see Figure 2.2a).
- Benign calcifications: larger than calcifications associated with malignancy, usually coarser and rounder with smooth margins and more easily seen (see Figure2.2b).
- Calcifications associated with malignancy: usually very small (see Figure 2.3a).
- Architectural distortion: defined as a focal interruption of the normal mammographic pattern of lines (converging at the nipple), usually presenting a star-shaped distortion, with no visible defined mass (see Figure 2.3b).



(a) Mass [6] (b) Benign calcifications [7]

Figure. 2.2: Mass and Benign calcifications

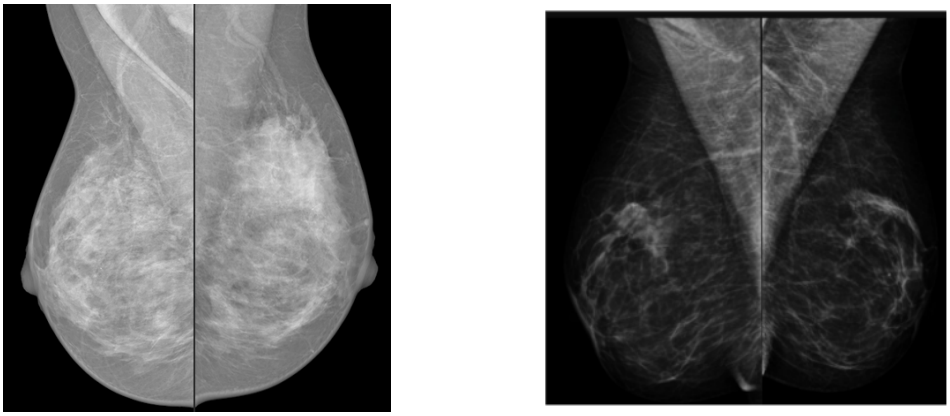


(a) Malignant calcification [7] (b) Architectural distortion [8]

Figure. 2.3: Malignant calcification and Architectural distortion

- Asymmetry: has no convex outer edges of a mass and can be represented in three ways [6]:
  - Size asymmetry (difference in volume between the right and left breast) (see Figure 2.4a);
  - Focal asymmetry (unilateral and localised area of parenchyma) (see Figure 2.4b);
  - Global asymmetry (difference in the amount of parenchyma between the right and left breast) (see Figure 2.5).

The ACR classifies mammograms according to various dimensions such as(ACR) [10] breast density (see Figure 2.6), existence or not of asymmetries (see Figure 2.7), BI-RADS classification (see Figure 2.8), existence or not of calcifications (see Figure 2.9), existence or not of distortions (see Figure 2.10), laterality (right or left breast) (see Figure 2.11), existence or not of masses (see Figure 2.12) and image incidence (Cranio-Caudal (CC) or Medio-Lateral Oblique (MLO)) (see Figure 2.13) [6].



(a) Size asymmetry [6] (b) Focal asymmetry [9]

Figure. 2.4: Size asymmetry and Focal asymmetry

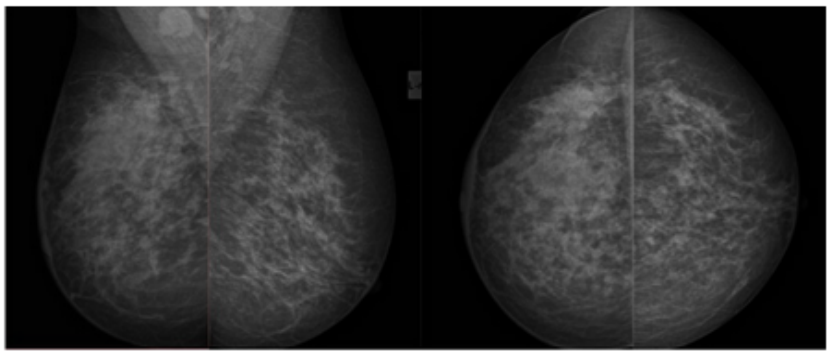


Figure. 2.5: Global asymmetry [9]

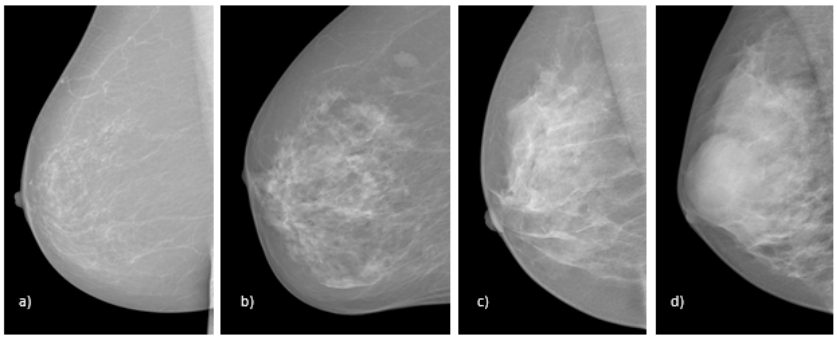


Figure. 2.6: Mammogram examples with different densities: a) ACR breast density 1; b) ACR breast density 2; c) ACR breast density 3; d) ACR breast density 4 [6]

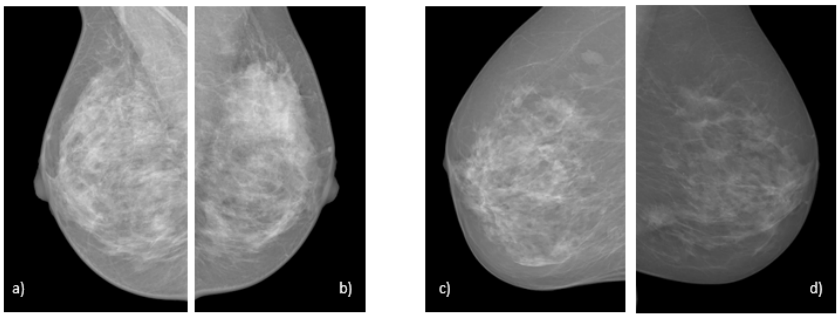


Figure. 2.7: Mammogram examples with and without asymmetry: a) and b) asymmetries; c) and d) no asymmetries [6]

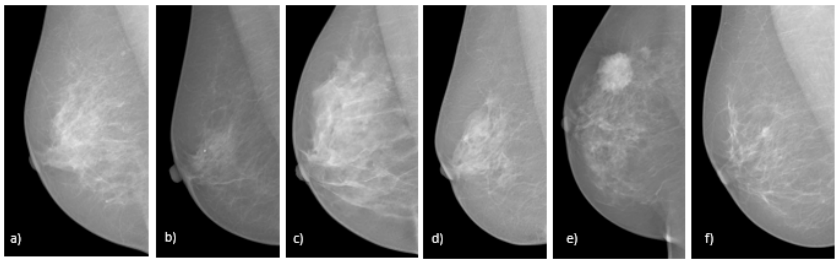


Figure. 2.8: Mammogram examples with different BI-RADS classification: a) BI-RADS 1; b) BI-RADS 2; c) BI-RADS 3; d) BI-RADS 4; e) BI-RADS 5; f) BI-RADS 6 [6]

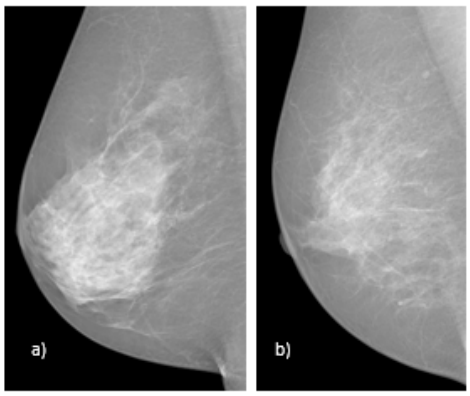


Figure. 2.9: Mammogram examples with and without calcifications: a) calcification; b) no calcification [6]

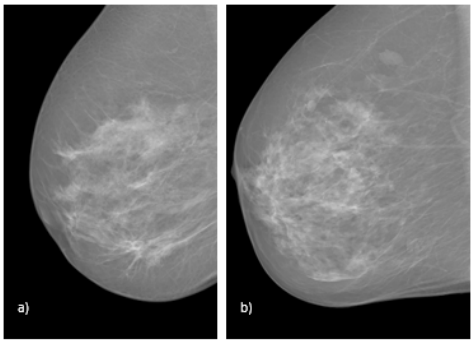


Figure. 2.10: Mammogram examples with and without distortions: a) distortions; b) no distortions [6]

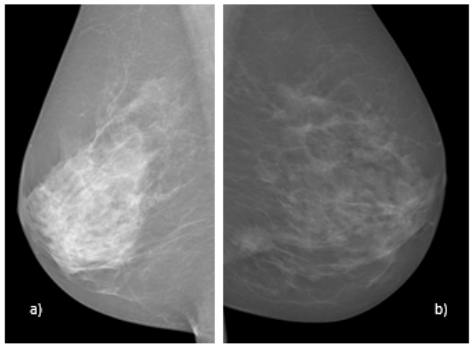


Figure. 2.11: Mammogram examples of different lateralities: a) Left breast; b) Right breast [6]

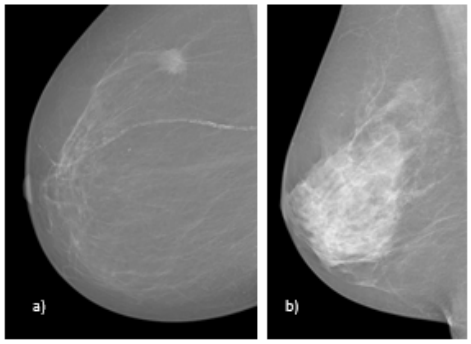


Figure. 2.12: Mammogram examples with and without masses: a) masses; b) no masses [6]

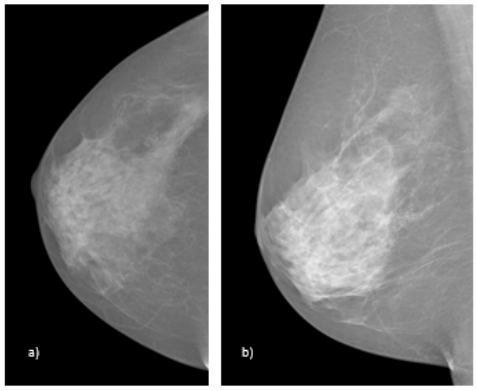


Figure. 2.13: Mammogram examples of different views: a) CC; b) MLO [6]

Mammography involves recording two incidences for each breast [6]:

- the CC of the right breast (Figure 2.14(a)) and the left breast (Figure 2.14(b)), which is done from top to bottom;
- the MLO of the right breast (Figure 2.14(c)) and the left breast (Figure 2.14(d)), which is done laterally.

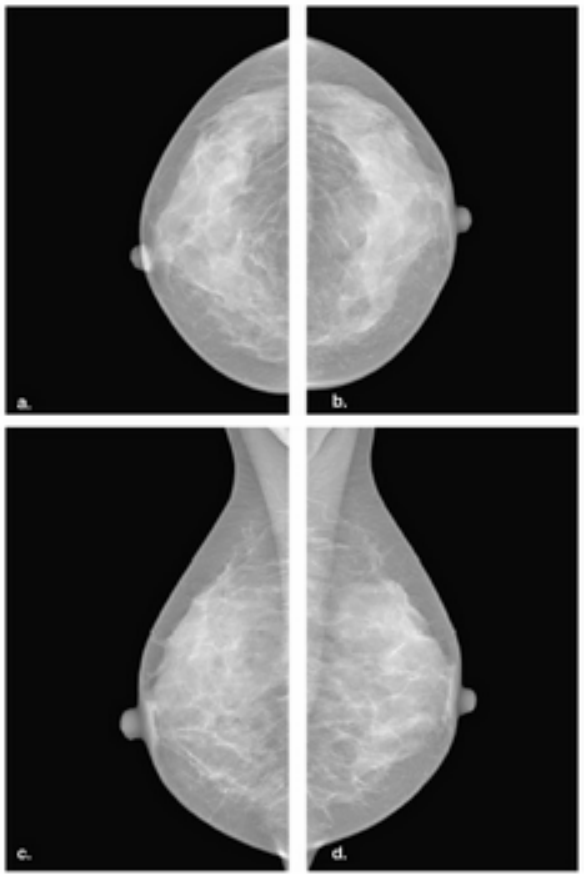


Figure. 2.14: Examples of mammograms [6]

2.3 Models

Several pre-trained networks have been used throughout this dissertation, so it is necessary to know a little more about them.

Recently, there has been a remarkable rise in machine learning applications focused on image recognition [11]. These applications play key roles in the creation of vision systems for autonomous vehicles, the detection of anomalies in medical images and even the ability to distinguish between different breeds of dog [11]. Such image classifiers are commonly built on the architecture of Convolutional Neural Networks (CNNs), known more informally as CNNs [11].

There are various types of neural networks, which are used for various use cases and types of data [12]. Residual Network 50 (RESNET-50) is a CNN and, as its name suggests, contains 50 layers. It is possible to load the pre-trained network onto more than a million images from the ImageNet database, thus allowing it to classify images from more than 1000 categories. This network learns very rich features from the images it classifies. It has an input image size of 224 by 224 [13]. Visual Geometry Group 16 (AVGG16) is a CNN that supports 16 layers. It can achieve a test accuracy of 92.7% on a dataset of over 14 million images in 1000 object classes [14]. InceptionV3 is a CNN that supports 48 layers, helping to analyse images and detect objects. In addition, by loading the pre-trained version of the network, it is possible to classify images into 1000 object categories. Finally, it has an input image size of 299 by 299 [15]. InceptionResnetV2 is a CNN that supports 164 depth layers and is trained on one million images from ImageNet. It can classify images into 1000 different categories. Presently, it has an input image size of 299 by 299 [16].

A paper (see Annex E - Industrial Seminars paper) was written as part of the Industrial Seminars course in the 2022/2023 academic year, where it was noted that there are several studies that choose to use different algorithms to classify breast cancer. The Support vector machine (SVM) algorithm is one of the best-known classification algorithms and one of the most widely used in breast cancer diagnosis, as it performs well in the vast majority of cases. It is an algorithm [3] that aims, in an N-dimensional space (N being the number of features), to find a hyperplane that classifies the data points, this hyperplane being a line. The "Kernel Trick" [17] is a strategy that enables the SVM to recognise a hyperplane capable of separating data with the largest possible margin, even when this data cannot be separated linearly in its original space. The support vectors [4] are the coordinates of each type of data, i.e. SVM can be considered to be the source between the two classes of data using the hyperplane. As for the advantages [4] of this algorithm, it is a generalised model because it has low risk of overfitting and can still generate accurate classification results due to its robustness. The disadvantages [4] are that the training time is long when using a relatively large data set and that it is dif-

ficult to interpret the model and its impact. The K-nearest neighbours algorithm[5], also known as KNN or k-NN, is an algorithm that is most often used for classification, although it can also be used for regression. The aim of this algorithm is to identify the closest neighbours points to the query point, so that a certain class can be assigned to that point. The k value defines the number of neighbours needed to classify a given point, and the choice of k value will depend on the amount of data. As for the advantages of this algorithm [6], it is very easy to implement and it is also considered a lazy learning algorithm because it does not require any training before making predictions, which means it is much faster than the SVM, which requires training. As for the disadvantages of the algorithm [6], they include the high cost of predicting large data sets and the low performance of categorical variables. The K-means algorithm [7] is a clustering algorithm which, within a data set, groups similar objects into clusters so that they are distinct from other clusters, reducing the distances between points. It can be considered as the discovery of clusters that do not have a label. The advantages of this algorithm [8] are that objects are automatically assigned to a group and the location of the group in space can vary, which means that initial dependency conditions must be established. As for the disadvantages [8], it is considered an algorithm that initially has to decide how many groups are needed and also the fact that the objects must necessarily belong to a group.

2.4 Summary

It is important to highlight the importance of mammography in the early detection of breast cancer, allowing patients to lead a normal life. Despite advances, the high mortality rate drives research into the disease. Breast cancer is characterised by the uncontrolled growth of breast tissue and can be invasive, with the potential to spread. The BI-RADS classification is introduced, classifying mammogram results into six levels. In suspicious cases (categories 4 and 5), a biopsy is essential for confirmation. The section on models highlights neural networks such as RESNET-50, AVGG16 and InceptionV3, used in image recognition. Algorithms such as SVM, KNN and K-means are also discussed, each with their own distinct characteristics. In summary, the chapter covers breast cancer detection, its nature and the use of models and algorithms in classification studies, providing a comprehensive and crucial overview for understanding this disease.

3 State of the art

This section is divided into two sections: the first highlights studies related to mammogram retrieval and consequently reviews the literature; the second section discusses software that competes with the software developed.

3.1 Retrieval

Initially, two studies provided by the supervisor of this project were considered, which would contribute greatly to the development of the literature review. Subsequently, in order to find the best studies developed, a search was made in the search engine "*Google Scholar*" with the phrase "intitle:Retrieval + intitle:mammography", obtaining a set of 43 results. The most relevant studies were then selected (see Annex A - Selection of papers). Finally, only studies with ten or more citations were selected. In order to complement the state of the art, 7 papers were selected, which were used in the preparation of the paper for the industrial seminars course in the current academic year 2022/2023, leaving a total of 14 studies, briefly described next.

In [18], an optimised classifier was used to develop a method for retrieving mammograms. The data used comes from the Mini-MIAS ('Mammography Image Analysis Society') database extracting 216 images, of which 51 are malignant, 63 benign and 102 are normal images. The proposed work uses a new Changed Weiner Channel (MWF) filter during preprocessing. A collection of features is extracted from the Regions of Interested (ROI). In particular, 12 features are extracted during feature extraction. The classifiers used were optimised MANFIS, optimised by Artificial Bee Colony (ABC) algorithm, ANFIS classifier, Artificial Neural Network (ANN), Naive Bayes (NB) and SVM. The best result was obtained by ABC-MANFIS with an accuracy of 0.98148. This method was developed using MATLAB software. The stated study is significant because it exhibits a very high accuracy but has not been put to the test in a clinical setting.

In [19], a Computer-Aided Diagnostics (CAD) system was developed to support breast cancer diagnosis. This system is based on retrieving images of relevant mammograms based on the respective pathology and also uses the retrieved cases to improve the performance of the classifier. The data used was the Digital Database for Screening Mammography (DDSM), developed by Massachusetts General Hospital, University of South Florida, consisting of 4300 breast cancer images. The decision support system

exploits the added value of local feature descriptors over global feature descriptors. To represent an image through global features, a single descriptor was required, while to describe an image locally, several descriptors were needed. These include Scale-invariant Feature Transform (SIFT), Speeded Up Robust Feature (SURF) and Local Binary Pattern (LBP). A weighted KNN was used as classifier. In terms of results, LBP outperformed SIFT and SURF, with a accuracy of approximately 0.9. This paper shows that CADs using Content-based image retrieval (CBIR) improve decision making, moreover, the recovery results are promising when compared to existing techniques.

In [20], a method for diagnosis and retrieval of breast masses was developed, since, according to the authors, the existing methods fall far short in the retrieval stage. The dataset was the DDSM, from which 11553 ROIs were extracted. To determine if there are any masses the retrieved ROIs are used, and a process that is performed by the SIFT performs the extraction and search of the features of the mammography ROIs. The experiment used the k-means classification technique which allows an accuracy of 90.8.

In [21], the goal was to introduce a method for assessing the risk of breast cancer malignancy in mammograms with microcalcifications (MCs), and to perform CBIR of the mammograms. The dataset used is the DDSM consisting of 87 mammograms with MCs. The evaluation method took into account the morphology of the MCs defined in the BI-RADS scale, and used shape and texture features as input to the SVM classifier in order to perform image retrieval. Given the experimental results, it was possible to verify that the retrieval was done effectively by the CBIR scheme with a mean average precision (MAP) of 0.60. The paper is interesting because it proposes a method for evaluating the risk of cancer based on the BI-RADS classifications, which serve as a standard reference system. One of the weaknesses of this approach is the low performance of the SVM. According to the authors, if the input features were customised, better retrieval performance could possibly be obtained for each BI-RADS.

A mammography ontology notation that enables mammograms to be interpreted from specialised vocabulary was created in [22]. A new dataset Dokuz Eylul University Mammogram Set (DEMS) was proposed, which contains digital mammograms that were created due to the development of a new ontology-based mammogram annotation and retrieval tool (MART). A CBIR system was developed which allows huge amounts of images to be searched from a given query based on visual similarity. In order to test the system a series of queries were executed in DEMS. A mathematical model was developed in order to calculate the similarity between two breast lesions and was further developed using Semantic Query-Enhanced Web Rule Language (SQWRL) and XQuery. The state of the art of this thesis is based on crisp logic, i.e. it is not able to face certain conditions, for this a new approach with the goal of modelling uncertainty was presented. This approach to model uncertainty aims to deduce BI-

RADS scores on mammograms and for this purpose experiments were done with two datasets, DEMS and DDSM. This method was evaluated with the NB classifier, using Rapid Miner, and the results showed accuracy and sensitivity values for the DEMS dataset of 0.86 and 0.60, respectively. For the DDSM dataset, the values of accuracy and sensitivity are 0.81 and 0.5, respectively.

In [23], an integration method for CBIR was developed, and an integrated approach is also implemented using PostgreSQL, an open-source Database Management System (DBMS), used in the mammogram retrieval system as a case study. The new method called breast contour segmentation allows the segmentation of breast mass in order to determine their precise contours. An ontology-based annotation tool was designed. A series of experiments were conducted on two different mammogram datasets DDSM and DEMS. The DDSM dataset contains digitised mammography films, while the DEMS dataset contains only digital mammography images. The experiments aimed to determine the best low-level feature, Machine Learning, and region selection methods for breast masses. The main groups of features were intensity, shape, texture and margin. Intensity represents the average grey colour level of a mass. The shape allows the identification of the magnitude of the mass. The texture contains information about the spatial distribution and variation of grey levels. Finally, the margin allows to identify the malignancy of a mass. According to the results of the low-level performance evaluation, it was possible to conclude that shape and texture features are the best groups of features to be used to identify properties in masses. The classifiers used in the experiments were K-NN, Random Forest (RF), NB, ANN, Linear Discriminant Analysis (LDA) and SVM. The NB classifier was the most successful classifier according to the experimental results.

Since CBIR is a promising tool for CAD, the objective of [24] was to build an SVM-based suspicious tissue pattern classification algorithm. The dataset consists of medical images taken from a total of 10,509 radiographs. Two experiments were carried out, splitting the data into 12 classes and into 20 classes In the 12-class problem, three pathology and four tissue density categories were considered. In the 20-class problem, two categories of distinct types of lesions were to be distinguished. Two-dimensional principal component analysis (2DPCA) was used to extract the features. The features are extracted from small patches of 128 x 128 pixels and classified by an SVM. As for the results of the experiments, an accuracy of 61.6% was obtained for the 12-class problem and 52.1% for the 20-class problem.

In [25] a CAD on a screening mammogram was developed. The CBIR methodology is used for tissue and lesion classification. The dataset combines a number of open and research-friendly datasets. A total of 10,509 mammography images were gathered from a number of archives, including Lawrence Livermore National Laboratory, DDSM, and the Mammographic Image Analysis Society (MIAS), Department of Radi-

ology, Rheinisch-Westfälische Technische Hochschule (RWTH), Lawrence Livermore National Laboratory(LLNL), Aachen University, Germany. If an equal distribution is necessary, this data supports experiments with up to 12 classes and 233 images per class. In terms of feature extraction, the Singular Value Decomposition (SVD) method, 2DPCA, and principle component analysis (PCA) were employed to characterise breast density and lesions. The density of breast tissue, with or without a lesion can be characterised by brightness and texture attributes. The features extracted using 2DPCA can represent the texture through a low-dimensional feature vector. As PCA and SVD were already known to be capable of representing texture and reducing the dimensionality of the feature vector, their performance was compared to that of the 2DPCA technique for the characterisation of breast and lesions. For the task of image retrieval, SVM was the chosen classifier. The 2DPCA technique achieved the best accuracy of 80.07%. This study can help CBIR-CAD of mammograms, offering a system that can help radiologists in diagnosis or as a pre-processing step of CAD systems to classify breast lesions.

A CBIR system named MammoSVx is introduced in [26] and supports the categorisation of breast tissue density. It may also be used to modify parameters for lesion segmentation and classification during the processing phase. The IRMA reference dataset consists of data from several datasets that are used to answer user queries, namely the DDSM, MIAS, LLNL and RWTH datasets. The reference dataset consists of over 10000 distinct mammograms. The feature extraction and selection process is done by SVD and histograms, allowing the breast density to be characterised by the image texture. The SVM determines the similarity of patterns in order to divide the BI-RADS tissue into four categories. As for the results of the experiments, an average accuracy of 82.14% was obtained. It was possible to conclude that the development of a CBIR system with SVD associated with SVM for breast density characterisation for image retrieval can help radiologists in their diagnosis.

In [27], a method based on image retrieval was conceived in order to develop a case-adaptive classifiers design in CAD. The dataset used is the DDSM which consists of 589 mammogram images, of which 331 are benign breast cancer and 258 are malignant. Image retrieval is initially performed by obtaining a set of images with characteristics relevant to the diagnostic case. Later these cases are used to optimise a classifier, with the goal of increasing the classification accuracy of the breast cancer case. A linear classifier, using logistic regression, and a non-linear SVM classifier were used. The non-linear SVM classifier obtained the best result with an accuracy of 0.8. The paper is quite realistic in that the features of the images are extracted from an algorithm that quantifies the lesions, instead of this detection being done manually.

In [28] the aim is to present a number of approaches from previous studies and compare them in order to identify which approaches best apply to CBIR-based CAD sche-

mes. Previous studies have suggested that the use of CBIR-based CADs can greatly improve or increase the performance and confidence of radiologists. The dataset presented in this paper consists of image data collected from clinical practices at the Radiology Department of one University Hospital, consisting of 1,500 reference images (or ROIs) of verified masses and 1,500 ROIs of false-positive masses (negative breast tissue). To evaluate the performance of CAD algorithms using CBIR approaches, evaluation indices such as visual similarity and clinical relevance were used. According to the previous studies discussed by the paper, commercially available CADs, perform very high in both microcalcifications and clusters detection, however, perform significantly worse in mass detection. Furthermore, the "visual aid" in CAD applications can be useful for radiologists. CAD development using CBIR approaches entails a number of problems such as feature selection, accurate region segmentation, and assembling the optimal reference dataset. In short, more research is still needed for CAD development.

The literature on advanced CAD system techniques, including mass detection, calcification, bilateral asymmetry, architectural distortion, and image retrieval, is thoroughly reviewed by Tang et al. up to 2009 [29]. The two most popular algorithms among the numerous works referenced are SVM and KNN. The authors emphasised the value of CAD systems and drew attention to the need to increase their performance in order to support clinical applications.

In [30], the development and evaluation of a CBIR system system for mammography are described. The dataset used is from the Department of Radiology at the University of Chicago and consists of 200 mammogram images, of which 58 images are of benign breast cancer and 46 images of malignant breast cancer. In this study, the SVM classifier and an Adaptive SVM (Ada-SVM) were used, and a comparison is made between them, as to the use of the discriminant adaptive nearest neighbours (DANN), which was used to do the retrieval of the relevant images. Finally, it was proven that Ada-SVM got the best accuracy with 0.7925 (Ada-SVM with N=9). The paper is quite interesting since it led to a reduced generalisation error. One of the weak points is the fact that the approach did not go through clinical evaluations, which would have been an added value in order to test this approach.

The goal of information retrieval is to extract from a database information that is pertinent to the user's query. The most challenging aspect of CBIR is defining the relevance or similarity, which guides the retrieval process. In [31], a novel approach was pursued, where similarity is acquired through training examples provided by human observers. The authors investigate the use of NN and SVM to predict the user's notion of similarity. The dataset was collected by the Department of Radiology at the University of Chicago and consisted of 76 mammograms where all of which contained clustered MCs and the goal would be to retrieve these mammograms relevant to those

in a query. The performance of the retrieval system is evaluated using precision-recall curves, calculated using a cross-validation procedure. The proposed hierarchical learning approach comprises a sequence of a binary classifier and a regression module, aiming to optimise both the effectiveness and efficiency of retrieval. The first phase works as a screening procedure where images that are very different from the query image are eliminated. then, the second phase allows the images from the previous phase to be compared with a query, with the goal of obtaining a number for retrieval, i.e. quantitatively measuring the similarity between the "surviving" image and the query. Four combinations were evaluated: linear Fisher discriminant and SVM (Fisher-SVM); linear SVM and SVM (SVM-SVM); a linear SVM with modified objective function (MSVM) and SVM (MSVM-SVM); and linear SVM with modified objective function and a General Regression Neural Network (GRNN) (MSVM-GRNN). The obtained results demonstrated that all learning-based networks outperform those based on the Euclidian distance, with the two-stage networks MSVM-SVM achieving the greatest performance.

Table 3.1 presents a brief summary of each of the papers analysed in the literature, in terms of the objectives of each approach, the algorithms or methods used, the datasets used and the results obtained.

3.2 Available Mammogram Retrieval Software

The retrieval of relevant mammograms is the ultimate goal of the software to be developed. Table 3.2 shows some competing software. In contrast to the software provided in Table 3.2, the created software thus far allows experts to submit ground truth data for the later building of retrieval models. To the best of our knowledge, no tools are currently available for this purpose.

3.3 Summary

The studies address various methodologies for retrieving mammograms and diagnosing breast cancer, using datasets such as Mini-MIAS, DDSM, and DEMS. Different classification techniques, such as classifier optimization, CAD, and CBIR systems, have been applied. The studies reveal significant advances in retrieving mammograms and detecting anomalies such as microcalcifications. In addition, the introduction of ontologies, ontology-based annotation tools, and BI-RADS are explored to improve the interpretation and assessment of cancer risk. Results vary in terms of accuracy, sensitivity and effectiveness, highlighting the importance of approaches such as SVM, KNN, and machine learning techniques in obtaining promising results. Some studies emphasize the need for clinical evaluations to validate their approaches. In summary, the studies

Table. 3.1: Literature review of retrieval works in breast cancer

Paper	Main Goal	Methods	Database	Main Results
[18]	Retrieval	ABC-MANFIS, ANFIS, ANN, NB, SVM	Mini-MIAS	ABC-MANFIS with accuracy of 0.98148
[19]	CAD based on CBIR	SIFT, SURF, and LBP descriptors	DDSM	Accuracy ≈ 0.9 using LBP
[20]	CAD based on CBIR	SIFT, k-means	DDSM	Accuracy of 90.8% with k=5
[21]	CBIR	SVM with shape and texture features	DDSM	Average accuracy $[0.57 - 0.60]$
[22]	Ontology-based annotation; CBIR	NB	DEMS and DDSM	Accuracy of 0.86 for DEMS and 0.81 for DDSM
[23]	CBIR	Mass segmentation, 26 low-level features, 6 classifiers	DDSM and DEMS	Best classifier: NB
[24]	CAD based on CBIR	2DPCA, SVM	Merged public datasets	Accuracy 61.6% (12-class); 52.1% (20-class)
[25]	CAD based on CBIR	2DPCA, SVD, SVM	Merged public datasets	80.07% using 2DPCA
[26]	CBIR for tissue density	SVD feature extraction, SVM	IRMA	Average accuracy of 82.14%
[27]	CAD	Linear and non-linear SVM	DDSM	Non-linear SVM with accuracy of 0.8
[28]	CBIR-based CAD comparison	n.a.	n.a.	CAD development needs further research
[29]	Literature review of CAD techniques	n.a.	n.a.	CAD performance needs improvement
[30]	CBIR	SVM and Ada-SVM	Chicago dataset	Accuracy of 0.7925
[31]	CBIR	ANN and SVM	Chicago dataset	Similarity learned from human observers

presented contribute to the development of mammography retrieval systems, offering valuable knowledge to improve diagnosis and clinical support.

Compared to the competing software presented in Table3.2, the created software allows experts to provide fundamental data for the subsequent construction of retrieval models, a unique capability to date. The table highlights PenFetch and a mammography

Table. 3.2: Competing Software

	PenFetch [32]	**Retrieval system for mammogram images [33]**
Capabilities	Automates retrieval of previous medical studies from PACS	Users can send RF to improve the system's understanding of information needs
Methods	Uses DICOM worklists to identify patients and images, requesting transfers based on predefined rules	Uses text-based methods for the initial search and augments the results with visual aids
Advantages	Simplifies workflow by sending relevant images to radiologists' stations in advance, minimising delays	Applications include computer-aided diagnosis, medical education and research
Disadvantages	Dependence on DICOM worklists. Limited to PACS environments	Specific focus on certain domains may limit wider applicability
User needs	When a radiology study is opened for reading and the relevant previous images are not available, it is necessary to wait for these images to be retrieved from the PACS to the workstation, resulting in wasted time and network resources, which delays the execution of the report. In addition, time is wasted in the imaging room as technicians wait for the previous images to appear on their workstations. All this and much more can be automated with PenFetch	The application of visual elements faces specific challenges due to the so-called "semantic gap". This gap refers to the difference between the information that can be readily extracted from images and the more abstract descriptions that are meaningful to users. In other words, there is a discrepancy between the immediate visual details and the higher-level interpretations that are relevant to people.

PACS: Picture Archiving and Communication System
RF: Relevance Feedback
DICOM: Digital Imaging and Communications in Medicine

image retrieval system, including their capabilities, methods, advantages and disadvantages. PenFetch automates the retrieval of previous medical studies from Picture Archiving and Communication System (PACS), while the alternative system allows users to submit Relevance Feedback (RF) to enhance the system's understanding of information needs.

4 Methods

This section presents the methodology adopted in the course of this project.
The methods applied in each iteration of the development of this work consist of three main steps (see Figure 4.1). The first step involves building several models in Python language using four different pre-trained networks, which allow the classification of mammograms according to different dimensions, namely, the existence of masses, the existence of calcifications, the existence of distortions, breast density, BI-RADS classification, laterality, view and the existence of asymmetries In the second step, the model that classified mammograms most accurately based on the various dimensions is selected. The third and final phase involves creating the final retrieval model using the previously built models; more specifically, the final model is created by combining these models using a weighted sum (Section 4.2). This section also includes the description of the experimental set-up (Section 4.2) and the used database (Section 4.1).

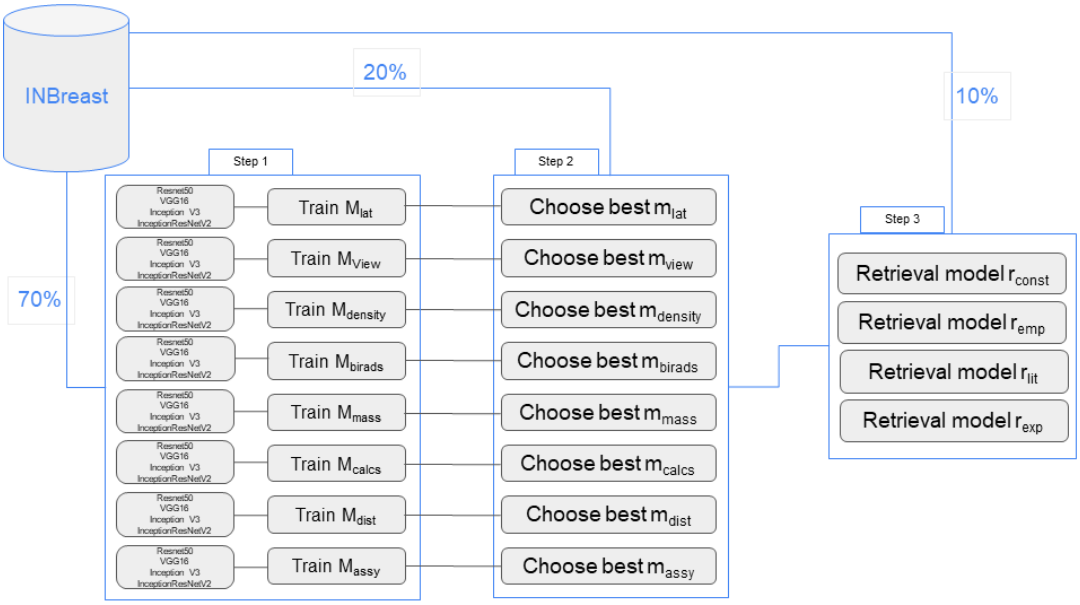


Figure. 4.1: Overview of the training methodology

4.1 Database

The INBreast database was acquired at the Breast Centre of the CHSJ, Porto, under the authorisation of the Hospital's Ethics Committee and the National Data Protection Commission [6, 34]. The images have a pixel size of 70 µm (microns) and a contrast re-

solution of 14 bits [6]. The image matrix was 3328 4084 or 2560 3328 pixels. They were saved in the Digital Imaging and Communications in Medicine (DICOM) format [6]. All confidential medical information was removed from the DICOM file [6]. The correspondence between images of the same patient is maintained with a randomly generated patient ID [6]. INBreast has full-field digital mammography (FFDM) images of screening, diagnostic and follow-up cases. Diagnosis is made when screening shows signs of an abnormality [6]. In the follow-up images, the cancer was previously detected and treated [6]. A total of 115 cases have been collected, of which 90 have two images (MLO and CC) of each breast and the remaining 25 cases are of women who have had a mastectomy and two images of just one breast have been included [6]. This adds up to a total of 410 images. Eight of the 91 cases with 2 images per breast also have images acquired at different times (follow-up) [6].

The database includes examples of normal mammograms, mammograms with masses, mammograms with calcifications, architectural distortions, asymmetries and images with multiple findings (Figure 4.2) [6].

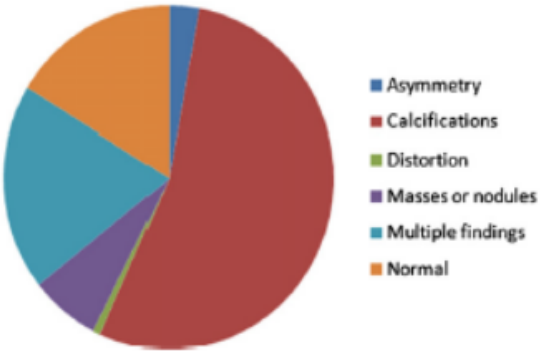
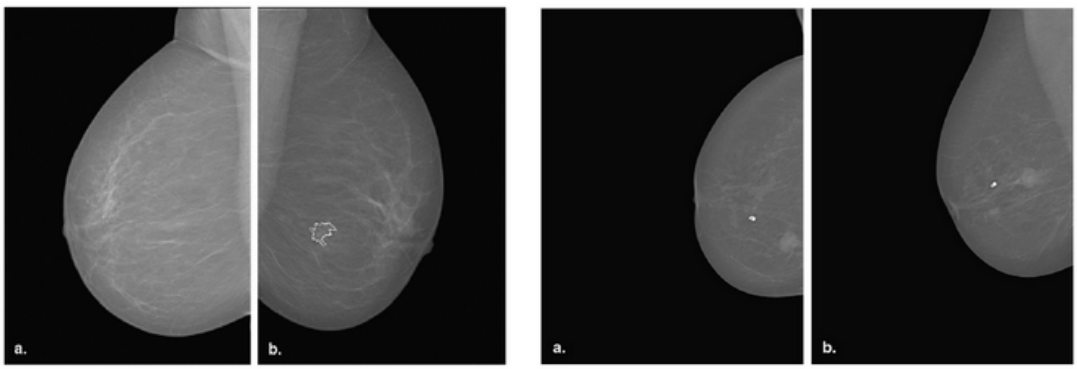


Figure. 4.2: Findings from the INBreast database [6]

The main feature of this dataset is the annotations. These were made by an expert in the field and validated by a second expert between April 2010 and December 2010 and were stored in XML [6]. They were made on OsiriX, an open source PACS workstation running on a Macintosh platform [6].

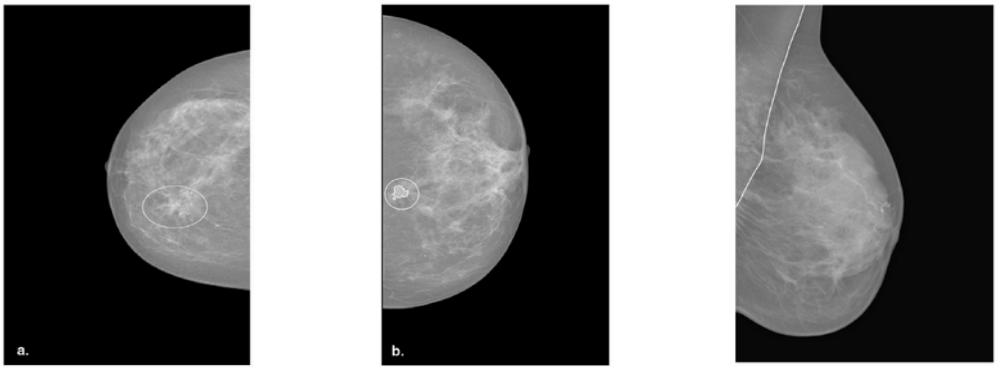
There are seven types of annotated dimensions:

- Asymmetry (see Figure 4.3a)
- Calcification (see Figure 4.3b)
- Cluster (see Figure 4.3c)
- Masses (see Figure 4.3d)
- Distortion (see Figure 4.3e)
- Spiculated region (see Figure 4.3f)
- Pectoral muscle (see Figure 4.3g)



(a) Asymmetry [6] (b) Calcification [6]

(c) Cluster [6] (d) Masses [6]



(e) Distortion [6] (f) Spiculated region [6] (g) Pectoral muscle [6]

Figure. 4.3: Annotated dimensions available in InBreast

Information is also provided on the patient's age at the time of image acquisition, family history, ACR breast density annotation and BI-RADS classification. The distribution of the BI-RADS classification in the database is shown in Figure 4.4 [6].

4.2 Models

This section highlights all the classifiers used to create the models and how the final retrieval model was built.

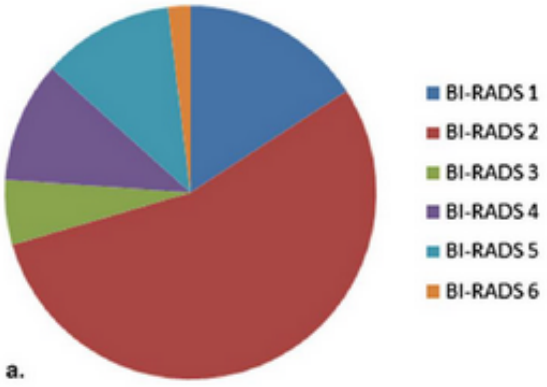


Figure. 4.4: BI-RADS classification [6]

4.2.1 Classifiers

Four pre-trained networks were tested, ResNet50, VGG16, InceptionV3 and Inception-ResNetV2. These four types of pre-trained networks were chosen because they were the most interesting and there was already some prior knowledge of them. Thus, four models were created for each dimension (masses, calcifications, distortions, breast density, BI-RADS, laterality, view and asymmetries), each with a different pre-trained network. In total, 32 models were created.

Initially, the preprocessing of the images is performed, the image is resized to 224×224 pixels and converted from RGB to BGR. Next, each colour channel is zero-centred concerning the ImageNet dataset, without scaling. Two approaches are adopted in each model:

- In the first approach all layers of the network are frozen, except for the final fully connected layers. By keeping the pre-trained weights intact and only fine-tuning the final layers, the learned features are leveraged, while adapting the model to each specific task;
- In the second approach the entire model is trained with a low learning rate. By allowing the entire network to be updated, the potential for further improvements over the previous approach is enabled.

The models apply the Adam optimiser and Sparse Categorical Cross-entropy loss. After that, the models are compiled and trained for 25 epochs.

4.2.2 Retrieval Model

The retrieval model leverages the individual models described in Section 4.2.1. It is a weighted sum and has the following generic equation:

$$d^{probe,i} = w_{\,lat} * d^{\,probe,i}_{lat} + w_{\,view} * d^{\,probe,i}_{view} + w_{\,density} * d^{\,probe,i}_{density} + w_{\,birads} * d^{\,probe,i}_{birads} + w_{\,mass} * d^{\,probe,i}_{mass} + w_{\,calcs} * d^{\,probe,i}_{calcs} + w_{\,dist} * d^{\,probe,i}_{dist} + w_{\,assy} * d^{\,probe,i}_{assy},$$

$i = 1 : 410 \backslash \{probe\}$

where w_{dim} is the weight attributed for each dimension (laterality, CC or MLO view, breast density, BI-RADS class, existence or not of masses, existence or not of calcifications, existence or not of distortions, and existence or not of asymmetry).

Different distance metrics can be chosen to be included in the above formula. The present implementation behaves accordingly to:

$$d_{dim}^{probe,i} = \begin{cases} 1, if\ \tilde{V}_{dim}^{probe} == V_{dim}^{i} \\ 0, else \end{cases}$$

where dim is each of the considered targets (dimensions), V_{dim}^{i} is the ground truth value of the attribute as given in the INBreast database (see Section 4.1) for the image i and $\tilde{V}_{dim}^{probe}$ is the result of the classifier (see Section 4.2.1) when applied to the probe image.

The above distance is computed between the probe image and all of the 410 images in the INBreast dataset, excluding the probe image itself.

Section 6.3 shows four versions of the weights. For the r_{const} model, the weights have been set to a constant value. In the r_{emp} model, the weights are defined empirically. In the r_{lit} model, the weights are defined according to the available literature. Finally, in the r_{exp} model, the weights are defined according to the results obtained by the expert who collected the ground truth using the developed software.

4.2.3 Experimental Setting

In conducting this study, Tensorflow, version 2.11.0, was used. The programming environment for importing the libraries was Google Colab with the NVIDIA T4 Tensor Core GPU. Two approaches were taken, in which the data was divided into 70% for training, 20% for validation and 10% for testing. Two training methodologies were employed. In the first approach, images were randomly allocated across the training, validation, and test sets. In the second approach, having in mind that there may be more that one image per patient, and in order to avoid overfitting, the random assignment of the images was made in a way that images of the same patient are in the same set.

4.3 Summary

Initially, several models are built in Python using four different pre-trained networks (ResNet50, VGG16, InceptionV3 and InceptionResNetV2 resulting in 32 models, each for a specific dimension) to classify mammograms in relation to different dimensions. In the second stage, the best model (with the best accuracy) for each dimension is selected. The final stage involves creating the final retrieval model by combining the models

built previously, using a weighted sum. This chapter also covers the experimental setup, details of the INBreast database and a description of the models used. The models were trained in two approaches: one in which only the final layers were adjusted, keeping the pre-trained weights, and another in which the entire model was trained with a low learning rate. The programming environment included the use of TensorFlow 2.11.0 on Google Colab with an NVIDIA T4 Tensor Core GPU. Finally, two divisions of the data were carried out, one random and the other taking into account the possibility of multiple images per patient.

5 Software developed

This section describes the creation of an application designed to collect real data for the development of a mammogram retrieval system.

5.1 Outsystems

OutSystems [35], a Portuguese company founded in 2001, arose from the need to overcome challenges in software projects. Differentiating itself from the traditional "waterfall" approach, the company proposed the idea of making software changes fast and economical, regardless of the size of the application. This led to the development of the OutSystems platform, which enables agile, robust and cost-effective changes at any stage of the application life-cycle. This innovative approach aims to avoid the problems associated with trying to plan perfectly from the outset, providing flexibility and efficiency in software development. OutSystems is a contemporary application platform designed to significantly speed up the development of a company's most crucial applications, while providing unparalleled levels of flexibility and efficiency.

The software was created using the OutSystems platform due to the fact that it provided a fast software solution and also due to previous knowledge of this platform. The complete application is constructed within a unified development environment, employing a low-code approach. This includes the front-end, back-end, database and integration with pre-existing systems or services such as Simple Object Access Protocol (SOAP), Representational State Transfer (REST) and Systemanalyse Programmentwicklung (SAP).

The way the software works involves interaction between different components, as illustrated below.

Interaction with the user

The graphical [35] representation shown in Figure 5.1 shows the dynamic interaction of the user when making a request to the server. This figure highlights the flow of actions that takes place from the user's initial interaction with the application to the corresponding processing on the server during the request.

The user starts the process by opening their browser, where they:

1. Enters a Uniform Resource Locator (URL) or clicks on a link on the web page;

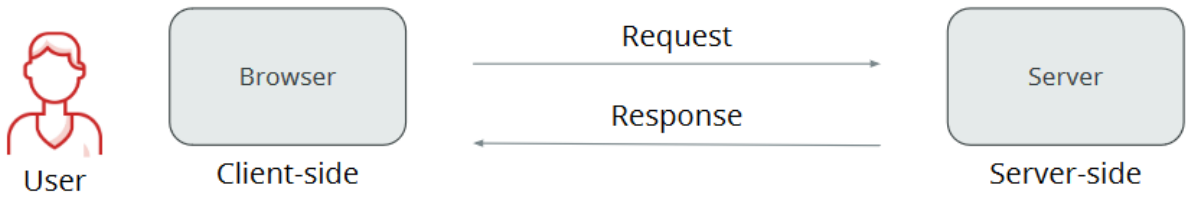


Figure. 5.1: User interaction [35]

2. The browser (client side) sends a request to the server;
3. The server (server side) responds to the request;
4. The browser receives the response and displays the web page to the user.

Web requests and response

The visual representation shown in Figure 5.2 outlines the interaction between the browser and the server, offering a graphical view of the communication between these two components [35].

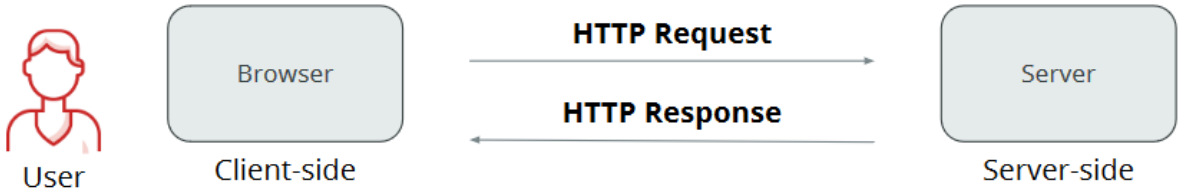


Figure. 5.2: WEB_Response_Requests [35]

Communication between the browser and the web server takes place via the Hypertext Transfer Protocol (HTTP) [35]:

1. The browser sends an HTTP Request when users interact with a web page, such as clicking on a link, submitting a form or performing a search.
2. The server waits for requests and responds by sending an HTTP Response back to the browser.
3. Successful HTTP responses contain the requested resource, such as an Hypertext Markup Language (HTML) page.

Server-side

Figure 5.3 represents the server's interaction in the process of responding to a request, visually outlining how the server behaves in this specific context [35].

The server carries out the following steps [35]:

- Listens to the incoming requests;

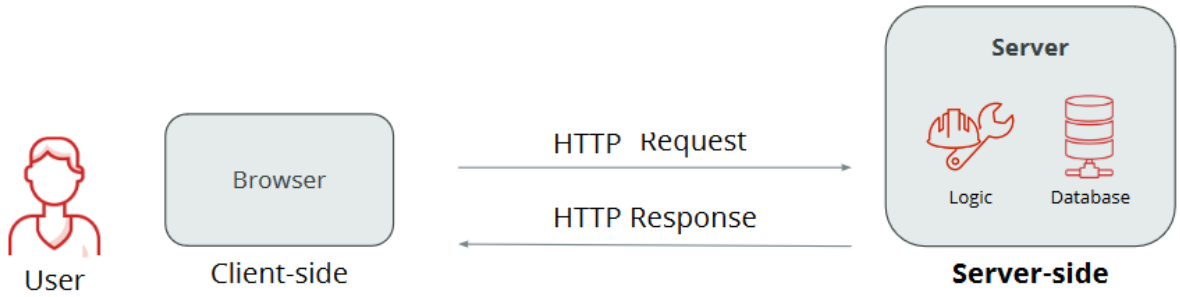


Figure. 5.3: Server-side [35]

- Processes the requests:
  - Retrieves necessary data and stores relevant information;
  - Controls access to data and personalises responses.
- Sends the response back to the browser:
  - Dynamically builds the HTML page;
  - Provides other resources needed for that HTML page.

Client-Side

Figure 5.4 exemplifies how the user interacts in the process of responding to a request, providing a visual representation of the specific dynamics of this involvement [35].

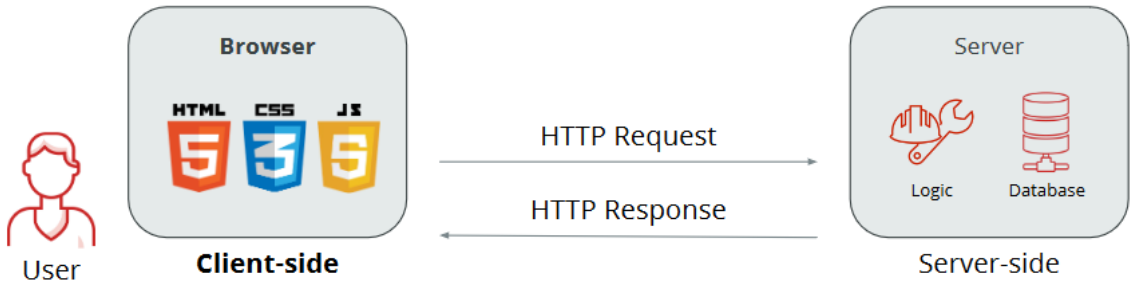


Figure. 5.4: Client-side [35]

On the client side:

- The code is executed in a browser;
- The browser renders the response to the end user, including:
  - Fundamental components of the web page (HTML);
  - Style of these components (Cascading Style Sheet (CSS));
  - Behaviour and interactivity of these components (JavaScript (JS)).

5.2 Requirements analysis

To be able to develop the application, an initial list of requirements was identified:

ID01- View probe image: The software must allow the visualisation of the probe image that will later be analysed and associated with a relevant image from the database.

ID02- View images in the database: The software shall allow the visualisation of the various images of the database, images that are tagged according to the several dimensions of the mammograms.

ID03- Filter images based on dimensions: The software should allow the filtering of the images in the database according to their dimensions such as breast density, existence or not of asymmetries, BI-RADS, existence or not of calcifications, existence or not of distortions, laterality (right or left breast), existence or not of masses and image incidence CC or MLO).

ID04- Select an image from the database: For each probe image, the software must allow the selection of one image from the database, the one considered by the expert as the most relevant.

ID05- Visualise registrations: The software must allow the visualisation of all the registrations made in the application.

During the development of the software, several improvements were suggested by the expert. These include:

ID06- The probe image cannot be associated with it itself:: The software must ensure that the probe image does not appear in the database images to be selected.

ID07- The probe image must always be visible: The software should allow the probe image to be visible at all times, i.e. it should always remain visible while selecting the image from the database.

ID08- Zoom for Enhanced Viewing: The software should allow zooming in on images, both probe images and database images.

ID09- Navigation to Next Probe Image: The software should allow to skip the current probe image and move on to the next probe image while still having access to the prior probe image.

ID10- Filter selection confirmation: The software should include a functionality to generate a warning, allowing for verification of whether the filter selection has been finalised or not.

The requirements are pictured in the Use Case Diagram of Figure 5.5.

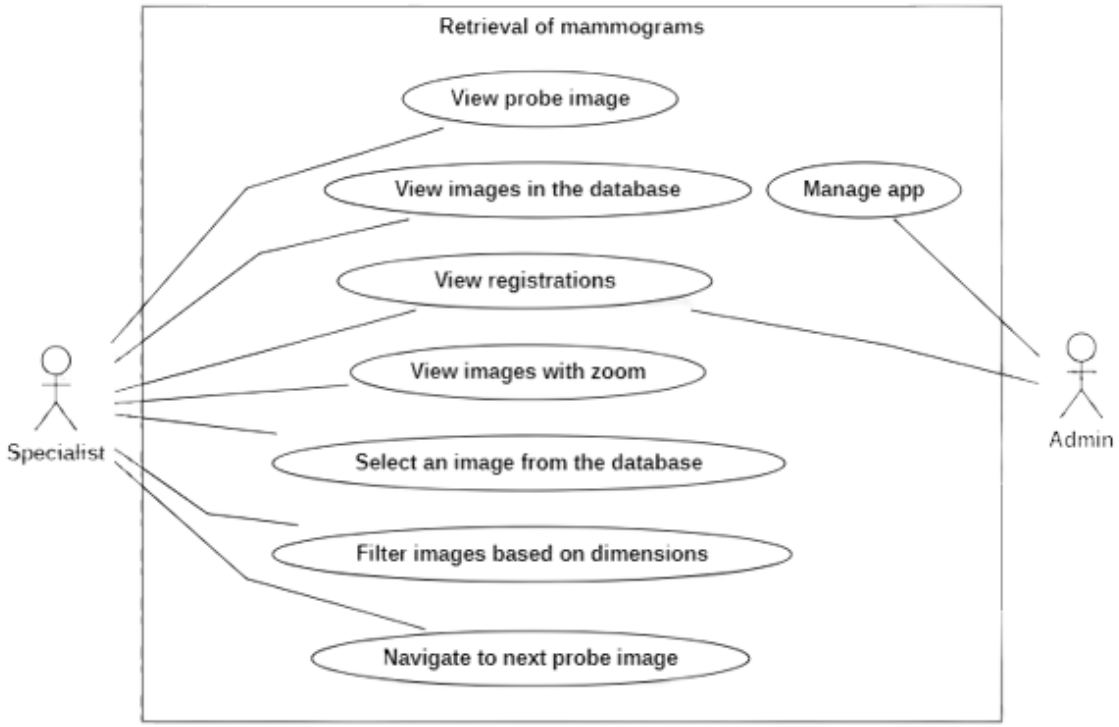


Figure. 5.5: Use case diagram

5.3 Application database

The data model, shown in Figure 5.6, was created to represent the information that is necessary for the application to meet its objectives.

In Outsystems, database tables are called Entities and Static Entities, so a database was built with three entities (Probe_Image, ImagesDataBase and User) and eight static entities (Distortions, ACR, BiRads, Mass, Calcification, Asymmetrys, View and Laterality). The Probe_Image entity represents all the probe images that will be associated with an image in the database. The ImagesDataBase entity represents all the images of the database, where only one of these, in each interaction, will be selected as the relevant image. The User entity is a default entity of the OutSystems platform and represents the registration of users on the platform. Static entities are entities that have a predefined set of possible values and in this context represent the dimensions of the mammograms. Static entities have a set of values ("Records") as shown in Figure 5.7.

5.4 Implementation

This section describes the entire implementation of the software. The complete instruction manual for the software can be found in Annex F - Software operating instructions.

The application starts with a login page (Figure 5.8) where the expert can fill in the

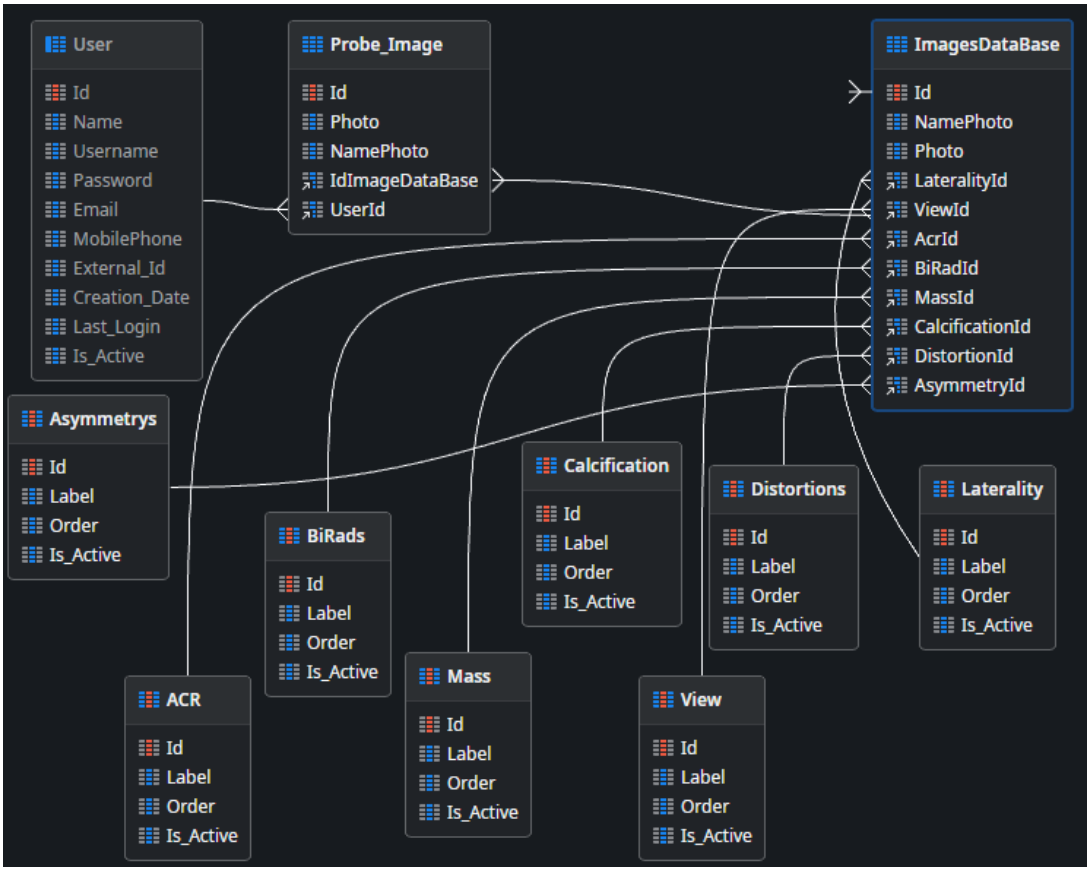


Figure. 5.6: Database

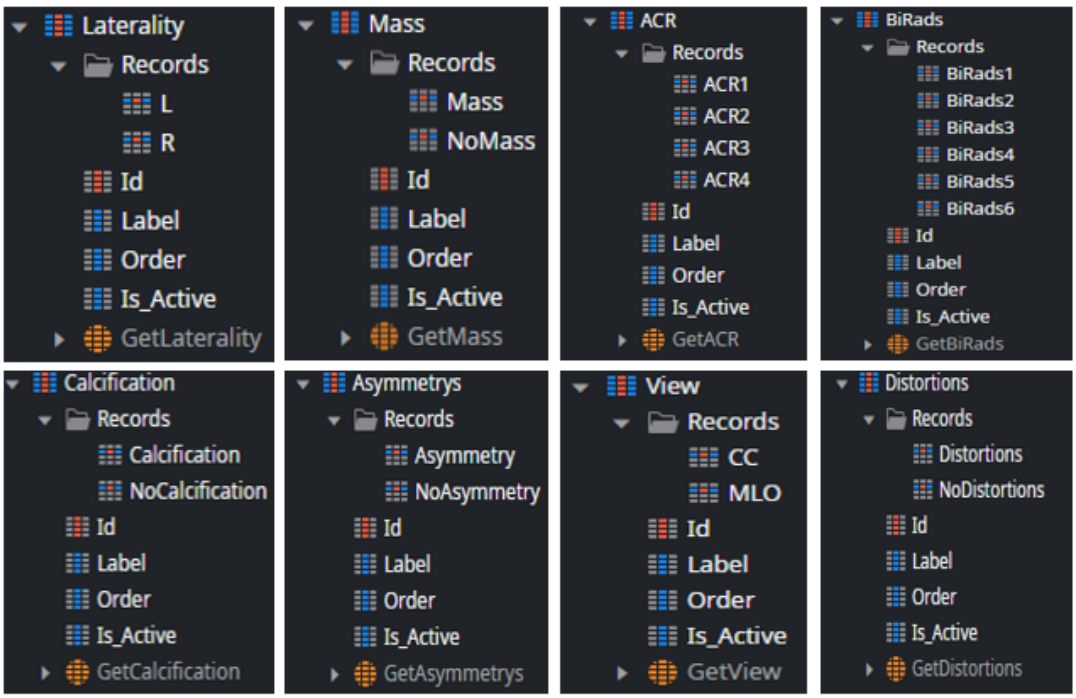


Figure. 5.7: Static Entities Records

form with their credentials to access the software.

Figure. 5.8: Login page

If the provided software access credentials are valid, the user will be directed to the homepage. There, the user only needs to click the "Start" button, which will then take them to a new page (Figure 5.9). This newly accessed page enables image retrieval, the specialist associates the probe image (highlighted in blue in Figure 5.9) with the most relevant image on the right (highlighted in orange in Figure 5.9). To assist the expert in their selection process, the images on the right (highlighted in orange in Figure 5.9) can be filtered using available filters (highlighted in green in Figure 5.9). These filters allow the selection of images with specific dimensions of the mammogram (as seen in Figure 5.7) exclusively. After conducting a thorough analysis, the expert can select the most relevant image by selecting the respective checkbox, followed by clicking the "Save" button to save the record in the database.

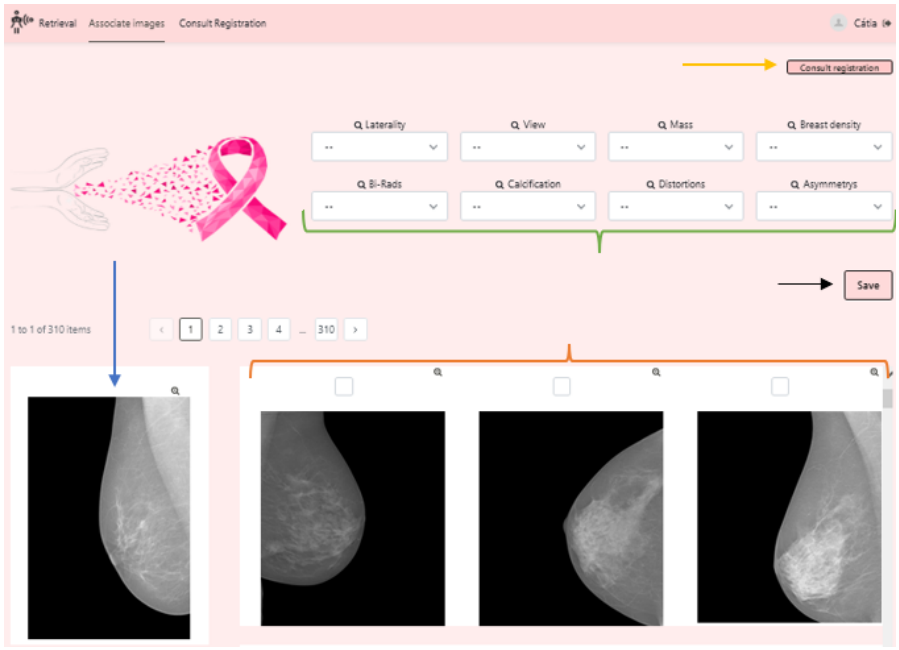


Figure. 5.9: Mammogram retrieval page

After the association procedure is done, the expert is led to a new page (see Figure 5.10), where all records made can be viewed in detail. The probe image and the image selected by the expert as most relevant, along with their dimensions, are displayed on this page.

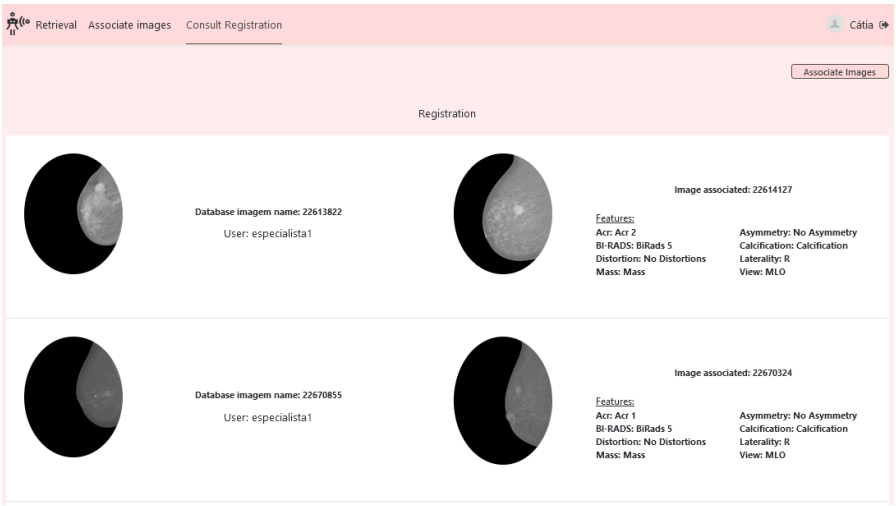


Figure. 5.10: Registry example

5.5 Ground truth input analysis

This section analyses the inputs provided by an specialist. After the software development was completed, it was necessary to collect ground truth using the software developed.

5.5.1 Analysis

When developing the software, 410 mammogram images from the INBreast database were added to the software so that they could be analysed and the most relevant image could be selected in relation to the probe image. In total, 102 records were made in the database, i.e. 102 images were analysed out of the 410 that existed. This number was low due to the limited capacity of the Outsystems platform database, since a free Outsystems platform plan was used, which meant that the software developed was slow in displaying the images, which meant that the specialist's work was not carried out under the best conditions. That said, it is necessary to analyse the ground truth provided by the specialist. Figure 5.11 shows an example of ground truth obtained by the expert.

Figure 5.12 shows the general distribution of the points on the graph, i.e. a view of how the similarities are distributed across the expert's inputs. It should be noted that there are a total of eight (8) dimensions for the mammograms (vertical axis), i.e. the following graph shows the count by input (line) of equal dimensions between the probe

Figure. 5.11: Ground Truth

image and the image chosen by the specialist, and this count can vary between 0 and 8 equal dimensions. In general, the graph shows an upward and downward trend and it is not possible to have just one trend, since the horizontal axis corresponds to the number of inputs that the expert made when collecting the ground truth. There is a variation in the similarity between the expert's inputs and those of the database. It is possible to see some patterns such as specific areas in the expert's inputs that indicate similarities (peaks) and differences (valleys). The graph clearly shows that there are never less than two dimensions in common and never more than six. When developing the software, images of the same patient were not excluded, i.e. when the specialist selects the most relevant image in relation to the probe image, it may happen that he selects an image of the same patient, a contralateral image, which consequently causes there to be at least one different characteristic.

According to Figure 5.13, the overall average distribution of the points in the graph is visible, i.e. a view of how the average similarities are distributed across the specialist's inputs. Compared to the average, which is the reference point, there are higher and lower points than the reference point. In short, it is possible to conclude that in the upper points there are more similarities relative to the average, and the lower points indicate that there are fewer similarities relative to the average.

According to Figure 5.14 it is possible to provide an overview of the distribution of the averages of equal values for each dimensions that the expert has chosen. It is possible to deduce that some columns (dimensions), for example Calcifications and Asymmetries, have consistently high bars, which indicates that the specialist is more likely to provide equal values in these dimensions.

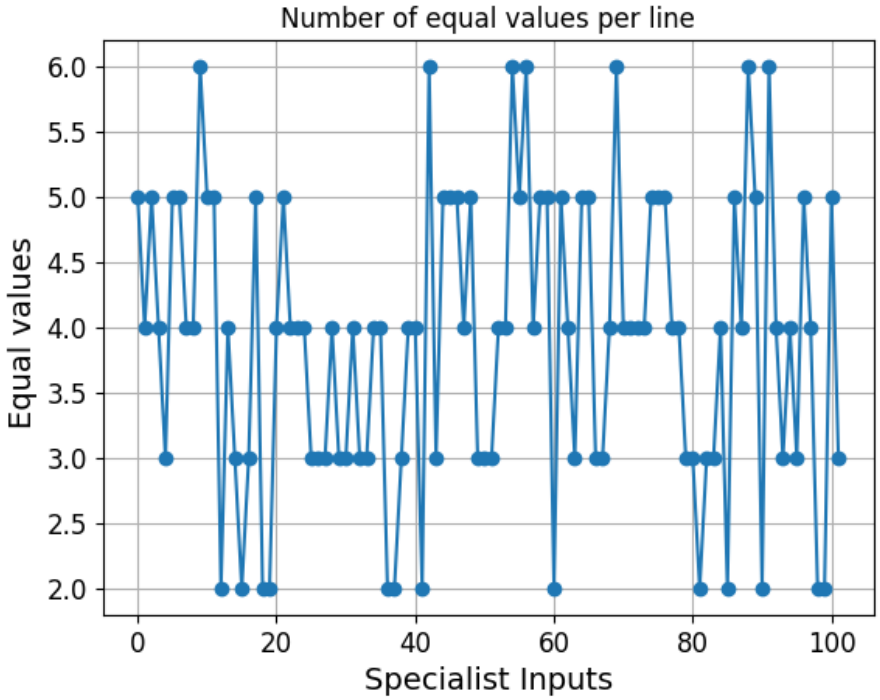


Figure. 5.12: Number of equal values per line

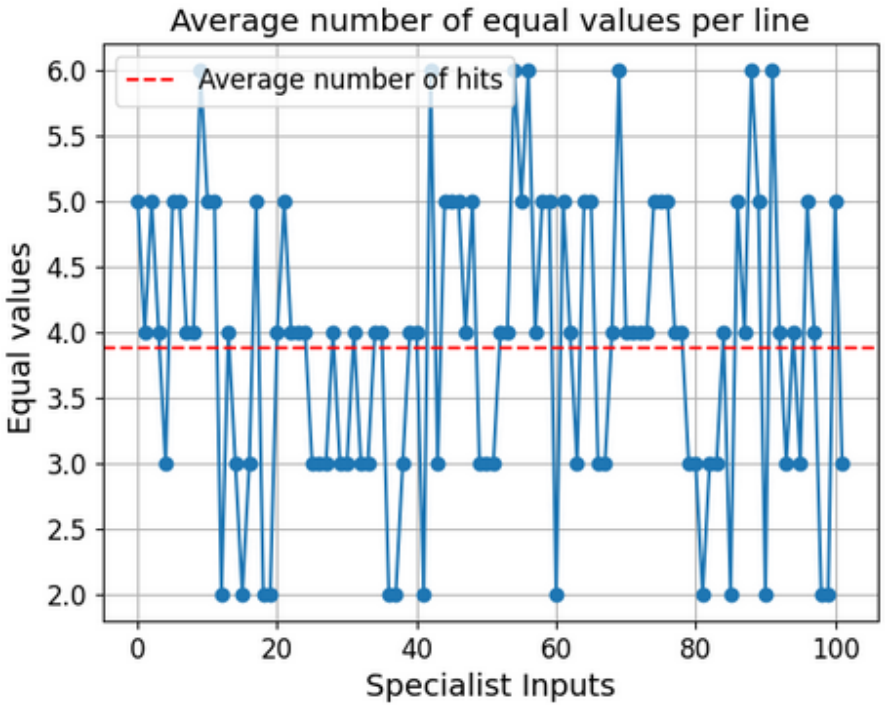


Figure. 5.13: Average number of equal values per line

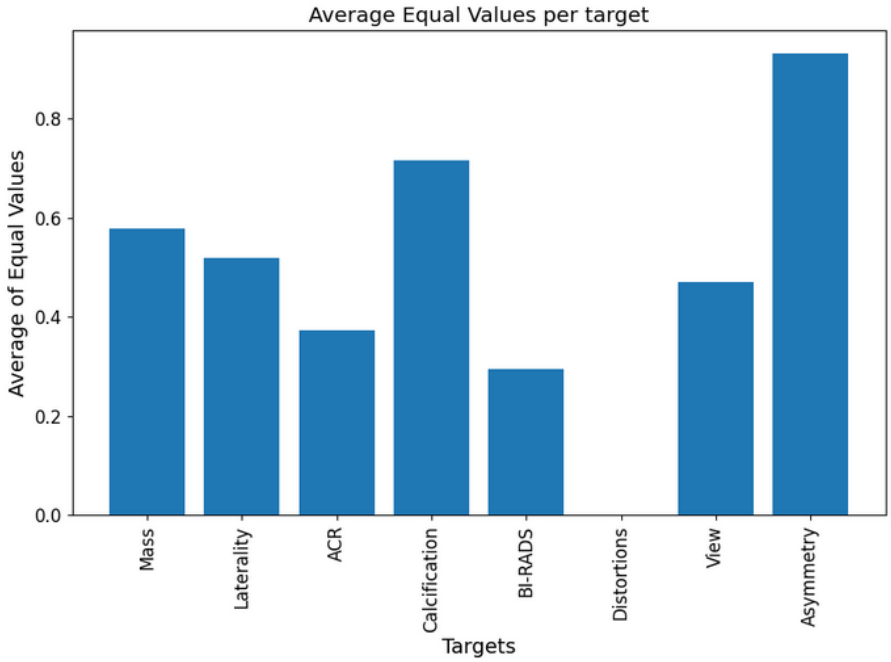


Figure. 5.14: Average Equal Values per dimensions

5.5.2 Software improvements

During the specialist's analysis, some difficulties arose, so the software was improved over time as the specialist carried out his task. All the suggested changes to the software are shown below:

1. The "probe" image must not appear on the right side at the time of association, i.e. the same image must not be compared to itself;
2. The probe image must not disappear on the left, it must always remain on the left while an image is selected on the right, i.e. the probe image must always be visible at all times;
3. Filters must allow the options themselves to be deselected;
4. The size of the images should be increased to make them easier to view. Allow zooming in on images, both the probe image and the images on the right;
5. Move the probe image to the front;
6. When a filter is selected, it is necessary to issue a warning that it was selected and did not give an error.

5.6 Summary

In this chapter, an overview is provided of software developed for collecting ground truth data intended for the development of a mammogram retrieval system. The software was built using the OutSystems platform, known for its low-code approach. The

initial requirements for the software are presented, such as the visualization of probe images, filtering of images based on dimensions, and viewing of records. Additional improvements suggested by an expert during the software development are also included. The entire implementation is described, starting from the software interface, through a login page, and covering the process of associating probe images with relevant images from the database. Subsequently, the inputs provided by a specialist are analyzed through the developed software. Graphs depict the distribution of similarities between probe images and images considered relevant. The specialist also highlights the need for improvements to ensure more effective software.

6 Results

This section is subdivided into four parts. Sections 6.1 and 6.2 present the evaluation of the individual classifiers. While in 6.1 the results were achived with the random split of the data; the results in section 6.2 were obtained when the dataset was divided by patient. In Section 6.3, an exploratory data analysis of the retrieval models, according to four versions of the weights (r_{const}, r_{emp}, r_{lit} and r_{exp}) is given. Finally, the last part (Section 6.4) contains the formal evaluation of the retrieval models, using the ground truth provided by the specialist (see Section 5.5)

6.1 Individual classifiers

For experimental purposes, two types of approaches (see Section 4.2.1) were considered:

- Training only the final fully connected layers (trainable = false);
- Training the networks from scratch (trainable = true).

In Figures 6.1, 6.2, 6.3, and 6.4, the metrics "Loss" and "Accuracy" refer to the training set, while "Val_accuracy" and "Val_loss" are results in the 20% of validation set. In Figure 6.5 the metrics "Loss" and "Accuracy" refer to the test set.

Based on the obtained results (Figure 6.1), it can be concluded that the InceptionV3 pre-trained network achieved the highest accuracy of 96% among the models used for classifying mammograms based on the BI-RADS dimension. For the breast density dimension (ACR), the network that obtained the best performance was InceptionResNetV2 with an accuracy of 98%.

	BI-RADS				ACR			
	Loss	Accuracy	Val_loss	Val_accuracy	Loss	Accuracy	Val_loss	Val_accuracy
	RESNET50				RESNET50			
treinable = false	1.4491	0.5041	1.4147	0.4945	1.3538	0.3415	1.2747	0.3556
treinable = true	0.5730	0.8211	1.5515	0.4945	0.3987	0.9024	1.3151	0.3556
	VGG16				VGG16			
treinable = false	1.4504	0.4553	1.4198	0.4945	1.3646	0.3496	1.2698	0.3556
treinable = true	1.4498	0.5122	1.4100	0.4945	1.3348	0.3008	1.2512	0.3556
	InceptionResNetV2				InceptionResNetV2			
treinable = false	1.4226	0.4797	1.4109	0.4945	1.3215	0.3171	1.2650	0.3556
treinable = true	0.3047	0.9024	1.5363	0.4945	0.0897	0.9837	1.4237	0.3778
	InceptionV3				InceptionV3			
treinable = false	1.3807	0.4797	1.3740	0.4945	1.3523	0.3496	1.2738	0.3333
treinable = true	0.1242	0.9593	1.5639	0.2088	0.0884	0.9593	1.6155	0.3111

Figure. 6.1: BI-RADS and breast density results.

Regarding the results obtained by the models that classify the mammograms according to the existence or not of masses (Figure 6.2), it is possible to say that the pre-trained

network InceptionV3 was the one that obtained the best results with an accuracy of 100%. The network that achieved the greatest results for the dimension calcification was InceptionV3, which also had a 100% accuracy rate.

	Mass				Calcifications			
	Loss	Accuracy	Val_loss	Val_accuracy	Loss	Accuracy	Val_loss	Val_accuracy
	RESNET50				RESNET50			
treinable = false	0.6069	0.7236	0.6392	0.6703	0.6096	0.7398	0.5567	0.7692
treinable = true	0.6663	0.8130	0.8439	0.6703	0.0665	0.9837	0.5566	0.7692
	VGG16				VGG16			
treinable = false	0.6201	0.7154	0.6353	0.6703	0.6164	0.7236	0.5840	0.7692
treinable = true	0.6067	0.7236	0.6368	0.6703	0.6288	0.7398	0.5489	0.7692
	InceptionResNetV2				InceptionResNetV2			
treinable = false	0.5896	0.7317	0.6707	0.6703	0.6199	0.7317	0.5418	0.7692
treinable = true	0.0619	0.9756	0.6443	0.6703	0.0208	1.0000	0.6229	0.7692
	InceptionV3				InceptionV3			
treinable = false	0.5597	0.7317	0.6465	0.6703	0.5604	0.7398	0.5219	0.7692
treinable = true	0.0135	1.0000	0.8536	0.6703	0.0139	1.0000	0.5798	0.7692

Figure. 6.2: Mass and calcifications results.

From the results obtained by the models that classify the mammograms according to the view (Figure 6.3), it is possible to deduce that the pre-trained network InceptionResNetV2 was the one that obtained the best results with an accuracy of 100%. With a 100% accuracy rate, InceptionV3 achieved the best performance for the laterality dimension.

	View				Laterality			
	Loss	Accuracy	Val_loss	Val_accuracy	Loss	Accuracy	Val_loss	Val_accuracy
	RESNET50				RESNET50			
treinable = false	0.7305	0.4146	0.6898	0.6889	0.7069	0.5203	0.6918	0.4945
treinable = true	0.0580	0.9756	0.7659	0.4778	0.0027	1.0000	0.8133	0.5275
	VGG16				VGG16			
treinable = false	0.7242	0.5122	0.7005	0.4778	0.7806	0.4309	0.6942	0.4725
treinable = true	0.6945	0.4959	0.6339	0.4778	0.6929	0.5122	0.6926	0.5275
	InceptionResNetV2				InceptionResNetV2			
treinable = false	0.7070	0.4878	0.6884	0.5667	0.6911	0.5447	0.6933	0.4725
treinable = true	0.0074	1.0000	0.6876	0.5222	0.0081	1.0000	0.7826	0.4725
	InceptionV3				InceptionV3			
treinable = false	0.6418	0.6423	0.6244	0.8000	0.4107	0.8862	0.3775	0.9560
treinable = true	0.0079	1.0000	0.6493	0.6222	0.0047	1.0000	0.6349	0.5934

Figure. 6.3: View and laterality results.

With regard to the classification based on dimensional asymmetries, it is verified that the pre-trained ResNet50 network is able to obtain the best results, with an accuracy of 100% (Figure 6.4). Among the tested networks, InceptionV3 demonstrated superior performance in handling dimensional distortions, achieving a remarkable accuracy of 100%.

	Asymmetries				Distortions			
	Loss	Accuracy	Val_loss	Val_accuracy	Loss	Accuracy	Val_loss	Val_accuracy
	RESNET50				RESNET50			
treinable = false	0.2417	0.9431	0.1806	0.9560	0.0674	0.9919	0.1194	0.9780
treinable = true	0.0110	1.0000	0.2132	0.9560	0.0123	0.9919	0.1549	0.9780
	VGG16				VGG16			
treinable = false	0.2356	0.9350	0.1819	0.9560	0.0569	0.9919	0.1307	0.9780
treinable = true	0.2322	0.9431	0.2287	0.9560	0.0574	0.9919	0.1217	0.9780
	InceptionResNetV2				InceptionResNetV2			
treinable = false	0.2677	0.9431	0.1800	0.9560	0.0558	0.9919	0.1160	0.9780
treinable = true	0.0429	0.9837	0.4145	0.9560	0.0078	1.0000	0.1907	0.9780
	InceptionV3				InceptionV3			
treinable = false	0.2507	0.9431	0.1836	0.9560	0.0802	0.9919	0.1262	0.9780
treinable = true	0.0112	1.0000	0.1743	0.9560	0.0047	1.0000	0.1339	0.9780

Figure. 6.4: Asymmetries and distortions results.

Figure 6.5 illustrates the test set's outcomes. For the BiRads dimension, the networks ResNet50 and VGG16 were able to obtain the best results with an accuracy of 66%.

For the breast density dimension (represented by ACR in Figure 6.5) it is the VGG16 network that obtains the best result with an accuracy of 41%. For the dimension asymmetries, the network that obtained the best result was VGG16 with an accuracy of 100%, and for the dimension distortions, it is ResNet5 that obtained the best performance, with an accuracy of 100%. For the dimension masses, the network ResNet50 obtained the best result with 81% accuracy. For the dimension calcification and view the inceptionResNetV2 network is the one that obtains the best performance with an accuracy of 81% and 94%, respectively. Finally, for the dimension laterality, it is the VGG16 network that obtains the best result with an accuracy of 44%. Overall, the test results were reasonable, having the networks behaved differently in relation to each dimension.

BI-RADS		ACR		Asymmetries		Distortions	
Loss	Accuracy	Loss	Accuracy	Loss	Accuracy	Loss	Accuracy
RESNET50		RESNET50		RESNET50		RESNET50	
1.5804	0.6562	1.5725	0.3750	0.0690	1.0000	0.0011	1.0000
VGG16		VGG16		VGG16		VGG16	
1.5006	0.6562	1.4786	0.4062	0.3195	1.0000	0.0998	1.0000
InceptionResNetV2		InceptionResNetV2		InceptionResNetV2		InceptionResNetV2	
2.6745	0.6562	2.3488	0.2812	7.0141	1.0000	3.1124	1.0000
InceptionV3		InceptionV3		InceptionV3		InceptionV3	
1.8360	0.5000	4.0283	0.2812	4.4703	1.0000	1.0803	1.0000
Masses		**Calcification**		**View**		**Laterality**	
Loss	Accuracy	Loss	Accuracy	Loss	Accuracy	Loss	Accuracy
RESNET50		RESNET50		RESNET50		RESNET50	
0.5750	0.8125	0.7292	0.7812	0.9862	0.5312	0.7816	0.4375
VGG16		VGG16		VGG16		VGG16	
0.7831	0.2187	0.6802	0.5937	1.1091	0.5312	0.6956	0.4062
0.6334	0.8125	0.5382	0.7812	0.9625	0.4687	0.7080	0.4375
InceptionResNetV2		InceptionResNetV2		InceptionResNetV2		InceptionResNetV2	
0.8398	0.6875	0.8050	0.8125	0.2266	0.9375	1.0682	0.5625
InceptionV3		InceptionV3		InceptionV3		InceptionV3	
1.7035	0.8125	0.9828	0.7812	1.5004	0.4687	3.6191	0.5000

Figure. 6.5: Results obtained in the test set.

6.2 Division by patient

In order to complement the work, the images (410) were divided by patient, as each patient has between two to four associated images. Building upon the results obtained from the previous models (see Section 6.1), and considering the outcomes achieved by the best networks for each dimension, a new training was conducted for each dimension using the networks that demonstrated the best performance in their respective dimensions. In Figures 6.6, the metrics "Loss" and "Accuracy" refer to the training set, while "Val_accuracy" and "Val_loss" are results in the 20% of validation set. In Figure 6.7 the metrics "Loss" and "Accuracy" refer to the test set.

Regarding the results of the test set, the "View" dimension obtained an accuracy of 73.88%, while the "Laterality" dimension obtained an accuracy of 65.88%. The detection of "Mass" shows a robust performance, achieving an accuracy of 81.25%. Notably, the same model achieves 100% accuracy in the "Distortions" category. As for the "Assymmetries" dimension, it performs remarkably well with a perfect accuracy of 100%. In the "Bi-Rads" dimension, the accuracy was only 6.25%. In the "ACR" dimension, it

View - InceptionResNetV2				
	Loss	Accuracy	Val_loss	Val_accuracy
Treinable=False	0.6892	0.5157	0.6864	0.5444
Treinable=True	0.0093	0.9965	0.6945	0.5222

Assymmetries - ResNet50				
	Loss	Accuracy	Val_loss	Val_accuracy
Treinable=False	0.1773	0.9652	0.1854	0.9560
Treinable=True	0.0657	0.9756	0.2130	0.9560

Laterality - InceptionV3				
	Loss	Accuracy	Val_loss	Val_accuracy
Treinable=False	0.1635	0.9652	0.1301	0.9778
Treinable=True	0.0019	1.0000	0.8202	0.5889

Bi-Rads - InceptionV3				
	Loss	Accuracy	Val_loss	Val_accuracy
Treinable=False	1.3728	0.5345	1.3837	0.4945
Treinable=True	0.1394	0.9690	1.5889	0.4835

Mass - InceptionV3				
	Loss	Accuracy	Val_loss	Val_accuracy
Treinable=False	0.5526	0.7561	0.6605	0.6703
Treinable=True	0.0049	1.0000	0.9285	0.6593

ACR - InceptionResNetV2				
	Loss	Accuracy	Val_loss	Val_accuracy
Treinable=False	1.3694	0.3241	1.3096	0.3043
Treinable=True	0.1458	0.9759	1.9740	0.3913

Distortions - InceptionV3				
	Loss	Accuracy	Val_loss	Val_accuracy
Treinable=False	0.0344	0.9965	0.1602	0.9780
Treinable=True	0.0050	1.0000	0.1902	0.9780

Calcifications - InceptionV3				
	Loss	Accuracy	Val_loss	Val_accuracy
Treinable=False	0.5318	0.7561	0.5209	0.7692
Treinable=True	0.2188	0.9338	0.7231	0.3297

Figure. 6.6: Results of patient-wise division.

performs relatively poorly with an accuracy of 25%. Finally, when it comes to detecting "Calcifications", it has a moderate accuracy of 78.12%, which suggests a satisfactory performance.

View - InceptionResNetV2	
Loss	Accuracy
2.3246	0.7688

Assymmetries - ResNet50	
Loss	Accuracy
0.0209	1.0000

Laterality - InceptionV3	
Loss	Accuracy
0.3020	0.6588

Bi-Rads - InceptionV3	
Loss	Accuracy
1.4696	0.0625

Mass - InceptionV3	
Loss	Accuracy
7.7452	0.8125

ACR - InceptionResNetV2	
Loss	Accuracy
6.3784	0.2500

Distortions - InceptionV3	
Loss	Accuracy
1.7830	1.0000

Calcifications - InceptionV3	
Loss	Accuracy
3.7794	0.7812

Figure. 6.7: Results obtained in the test set for the division by patient.

6.3 Retrieval models

Four versions of the weights were attempted:

- In the model r_{const}, all of the weights are equal.
- In the model r_{emp}, the weight values were empirically defined as: $w_{lat} = 0.000$, $w_{view} = 0.000$, $w_{density} = 0.167$, $w_{birads} = 0.167$, $w_{mass} = 0.167$, $w_{calcs} = 0.167$, $w_{dist} = 0.167$, $w_{assy} = 0.167$. The weights were assigned these values because both laterality and view have less importance and value in the detection of breast cancer as they end up being less informative dimensions, while density, BI-RADS, mass, calcifications, distortions and asymmetries can greatly help in the identification of this disease due to their presence or absence on the mammogram.
- The model r_{lit} is based on existing studies. For w_{lat} the value 0.018 was set because according to [36], the left breast is associated with a more aggressive biology. For w_{view}, the value 0.000 was set since no references were found. For $w_{density}$,

the value 0.172 is set because, according to [37], mammographic breast density is a strong reproducible risk factor for breast cancer. For w_{birads}, the value 0.172 is set because according to [38] BI-RADS provides a standardised terminology of lesion type and morphology as well as the enhancement features (enhancement pattern and signal intensity curves). He adds that the universal use of the BI-RADS lexicon has contributed to reducing inter-observer variability. The weight w_{mass} was set to 0.172 because according to [39] the ability to directly identify and visualise the characteristics of masses allows accurate and rapid diagnosis and prognosis. The value of w_{calcs} is 0.172 because the authors of [40] say that the identification of calcifications is one of the methods that allow the identification of breast cancer effectively. For w_{dist} the value 0.172 is set because it is found in [41] that architectural distortion is the third most suspicious appearance (after microcalcifications and masses), representing 6% of the abnormalities detected in screening mammography. Finally, w_{assy} is set to 0.122 because according to [42], pairing two opposite breasts to examine asymmetry improves the detection of local lesions.

- The creation of the model r_{exp} involved information gathered from an specialist.

  Therefore, the following weights were defined: $w_{lat} = 0.143$, $w_{view} = 0.140$, $w_{density} = 0.098$, $w_{birads} = 0.097$, $w_{mass} = 0.132$, $w_{calcs} = 0.118$, $w_{dist} = 0.142$, $w_{assy} = 0.130$.

The weights for each one of the four versions are summarised in Table 6.1.

Table. 6.1: Summary of weights in different models

Model	w_{lat}	w_{view}	$w_{density}$	w_{birads}	w_{mass}	w_{calcs}	w_{dist}	w_{assy}
r_{const}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
r_{emp}	0.000	0.000	0.167	0.167	0.167	0.167	0.167	0.167
r_{lit}	0.018	0.000	0.172	0.172	0.172	0.172	0.172	0.122
r_{exp}	0.143	0.140	0.098	0.097	0.132	0.118	0.142	0.130

For an illustration of the retrieval results, the probe image shown in Figure 6.8 was used. In Figure 6.8 the ground truth for each dimension is also shown. Some quantitative results of the final retrieval model according to the four approaches are presented in Table 6.2, where the similarity between the probe image and the most similar image (the first image of the top-5) is presented.

Table. 6.2: Retrieval results

	r_{const}	r_{emp}	r_{lit}	r_{exp}
$d^{probe,i}$	0.319	0.191	0.197	0.292

The model r_{const} obtained the value 0.319 and in Figure 6.9 it is possible to visualise the top-5 retrieved mammograms. The top-5 retrieved mammograms present six of

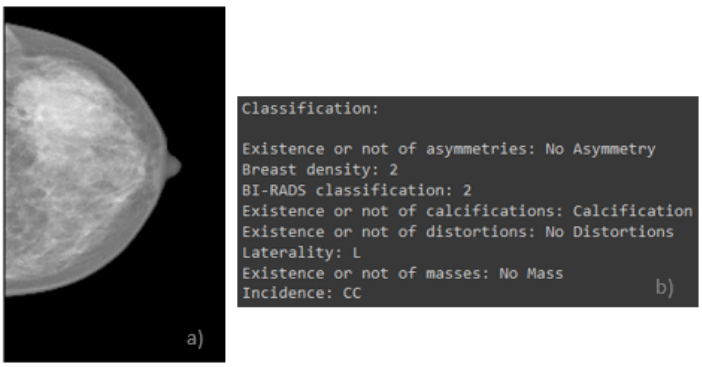


Figure. 6.8: Probe image along with the ground truth (ground truth is not used during retrieval)

the eight dimensions of the probe image. In this first approach, one would expect the dimensions laterality, view, mass, calcification and distortions to be the same as in the probe image, since in the distance calculation the weights have the same value, the same importance. It is clear that the use of a different distance metric (left for future work) would improve these results.

Figure. 6.9: Top-5 images retrieved by r_{const}

The model r_{emp} obtained the value 0.191 and in Figure 6.10 it is possible to visualise the top-5 retrieved mammograms. These five mammograms present two of the eight dimensions of the probe image. In this approach, since the weights that make up the distance formula have been defined empirically, it is possible to realise that laterality and view do not have the same importance, in addition it is the highest value presented in Table 6.2 for similarity.

The model r_{lit} obtained the value 0.197 and in Figure 6.11 it is possible to visualise the top-5 retrieved mammograms. These five mammograms present three of the eight dimensions of the probe image. In this approach, the weights were defined based on the literature, so the five retrieved mammograms are the result of this research.

Finally, the model r_{exp} obtained the value 0.292 and in Figure 6.12 it is possible to visualise the top-5 retrieved mammograms. These five mammograms present three of

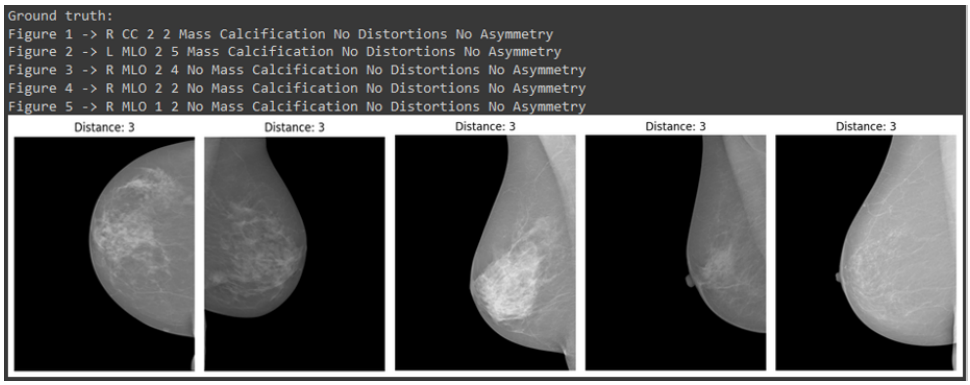


Figure. 6.10: Top-5 images retrieved by r_{emp}

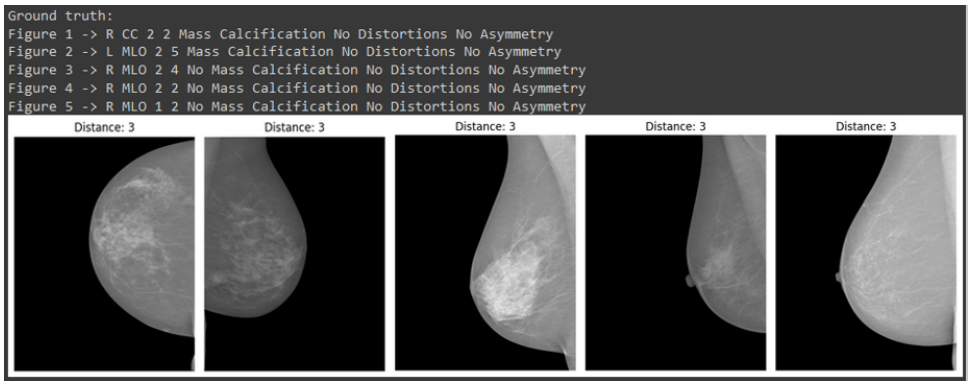


Figure. 6.11: Top-5 images retrieved by r_{lit}

the eight dimensions of the probe image. In this approach, which is based on specialist opinion, it is to be expected that the five mammograms displayed are the most relevant when given the probe image in Figure 6.8. The specialist's information is somewhat constrained because the work was not done under ideal circumstances, which is one of this approach's shortcomings. For instance, using a mammography reading screen would make it easier to understand each component of the mammograms.

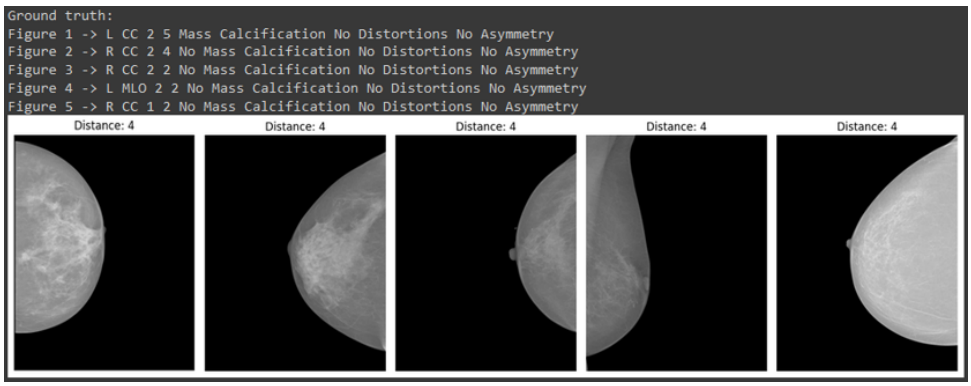


Figure. 6.12: Top-5 images retrieved by r_{exp}

In summary, the results of the final retrieval model were quite positive for the four different approaches, with the first approach standing out with six out of 8 dimensions

presented.

6.4 Evaluation

This section analyses the results obtained when evaluating the data provided by an specialist in the retrieval model, according to two approaches.

6.4.1 All images

This evaluation is carried out by finding out the position in which each retrieval model places the choice image (image chosen by the specialist) on a scale of 0 to 409, for 410 images.

The full evaluation results are in Annex G - Retrieval extended results. According to the results obtained, it is possible to conclude that r_{const} and r_{exp} have a very solid performance, with the majority of positions closer to 0, indicating a good positioning from 0 to 409. This suggests considerable effectiveness in classifying the images. r_{emp} and r_{lit} show a similar performance to r_{const} and r_{exp}, but with a slightly greater variation in the classifications, indicating a wider dispersion in the results. Finally, it can be seen that in r_{const} and r_{exp} the best classification is 0 and the worst 356, while in r_{emp} and r_{lit} the best classification is 0 and the worst 389. In summary, r_{const} and r_{exp} yield the best results when it comes to positioning the images on a scale from 0 to 409.

According to the box plot in Figure 6.13, it is noticeable that the box plots of r_{const} and r_{exp} have a different distribution from the others, meaning that the majority of the data consists of lower values. On the other hand, the box plots of r_{emp} and r_{lit} exhibit higher variability, with wider distributions, indicating higher values.

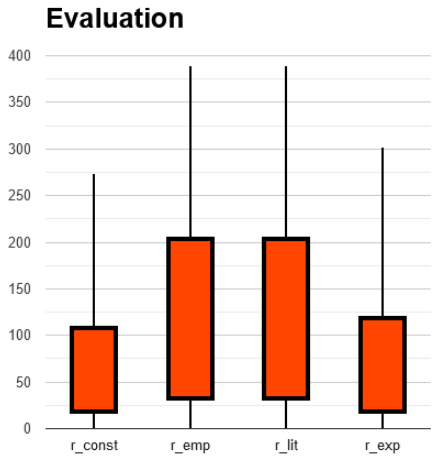


Figure. 6.13: Retrieval evaluation

6.4.2 Evaluation of division by patient

The results obtained in the evaluation also stemmed from the discovery of the position of the image chosen by the specialist on a scale from 0 to 409, except that the images were divided by patient. The results of this evaluation are presented in Annex H - Version 2 of the extended results retrieval. According to the results obtained, it is possible to conclude that r_{const} has a very solid performance, with most of the positions closer to 0, indicating a good positioning from 0 to 409. r_{emp}, r_{lit} and r_{exp} show a greater variation in positions, indicating a greater dispersion in the results. Finally, it can be seen that in r_{const} the best position is 1 and the worst is 330, while in r_{emp}, r_{lit} and r_{exp} the best position is 1 and the worst is 394. In short, r_{const} produces the best results when it comes to positioning the images on a scale from 0 to 409.

According to the box plot in Figure 6.14, it is noticeable that the box plots of r_{const} have a different distribution from the others, meaning that the majority of the data consists of lower values. On the other hand, the box plots of r_{emp}, r_{lit} and r_{exp} exhibit higher variability, with wider distributions, indicating higher values.

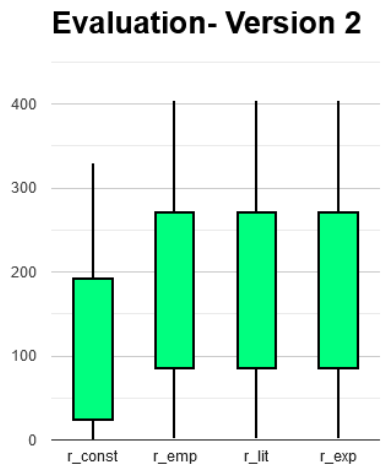


Figure. 6.14: Retrieval evaluation - Version 2

6.5 Summary

This chapter presents all the results. For individual classifiers, detailed analyses are provided for various dimensions such as BI-RADS, breast density, masses, calcifications, view, laterality, asymmetries, and distortions. The results show high accuracy for certain models, highlighting the effectiveness of the classifiers in detecting anomalies in mammograms. Another adopted process was "Division by patient", where images are grouped by patients, and a new training is carried out based on the models that performed best in the evaluation of individual classifiers. The results show varied accuracies

for different dimensions, indicating the models' ability to adapt to patient-specific data. Four variations of weights (r_{const}, r_{emp}, r_{lit}, and r_{exp}) were introduced, highlighted their performance in retrieving relevant mammograms based on similarity metrics. The results indicate promising outcomes in different weight configurations, with the r_{const} approach standing out. Finally, an overall evaluation of the retrieval models is conducted using two approaches: considering all images and considering images divided by patient. The r_{const} model consistently demonstrates solid performance, with most positions closer to the ideal value, highlighting its effectiveness in positioning images on a scale of 0 to 409. The box plots illustrate the distribution of positions, highlighted the consistency and reliability of r_{const} compared to other models. These findings contribute valuable insights to the development of mammography retrieval systems, showcasing the potential to enhance breast cancer detection and clinical support in diagnosis.

7 Conclusion

Breast cancer is a widespread illness that affects a large number of people, with women being more at risk for developing it. Despite advancements in medical knowledge and treatments, there is still a critical need for extensive research in this field. This urgency arises from the fact that breast cancer continues to claim the lives of thousands of individuals each year. By investing in further research efforts, we can strive to improve prevention, early detection, and treatment strategies, ultimately reducing the devastating impact of this disease on countless lives. The present project is related to the aforementioned problem since early detection of breast cancer is essential for a favourable prognosis.

A retrieval model has been developed in which an image is an input into the system and a set of relevant images with known diagnoses and their respective histories are returned. Towards this end, multiple models were created, one for each dimension including breast density, asymmetries, BI-RADS classification, presence of calcifications, distortions, laterality (right or left breast), presence of masses, and image incidence (CC or MLO). Each model was trained with four pre-trained networks: ResNet50, VGG16, InceptionV3 and InceptionResNetV2, in order to understand which one performed best. In order to create the final model retrieval model, we here proposed the use of a weighted sum of the aforementioned models. Four versions were then attempted, a first version where all the dimensions are equally important, a second version where each dimension was empirically weighted, a third where the weights were defined according to the literature, and a final one where the values of the weights were defined by a specialist.

A software tool was developed to gather data, thereby facilitating the creation of a mammography retrieval system. This software was developed according to initial requirements, and subsequently refined, to have a final quality version to facilitate the specialist's work. The disadvantages of the software are that it is not constantly fed with new mammogram images.

As for the results obtained by the individual classification models in the test phase, according to each dimension, the best accuracy achieved is 66% for BI-RADS classification, 41% for breast density, 98% for the presence of masses, 81% for the presence of calcifications, 94% for view, 44% for laterality, 100% for asymmetries, and 100% for distortions. The final retrieval models behaved differently, since the weights attributed to each dimension were defined in different ways, i.e. each dimension had a specific

importance. The r_{const} of the final retrieval model managed to achieve six out of the eight dimensions.

An analysis was made of the results obtained in the evaluation carried out on the data provided by an specialist in the retrieval model. This evaluation consisted of discovering the position in which the retrieval model places the image chosen by the specialist on a scale of 0 to 409. It was possible to conclude that r_{const} and r_{exp} obtained the best classification results for images on a scale of 0 to 409.

Several future directions are foreseen. Concerning the classifiers, two training methodologies were used, freeze all layers except the last one, and train all the layers. It is planned to test an intermediate solution, by iteratively unfreezing the layers of the classifiers. For the retrieval models it is planned, for example, to experiment with different distance metrics. Another line to pursue is to automatically derive the weights of each dimension in the final weighted sum model. To achieve this objective, we can leverage artificial intelligence techniques such as the training of neuronal networks or the use of evolutionary algorithms such as genetic algorithms, or Particle Swarm Optimisation type algorithms. As for the software, the current version has been developed to be used by an specialist to collect ground truth information [43]. The primary future focus centres on software enhancement, specifically the creation of a version suitable for clinical application, incorporating retrieval models.

Bibliographical references

[1] M. Arnold, E. Morgan, H. Rumgay, A. Mafra, D. Singh, M. Laversanne, J. Vignat, J. R. Gralow, F. Cardoso, S. Siesling, and I. Soerjomataram, "Current and future burden of breast cancer: Global statistics for 2020 and 2040," *The Breast*, vol. 66, pp. 15–23, 2022.

[2] L. P. C. o. Cancro, "Cancro da Mama : Liga Portuguesa Contra o Cancro," access date: 2023-01-31. [Online]. Available: http://www.ligacontracancro.pt/cancro-da-mama/

[3] S. Bessa, I. Domingues, J. S. Cardosos, P. Passarinho, P. Cardoso, V. Rodrigues, and F. Lage, "Normal breast identification in screening mammography: a study on 18 000 images," in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2014, pp. 325–330.

[4] L. P. C. o. Cancro. Detecção - cancro da mama : Liga portuguesa contra o cancro, access date: 2023-05-16. [Online]. Available: http://www.ligacontracancro.pt/cancro-da-mama-deteccao/

[5] "Cancro da Mama – Cancro Online," access date: 2023-11-15. [Online]. Available: https://www.cancro-online.pt/site/cancro-da-mama/

[6] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: Toward a Full-field Digital Mammographic Database," *Academic Radiology*, vol. 19, no. 2, pp. 236–248, 2012, publisher: Elsevier.

[7] V. Sharma, R. K. Rajasekaran, and S. Badhrinarayanan, "Visualization of Data Mining Techniques for the Prediction of Breast Cancer with High Accuracy Rates," *Journal of Computer Science*, vol. 15, no. 1, pp. 118–130, 2019.

[8] A. Stanislavsky, "Radial scar | Image | Radiopaedia.org," access date: 2023-12-07. [Online]. Available: https://radiopaedia.org/images/41818958?case_id=31417

[9] F. d. C. M. Unicamp, "Mamografia - assimetrias," access date: 2023-12-07. [Online]. Available: https://drpixel.fcm.unicamp.br/es/node/34

[10] D. B. Kopans, J. E. Meyer, and N. Sadowsky, "Breast imaging," *New England Journal of Medicine*, vol. 310, no. 15, pp. 960–967, 1984.

[11] A. Dennanni, "Acelere seus projetos de Deep Learning com Redes Neurais pré-treinadas," 2018, access date: 2023-

11-15. [Online]. Available: https://medium.com/neuronio-br/acelere-seus-projetos-de-deep-learning-com-redes-neurais-pre-treinadas-b5c43bc590e1

[12] "What are Convolutional Neural Networks? | IBM," access date: 2023-11-15. [Online]. Available: https://www.ibm.com/topics/convolutional-neural-networks

[13] "Rede neural convolucional ResNet-50 - MATLAB resnet50," access date: 2023-11-15. [Online]. Available: https://www.mathworks.com/help/deeplearning/ref/resnet50.html

[14] "Understanding VGG16: Concepts, Architecture, and Performance," access date: 2023-11-15. [Online]. Available: https://datagen.tech/guides/computer-vision/vgg16/

[15] "Rede neural convolucional Inception-v3 - MATLAB inceptionv3," access date: 2023-11-15. [Online]. Available: https://www.mathworks.com/help/deeplearning/ref/inceptionv3.html

[16] "Rede neural convolucional Inception-ResNet-v2 pré-treinada - MATLAB inceptionresnetv2," access date: 2023-11-15. [Online]. Available: https://www.mathworks.com/help/deeplearning/ref/inceptionresnetv2.html

[17] S. Yadav, "What is Kernel Trick in SVM ? Interview questions related to Kernel Trick," 2023, access date: 2023-11-28. [Online]. Available: https://medium.com/@Suraj_Yadav/what-is-kernel-trick-in-svm-interview-questions-related-to-kernel-trick-97674401c48d

[18] S. J. Rayen and R. Subhashini, "An Efficient Mammogram Image Retrieval System Using an Optimized Classifier," *Neural Processing Letters*, vol. 53, no. 4, pp. 2467–2484, 2021.

[19] K. Kiruthika, D. Vijayan, and L. R., "Retrieval Driven Classification for Mammographic Masses," in *2019 International Conference on Communication and Signal Processing (ICCSP)*, 2019, pp. 0725–0729.

[20] M. Jiang, S. Zhang, H. Li, and D. N. Metaxas, "Computer-Aided Diagnosis of Mammographic Masses Using Scalable Image Retrieval," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 783–792, 2015.

[21] L. Tsochatzidis, K. Zagoris, M. Savelonas, N. Papamarkos, I. Pratikakis, N. Arikidis, and L. Costaridou, "Microcalcification oriented content-based mammogram retrieval for breast cancer diagnosis," in *IEEE International Conference on Imaging Systems and Techniques (IST)*, 2014, pp. 257–262.

[22] H. Bulu, "Ontology-based medical image annotation and retrieval," Thesis, DEÜ Fen Bilimleri Enstitüsü, 2013.

[23] T. Berber, "Integration of content-based image retrieval and database management system: A case study with digital mammography," Thesis, DEÜ Fen Bilim-

leri Enstitüsü, 2013.

[24] T. M. Deserno, M. Soiron, J. E. E. d. Oliveira, and A. d. A. Araújo, "Computer-aided diagnostics of screening mammography using content-based image retrieval," in *Medical Imaging: Computer-Aided Diagnosis*, vol. 8315. SPIE, 2012, pp. 647–655.

[25] T. Deserno, M. Soiron, J. Oliveira, and A. Araujo, "Towards Computer-Aided Diagnostics of Screening Mammography Using Content-Based Image Retrieval," in *24th SIBGRAPI Conference on Graphics, Patterns and Images*, 2011, pp. 211–219.

[26] J. E. E. de Oliveira, A. de Albuquerque Araújo, and T. M. Deserno, "Content-based image retrieval applied to BI-RADS tissue classification in screening mammography," *World Journal of Radiology*, vol. 3, no. 1, pp. 24–31, 2011.

[27] H. Jing and Y. Yang, "Image retrieval for computer-aided diagnosis of breast cancer," in *IEEE Southwest Symposium on Image Analysis & Interpretation (SSIAI)*, 2010, pp. 9–12.

[28] B. Zheng, "Computer-Aided Diagnosis in Mammography Using Content-Based Image Retrieval Approaches: Current Status and Future Perspectives," *Algorithms*, vol. 2, no. 2, pp. 828–849, 2009.

[29] J. Tang, R. M. Rangayyan, J. Xu, I. E. Naqa, and Y. Yang, "Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 2, pp. 236–251, 2009.

[30] R. M. Nishikawa, L. Wei, and Y. Yang, "Retrieval-driven microcalcification classification for breast cancer diagnosis," in *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2007, pp. 1260–1263.

[31] I. El-Naqa, Y. Yang, N. Galatsanos, R. Nishikawa, and M. Wernick, "A similarity learning approach to content-based image retrieval: application to digital mammography," *IEEE Transactions on Medical Imaging*, vol. 23, no. 10, pp. 1233–1244, 2004.

[32] PenRad, "PenFetch Automatic Smart Image Transfer and More. Eliminate Waste." 2008, access date: 2023-08-07. [Online]. Available: http://penrad.comcastbiz.net/pdf/PENFETCH_2%20Page.pdf

[33] V. Sharma, "Mammogram Image Retrieval System Using Texture and Semantic Features," *J of Physics: Conf Ser*, vol. 2267, no. 1, p. 012071, 2022.

[34] I. Domingues, J. S. Cardoso, I. Amaral, I. Moreira, P. Passarinho, J. Santa Comba, R. Correia, and M. J. Cardoso, "Pectoral muscle detection in mammograms based on the shortest path with endpoints learnt by SVMs," in *Annual International Conference of the IEEE Engineering in Medicine and Biology*, 2010, pp. 3158–3161.

[35] "Low-code de alto desempenho para desenvolvimento de aplicativos | OutSystems," acess date: 2023-11-15. [Online]. Available: https://www.outsystems.com/pt-br/

[36] Y. Abdou, M. Gupta, M. Asaoka, K. Attwood, O. Mateusz, S. Gandhi, and K. Takabe, "Left sided breast cancer is associated with aggressive biology and worse outcomes than right sided breast cancer," *Scientific Reports*, vol. 12, no. 1, p. 13377, 2022.

[37] A. Anandarajah, Y. Chen, G. A. Colditz, A. Hardi, C. Stoll, and S. Jiang, "Studies of parenchymal texture added to mammographic breast density and risk of breast cancer: a systematic review of the methods used in the literature," *Breast Cancer Research*, vol. 24, no. 1, p. 101, 2022.

[38] L. Meng, X. Zhao, J. Guo, L. Lu, M. Cheng, Q. Xing, H. Shang, B. Zhang, Y. Chen, P. Zhang, and X. Zhang, "Improved Differential Diagnosis Based on BI-RADS Descriptors and Apparent Diffusion Coefficient for Breast Lesions: A Multiparametric MRI Analysis as Compared to Kaiser Score," *Academic Radiology*, 2023.

[39] E. Cuypers, B. S. R. Claes, R. Biemans, N. G. Lieuwes, K. Glunde, L. Dubois, and R. M. A. Heeren, "'On the Spot' Digital Pathology of Breast Cancer Based on Single-Cell Mass Spectrometry Imaging," *Analytical Chemistry*, vol. 94, no. 16, pp. 6180–6190, 2022.

[40] S. Chaudhury, M. Rakhra, N. Memon, K. Sau, and M. T. Ayana, "Breast Cancer Calcifications: Identification Using a Novel Segmentation Approach," *Computational and Mathematical Methods in Medicine*, vol. 2021, p. e9905808, 2021.

[41] X. Chen, Y. Zhang, J. Zhou, X. Wang, X. Liu, K. Nie, X. Lin, W. He, M.-Y. Su, G. Cao, and M. Wang, "Diagnosis of architectural distortion on digital breast tomosynthesis using radiomics and deep learning," *Frontiers in Oncology*, vol. 12, 2022.

[42] Y. Guan, X. Wang, H. Li, Z. Zhang, X. Chen, O. Siddiqui, S. Nehring, and X. Huang, "Detecting Asymmetric Patterns and Localizing Cancers on Mammograms," *Patterns*, vol. 1, no. 7, p. 100106, 2020.

[43] C. Roriz, V. Vasconcelos, and I. Domingues, "Software for mammogram image retrieval," in *29th edition of the Portuguese Conference on Pattern Recognition* (*RECPAD*), 2023.

Annexes

Annex A - Selection of papers

Three sheets were created in Microsoft Excel. In the first (see Figure 7.1), all the studies were entered, including the respective title, year of publication, number of citations of the study, Digital Object Identifier(DOI) or hyperlink, whether or not the study was selected, and finally the respective observations, i.e. the reason for exclusion. On the second sheet (see Figure 7.2), papers containing citations are selected/excluded. The third sheet (see Figure 7.3) selects/excludes papers that do not contain any citations.

1	**Pesquisa googe scholar-->** intitle:Retrieval + intitle:mammography					
2	**43 resultados apresentados**					
3	**Artigo**	**Ano**	**Nº de citações**	**DOI ou Link**	**Excluir**	**OBS**
4	A similarity learning approach to content-based image retrieval: application to digital mammography	2004	354	10.1109/TMI.2004.834601	-	Acesso ao pdf
5	Computer-aided diagnosis in mammography using content-based image retrieval approaches: current status and future perspectives	2009	74	https://doi.org/10.3	-	Acesso ao pdf
6	Towards computer-aided diagnostics of screening mammography using content-based image retrieval	2011	43	10.1109/SIBGRAPI.2011.40	-	Acesso ao pdf
7	Automated and effective content-based image retrieval for digital mammography	2015	23	https://doi.org/10.1145/2	X	Sem acesso
8	Content-based image retrieval applied to BI-RADS tissue classification in screening mammography	2011	51	https://www.ncbi.n	-	Acesso ao pdf
9	Computer-aided diagnostics of screening mammography using content-based image retrieval	2012	30	https://doi.org/10.1117/12	-	Acesso ao pdf
10	Content-based image retrieval for digital mammography	2002	31	10.1109/ICIP.2002.1038924	-	Acesso ao pdf
11	Voice-activated retrieval of mammography reference images	1998	24	https://doi.org/10.1	-	Acesso ao pdf
12	Ontology-based mammography annotation and case-based retrieval of breast masses	2012	16	https://doi.org/10.1016/j.c	X	Sem acesso
13	Storage, transmission, and retrieval of digital mammography, including recommendations on image compression	2006	17	https://doi.org/10.1016/j.i	X	Sem acesso
14	Content-based image retrieval in mammography: using texture features for correlation with BI-RADS categories	2003	15	https://doi.org/10.1	X	Sem acesso
15	A mobile agent system for distributed mammography image retrieval	2002	6	MQ75169.pdf (uregina.ca)	-	Acesso ao pdf
16	Content-based image retrieval by similarity learning for digital mammography	2002	4	Content-based image retri	X	Sem acesso
17	The role of content-based image retrieval in mammography CAD	2014	3	https://doi.org/10.1	X	Sem acesso
18	Toward perceptually driven image retrieval in mammography: a pilot observer study to assess visual similarity of masses	2008	4	https://doi.org/10.1117/12	-	Acesso ao pdf
19	Secure and image retrieval based on multipurpose watermarking for mammography images database	2014	3	https://doi.org/10.5120/15	-	Acesso ao pdf
20	Storage and breast region segmentation for a non-distributed approach to clinical scale content-based image retrieval in mammography	2013	3	https://doi.org/10.1117/12	-	Acesso ao pdf
21	Integration of content-based image retrieval and database management system: A case study with digital mammography	2013	2	http://hdl.handle.net/20.5	X	Sem acesso
22	A boolean information retrieval system for mammography reports	2009	1	10.1109/BIYOMUT.2009.51	-	Acesso ao pdf
23	Ontology-based annotation and retrieval system for digital mammography images	2009	1	https://d1wqtxts1xzle7.clo	-	Acesso ao pdf
24	Content-based retrieval of calcification lesions in mammography	2010	1	10.1109/ICBBE.2010.5517	-	Acesso ao pdf
25	Assessment of content-based image retrieval approaches for mammography based on breast density patterns	2016	1	https://doi.org/10.3	-	Acesso ao pdf
26	Image retrieval based on histogram comparison for digital mammography workstations	2002	1	Microsoft Word - icbme200	-	Acesso ao pdf
27	Low level feature selection for a content based digital mammography image retrieval system	2009	1	10.1109/SIU.2009.5136428	-	Acesso ao pdf
28	Intelligent mammography retrieval engine	2007	-	10.1109/TENCON.2007.44	-	Acesso ao pdf
29	Content based mammography images retrieval using Ripley's K function	2009	-	https://doi.org/10.1145/16	-	Acesso ao pdf
30	Ontology-based Mammography Annotation and Similar Mass Retrieval with SQWRL	-	-	https://www.acade	-	Acesso ao pdf
31	Assessment and improvement of content-based image retrieval approaches for screening mammography analysis	2016	-	10.1109/SIU.2016.7496016	-	Acesso ao pdf
32	Low Level Feature Selection for a Content Based Digital Mammography Image Retrieval System	2009	-	10.1109/siu.2009.5136428	-	Acesso ao pdf

Figure. 7.1: Collection of papers.

1	Artigos com citações					
2	18 resultados apresentados					
3	**Artigo**	**Ano**	**Nº de citações**	**DOI ou Link**	**Excluir**	**OBS**
4	A similarity learning approach to content-based image retrieval: application to digital mammography	2004	354	10.1109/TMI.2004.834601	-	Inclui
5	Computer-aided diagnosis in mammography using content-based image retrieval approaches: current status and future pe	2009	74	https://doi.org/10.33	-	Inclui
6	Towards computer-aided diagnostics of screening mammography using content-based image retrieval	2011	43	10.1109/SIBGRAPI.2011.40	-	Inclui
7	Content-based image retrieval applied to BI-RADS tissue classification in screening mammography	2011	51	https://www.ncbi.nln	-	Inclui
8	Computer-aided diagnostics of screening mammography using content-based image retrieval	2012	30	https://doi.org/10.1117/12.9	-	Inclui
9	Content-based image retrieval for digital mammography	2002	31	10.1109/ICIP.2002.1038924	-	Inclui
10	Voice-activated retrieval of mammography reference images	1998	24	https://doi.org/10.10	-	Não inclui
11	A mobile agent system for distributed mammography image retrieval	2002	6	MQ75169.pdf (uregina.ca)	-	Não inclui
12	Toward perceptually driven image retrieval in mammography: a pilot observer study to assess visual similarity of masses	2008	4	https://doi.org/10.1117/12.7	-	Não inclui
13	Secure and image retrieval based on multipurpose watermarking for mammography images database	2014	3	https://doi.org/10.5120/1578	-	Não inclui
14	Storage and breast region segmentation for a non-distributed approach to clinical scale content-based image retrieval in m	2013	3	https://doi.org/10.1117/12.2	-	Não inclui
15	A boolean information retrieval system for mammography reports	2009	1	10.1109/BIYOMUT.2009.5130	-	Não inclui
16	Ontology-based annotation and retrieval system for digital mammography images	2009	1	https://d1wqtxts1xzle7.cloud	-	Não inclui
17	Content-based retrieval of calcification lesions in mammography	2010	1	10.1109/ICBBE.2010.5517206	-	Não inclui
18	Assessment of content-based image retrieval approaches for mammography based on breast density patterns	2016	1	https://doi.org/10.32	-	Inclui
19	Image retrieval based on histogram comparison for digital mammography workstations	2002	1	Microsoft Word - icbme2002.	-	Não inclui
20	Low level feature selection for a content based digital mammography image retrieval system	2009	1	10.1109/SIU.2009.5136428	-	Inclui

Figure. 7.2: Choice of papers with citations.

1	Artigos sem citações					
2	9 resultados apresentados					
3	**Artigo**	**Ano**	**Nº de citações**	**DOI ou Link**	**Excluir**	**OBS**
4	Intelligent mammography retrieval engine	2007	-	10.1109/TENCON.200	-	Não inclui
5	Content based mammography images retrieval using Ripley's K function	2009	-	https://doi.org/10.11	-	Não inclui
6	Ontology-based Mammography Annotation and Similar Mass Retrieval with SQWRL	-	-	https://www.ac	-	Não inclui
7	Assessment and improvement of content-based image retrieval approaches for screening mammography analysis	2016	-	10.1109/SIU.2016.74	-	Não inclui
8	An assessed digital mammography segmentation algorithm used for content-based image retrieval	2006	-	10.1109/ICOSP.2006.	-	Não inclui
9	A content-based digital mammography retrieval using inexact graph matching	2014	-	10.1109/IPAS.2014.7	-	Não inclui
10	Mamografi Raporları için Mantıksal Bilgi Erişim Sistemi A Boolean Information Retrieval System For Mammograph	2009	-	biyomut2009-with-cc	-	Não inclui
11	Sayısal Mamogram Erişim Sistemi için Alt Düzey Öznitelik Çıkarımı Low Level Feature Selection for a Content Base	2009	-	siu2009-libre.pdf (d1	-	Não inclui

Figure. 7.3: Choice of papers without citations.

Annex B - CIARP paper

Mammogram Retrieval based on the aggregation of image classifiers

Cátia Roriz[1], Inês C. Moreira[2][0000−0001−6868−1613], Verónica Vasconcelos[1,3][0000−0002−1638−1405], and Inês Domingues[1,4][0000−0002−2334−7280]

[1] Instituto Politécnico de Coimbra, Instituto Superior de Engenharia, Rua Pedro Nunes - Quinta da Nora, 3030-199 Coimbra, Portugal
[2] CHUSJ, ESS-IPP, Cintesis-FMUP
[3] INESC TEC Porto
[4] Centro de Investigação do Instituto Português de Oncologia do Porto (CI-IPOP): Grupo de Física Médica, Radiobiologia e Protecção Radiológica
{a21270590,veronica,ines.domingues}@isec.pt,icm@ess.ipp.pt

Abstract. Breast cancer is a serious health issue that affects people all over the world, particularly women. Early-stage diagnosis is essential for effective treatment and better patient outcomes. This paper presents a mammogram retrieval system based on the aggregation of image classifiers to aid specialists in diagnosing breast cancer. The system uses a retrieval model that combines the output of multiple classifiers, each targeting different dimensions related to breast cancer diagnosis. These dimensions include breast density, asymmetries, BI-RADS classification, calcifications, distortions, laterality, masses, and image incidence. Results obtained in the training set, for the models for each dimension, achieve, on average, an accuracy of about 99.3%, while in the test set, the average accuracy is around 78%. Four approaches were then developed for the final retrieval model, one assigning equal weights to every dimension, another with empirically defined weights, a third where the weights were defined according to the literature, and a final one where the values of the weights were defined by a specialist.

Keywords: Mammogram retrieval · Breast cancer · Image classification · Deep Learning.

1 Introduction

Breast cancer is a malignant disease that affects thousands of individuals worldwide, primarily women. Although it can also occur in men, they represent a minority of cases. By the year 2040 [3], a significant increase in the number of breast cancer cases is estimated, with more than 3 million new cases annually, representing a 40% increase. In addition, the number of deaths related to this disease is expected to increase by 50%, exceeding 1 million deaths per year. Breast cancer [3] is currently the most common type of cancer diagnosed, accounting for about 1 in 8 cancer cases worldwide. In 2020, approximately 2.3 million new cases of breast cancer were reported globally, resulting in about 685,000 deaths. However, it is important to note that these numbers vary significantly between countries and regions [3]. Although various approaches have been developed, much more effort has to be done to successfully fight this disease.

In Portugal [7], approximately 7,000 new cases of breast cancer are diagnosed annually, resulting in around 1,800 deaths among women. About 1% of breast cancer cases in Portugal are in men [7]. Given the recommended two-year screening interval for women aged 45 to 69, taking Portugal as an

example, a substantial number of mammograms require analysis each day [5]. The objective of this work was to develop a framework that helps the specialists to perform an early-stage diagnosis of breast cancer, facilitating prompt treatment for patients. Timely treatment significantly improves the chances of a swift recovery and prevents disease progression.

The main goal of the present work was to develop a medical aid tool that, when loading an image to analyse, returns a set of relevant images with known diagnoses. To achieve this main goal, a retrieval model that combines the output of several classifiers was developed. These individual classifiers had as targets the American College of Radiology (ACR) [21] breast density, existence or not of asymmetries, Breast Imaging-Reporting and Data System (BI-RADS) classification (see Figure 1 for some examples), existence or not of calcifications, existence or not of distortions, laterality (right or left breast), existence or not of masses and image incidence (Cranio-Caudal (CC) or Medio-Lateral Oblique (MLO)).

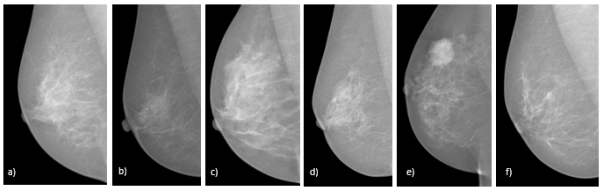


Fig. 1. Mammogram examples with different BI-RADS classifications: a) BI-RADS 1; b) BI-RADS 2; c) BI-RADS 3; d) BI-RADS 4; e) BI-RADS 5; f) BI-RADS 6

.

The main contributions of the present work include:

- The development of individual classifiers according to the eight dimensions of the mammograms.
- The proposal of a framework for mammogram image retrieval.
- The instantiation of the above framework, producing two different concrete retrieval models that combine the above mentioned eight classifiers.

This paper is organised as follows. The present Section presents an introductory reflection on the topic to be addressed. Section 2 provides a brief discussion of the existing studies on retrieval mammography. Section 3 describes the methodology adopted in this project's development and discusses all the work done, from the classifier's development to the creation of the final retrieval model and the database used. Section 4 deals with the results obtained and finally, Section 5 addresses final considerations and future work.

2 State of art

A literature review was made by highlighting the studies pertinent to the proposed theme. The search term in title section: "Retrieval + mammography" was used in the search engine "Google Scholar" to locate the best studies, yielding a collection of 45 results. As exclusion criteria, only studies with ten or more citations were selected, leading to the final selection of 14 studies.

In [26], an optimised classifier was employed to develop a method for mammogram retrieval. The study utilised the Mini-MIAS ("Mammography Image Analysis Society") database, extracting 216 images (51 malignant, 63 benign, and 102 normal). Preprocessing involved the use of a new Changed Weiner Channel (MWF) filter. From the region of interest (ROI), a collection of 12 features was extracted. The classifiers used were optimised MANFIS, optimised by Artificial Bee Colony (ABC) algorithm, ANFIS classifier, Artificial Neural Network (ANN), Naive Bayes (NB) and Support Vector Machine (SVM). The best result, achieved by ABC-MANFIS, yielded an accuracy of 0.98148. While the study demonstrates high accuracy, its validation in a clinical setting is yet to be conducted.

In [20], a Computer-aided Detection and/or Diagnosis (CAD) system was developed to aid breast cancer diagnosis. It's goal was to retrieve relevant mammogram images based on pathology and leverages these cases to enhance the classifier's performance. The system utilised the Digital Database for Screening Mammography (DDSM) containing 4300 breast cancer images. The decision support system employed local feature descriptors such as Scale-invariant Feature Transform (SIFT), Speeded Up Robust Feature (SURF) and Local Binary Pattern (LBP), which outperformed global features. A weighted K-Nearest Neighbours (KNN) classifier was used, achieving an accuracy of approximately 0.9. The study highlights the improved decision-making potential of Content-based image retrieval (CBIR)-based CAD systems and demonstrates promising retrieval results compared to existing techniques.

In [18], a novel method was proposed for diagnosing and retrieving breast masses. The study addressed the limitations of existing methods in the retrieval stage. The DDSM dataset was utilised, with 11553 ROIs extracted. The ROIs were used to identify masses, and SIFT was employed for feature extraction and search. The experiment employed k-means classification with an accuracy of 90.8% using k=5.

In [30], a method was proposed to assess the risk of breast cancer malignancy in mammograms with microcalcifications (MCs) and perform CBIR. The study used the DDSM dataset, which included 87 mammograms with MCs. The evaluation considered the morphology of MCs according to the BI-RADS scale and utilised shape and texture features for image retrieval using SVM. The CBIR scheme achieved effective retrieval with a mean average precision (MAP) of 0.60. The paper introduced a method that evaluates cancer risk based on BI-RADS classifications, a standardised reference system. However, the SVM's performance was noted as a weakness, suggesting that customised input features could potentially improve retrieval performance for each BI-RADS category.

In [6], a mammography ontology notation was developed to interpret mammograms using specialised vocabulary. The paper proposed the Dokuz Eylul University Mammogram Set (DEMS) dataset, consisting of digital mammograms created for the ontology-based mammogram annotation and retrieval tool (MART). The CBIR system allowed searching a large number of images based on visual similarity. The system was tested with queries in DEMS, and a mathematical model using Semantic Query-Enhanced Web Rule Language (SQWRL) and XQuery was used to calculate lesion similarity. The thesis focused on modelling uncertainty and deducing BI-RADS scores. Experiments were conducted with DEMS and DDSM datasets, evaluating the method with the NB classifier. The results showed accuracy and sensitivity values of 0.86 and 0.60, respectively, for DEMS, and 0.81 and 0.5, respectively, for DDSM.

In [4], an integration method for CBIR was developed and implemented using PostgreSQL as a case study for the mammogram retrieval system. The method introduced breast contour segmentation to accurately determine the contours of breast masses. An ontology-based annotation tool was also designed. Experiments were conducted on two mammogram datasets: DDSM and DEMS. The features included intensity, shape, texture, and margin, representing various aspects of the mass.

The performance evaluation concluded that shape and texture features were the most effective in identifying mass properties. The classifiers used in the experiments included KNN, Random Forest (RF), NB, ANN, Linear Discriminant Analysis (LDA), and SVM. Among them, the NB classifier achieved the best results.

In [12], an SVM-based suspicious tissue pattern classification algorithm was developed for CAD using CBIR. The dataset consisted of 10,509 radiographs, and two experiments were conducted: one with 12 classes and another with 20 classes. In the 12-class problem, three pathology and four tissue density categories were considered, while in the 20-class problem, two distinct types of lesions were distinguished. Features were extracted using two-dimensional principal component analysis (2DPCA) from small patches of 128 x 128 pixels, and SVM was used for classification. The experiments achieved an accuracy of 61.6% for the 12-class problem and 52.1% for the 20-class problem.

In [11], a CAD system for screening mammograms was developed using CBIR methodology. The dataset consisted of 10,509 mammography images collected from various archives, including Lawrence Livermore National Laboratory, DDSM, MIAS. The dataset supported experiments with up to 12 classes and 233 images per class. Feature extraction techniques such as Singular Value Decomposition (SVD), 2DPCA, and Principle Component Analysis (PCA) were employed to characterise breast density and lesions. The 2DPCA technique achieved the highest accuracy of 80.07% for image retrieval. This study contributes to CBIR-CAD of mammograms, providing a system that aids radiologists in diagnosis and classifying breast lesions.

In [25], MammoSVx, a CBIR system, is introduced for categorising breast tissue density and modifying parameters for lesion segmentation and classification. The IRMA reference dataset, comprising data from DDSM, MIAS and Lawrence Livermore National Laboratory (LLNL), contains over 10,000 distinct mammograms. Feature extraction and selection are performed using SVD and histograms, enabling characterisation of breast density based on image texture. The SVM determined pattern similarity to categorise BI-RADS tissue into four categories. The experimental results showed an average accuracy of 82.14%. This study demonstrated the potential of using SVD and SVM in a CBIR system for breast density characterisation, supporting radiologists in their diagnosis.

In [19], a method for CAD based on image retrieval was developed to design case-adaptive classifiers. The dataset used, DDSM, includes 589 mammogram images, with 331 benign breast cancer and 258 malignant cases. Image retrieval initially identified relevant images based on diagnostic case characteristics. These cases were then used to optimise a classifier, aiming to improve the classification accuracy for breast cancer cases. Logistic regression and a non-linear SVM classifier were employed, with the non-linear SVM achieving the highest accuracy of 0.8. The article highlighted the realistic approach of quantifying lesions through an algorithm rather than relying on manual detection of features.

In [31], the aim was to compare and identify the most applicable approaches for CBIR-based CAD schemes. The dataset used consists of 1,500 reference images of verified masses and 1,500 ROIs of false-positive masses collected from a university hospital's Radiology Department. Evaluation indices such as visual similarity and clinical relevance were used to assess the CAD algorithm performance using CBIR approaches. According to the discussed previous studies, commercially available CADs perform well in microcalcifications and clusters detection, but struggle with mass detection. The visual aid provided by CAD applications can be beneficial for radiologists. However, CAD development using CBIR approaches still faces challenges in feature selection, accurate region segmentation, and assembling the optimal reference dataset.

The literature on advanced CAD system techniques, including mass detection, calcification, bilateral asymmetry, architectural distortion, and image retrieval, was thoroughly reviewed by Tang et al. [29] up to 2009. The two most popular algorithms among the numerous works referenced are SVM and KNN. The authors emphasised the value of CAD systems and drew attention to the need to increase their performance in order to support clinical applications.

In [24], a CBIR system for mammography was developed and evaluated. The dataset consisted of 200 mammogram images, including 58 benign and 46 malignant breast cancer images, obtained from the Department of Radiology at the University of Chicago. The study compared the performance of SVM classifier and Adaptive SVM (Ada-SVM), with a focus on the use of discriminant adaptive nearest neighbours (DANN) for image retrieval. Ada-SVM achieved the highest accuracy of 0.7925 (Ada-SVM with N=9), resulting in reduced generalisation error. However, a weakness of the study is the lack of clinical evaluations, which could have provided valuable insights into the approach's effectiveness.

In [16], an approach was proposed to address the challenge of defining relevance or similarity in CBIR. The authors investigate the use of neural network(NN) and SVM to predict user's notion of similarity. 76 mammograms with clustered MCs from the Chicago dataset were used. The retrieval system's performance was evaluated using precision-recall curves through cross-validation. The proposed hierarchical learning approach involved a binary classifier and a regression module, aiming to optimise retrieval effectiveness and efficiency. The first phase screens out images that significantly differ from the query image. The surviving images from the first phase were then compared to the query in the second phase to quantitatively measure their similarity. Four combinations of classifiers were evaluated, with learning-based networks outperforming Euclidean distance-based approaches. The MSVM-SVM combination achieved the best performance among the evaluated combinations.

A summary is given in Table 1. In short, the above studies have contributed positively with an enormous amount of knowledge on the subject under study.

3 Methodology

This section presents the methodology adopted in the course of this project. The methods applied in each iteration of the development of this work consist of three main steps (see Figure 2). The first step involved building several models using four different pre-trained networks, which allow the classification of mammograms according to different dimensions, namely, the existence of masses, the existence of calcifications, the existence of distortions, breast density, BI-RADS classification, laterality, view and the existence of asymmetries (Section 3.2). In the second step, the model that classified mammograms most accurately based on the various dimensions was selected. The third and final phase involved creating the final retrieval model using the previously built models; more specifically, the final model was created by combining these models using a weighted sum (Section 3.3). This section also includes the description of the experimental set-up (Section 3.3) and the used database (Section 3.1).

3.1 Database

The INBreast database [23] has a total of 410 mammogram images. The images have a size of 3328 $\times$ 4084 or 2560 $\times$ 3328 pixels. The size of the pixel is 70 microns and the contrast resolution is 14 bits. The images were saved in Digital Imaging and Communications in Medicine (DICOM) format. All confidential medical information was removed from the DICOM file. This dataset's main feature

Table 1. Literature review of retrieval works in breast cancer

Paper	Main Goal	Methods	Database	Main Results
[26]	Retrieval	ABC-MANFIS, ANFIS, ANN, NB, SVM	Mini-MIAS	ABC-MANFIS with accuracy of 0.98148
[20]	CAD based on CBIR	SIFT, SURF, and LBP descriptors	DDSM	Accuracy ≈ 0.9 using LBP
[18]	CAD based on CBIR	SIFT, k-means	DDSM	Accuracy of 90.8% with k=5
[30]	CBIR	SVM with shape and texture features	DDSM	Average accuracy $[0.57 - 0.60]$
[6]	Ontology-based annotation; CBIR	NB	DEMS and DDSM	Accuracy of 0.86 for DEMS and 0.81 for DDSM
[4]	CBIR	Mass segmentation, 26 low-level features, 6 classifiers	DDSM and DEMS	Best classifier: NB
[12]	CAD based on CBIR	2DPCA, SVM	Merged public datasets	Accuracy 61.6% (12-class); 52.1% (20-class)
[11]	CAD based on CBIR	2DPCA, SVD, SVM	Merged public datasets	80.07% using 2DPCA
[25]	CBIR for tissue density	SVD feature extraction, SVM	IRMA	Average accuracy of 82.14%
[19]	CAD	Linear and non-linear SVM	DDSM	Non-linear SVM with accuracy of 0.8
[31]	CBIR-based CAD comparison	n.a.	n.a.	CAD development needs further research
[29]	Literature review of CAD techniques	n.a.	n.a.	CAD performance needs improvement
[24]	CBIR	SVM and Ada-SVM	Chicago dataset	Accuracy of 0.7925
[16]	CBIR	ANN and SVM	Chicago dataset	Similarity learned from human observers

is the annotations used as ground truth for the individual classifiers described in Section 3.2. These were made by an expert in the field and validated by a second expert between April 2010 and December 2010. Seven annotations are present, including asymmetries, clusters/microcalcifications, masses, distortions, spiculated regions, and pectoral muscles. The patient's age at the time of image acquisition, family history, breast density annotation, and BI-RADS categorisation are all also mentioned. This database was used for various purposes, including training models for distinguishing between malignant and benign breast tumours [13], developing methods for calcification detection [14], deriving algorithms for delineating the pectoral muscle [15], among others.

3.2 Classifiers

Four pre-trained networks were tested, ResNet50, VGG16, InceptionV3 and InceptionResNetV2. Thus, four models were created for each dimension (masses, calcifications, distortions, breast density, BI-RADS, laterality, view and asymmetries), each with a different pre-trained network. In total, 32 models were created.

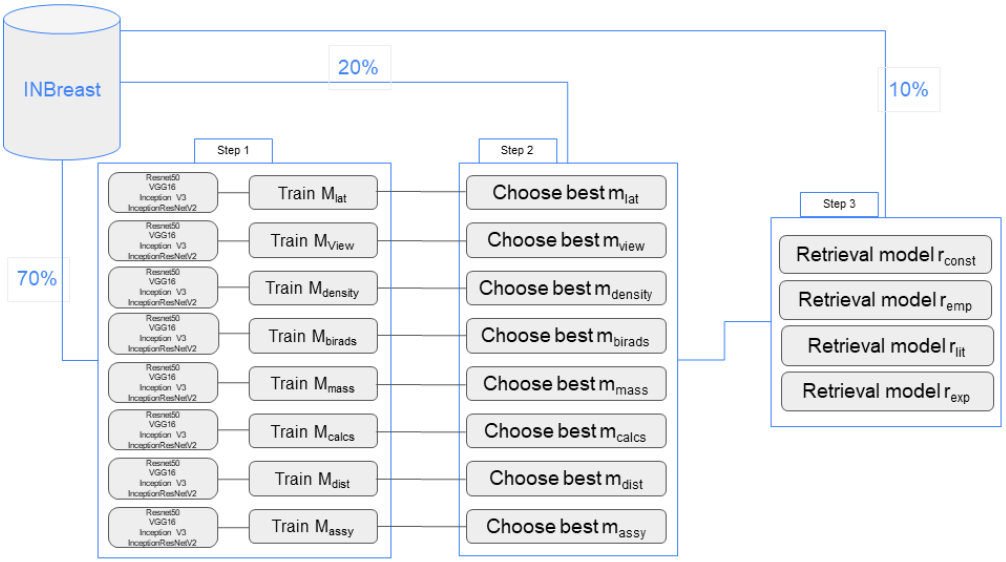


Fig. 2. Overview of the training methodology

Initially, the preprocessing of the images was performed, the images were resized to 224×224 pixels and converted from RGB to BGR. Next, each colour channel is zero-centred concerning the ImageNet dataset, without scaling.

Two approaches were adopted in each model:

- In the first approach, all layers of the network were frozen, except for the final fully connected layers. By keeping the pre-trained weights intact and only fine-tuning the final layers, the learned features are leveraged, while adapting the model to each specific task;
- In the second approach, the entire model was trained with a low learning rate. By allowing the entire network to be updated, the potential for further improvements over the previous approach is enabled.

Optimisation was performed using the Adam optimiser and Sparse Categorical Cross-entropy loss. After that, the models were compiled and trained for 25 epochs.

3.3 Retrieval model

The retrieval model used the individual models described in Section 3.2. It is a weighted sum and has the following generic equation:

$d^{probe,i} = w_{lat} * d_{lat}^{probe,i} + w_{view} * d_{view}^{probe,i} + w_{density} * d_{density}^{probe,i} + w_{birads} * d_{birads}^{probe,i} + w_{mass} * d_{mass}^{probe,i} + w_{calcs} * d_{calcs}^{probe,i} + w_{dist} * d_{dist}^{probe,i} + w_{assy} * d_{assy}^{probe,i}$,

$i = 1 : 410 \backslash \{probe\}$

$r_{const} = 1$

where w_{dim} is the weight attributed for each dimension (laterality, CC or MLO view, breast density, BI-RADS class, existence or not of masses, existence or not of calcifications, existence or not of distortions, and existence or not of asymmetry).

Different distance metrics can be chosen to be included in the above formula. The present implementation behaves accordingly to:

$$d_{dim}^{probe,i} = \begin{cases} 1, if\ \tilde{V}_{dim}^{probe} == V_{dim}^{i} \\ 0, else \end{cases}$$

where dim is each of the considered targets (dimensions), V_{dim}^{i} is the ground truth value of the attribute as given in the INBreast database (see Section 3.1) for the image i and $\tilde{V}_{dim}^{probe}$ is the result of the classifier (see Section 3.2) when applied to the probe image.

The above distance is computed between the probe image and all of the 410 images in the INBreast dataset, excluding the probe image itself.

In conducting this study, Tensorflow, version 2.11.0, was used. The programming environment for importing the libraries was Google Colab with the NVIDIA T4 Tensor Core GPU. Data was split into 70% for training, 20% for validation, and 10% for testing.

4 Results

This section is subdivided into two parts. In the first part (Section 4.1), the evaluation of the individual models for each dimension is presented. The second part (Section 4.2) illustrates the results achieved by the final retrieval models.

4.1 Individual classifiers

For experimental purposes, two types of approaches (see Section 3.2) were considered:

- Training only the final fully connected layers (trainable = false);
- Training the networks from scratch (trainable = true).

In Figures 3, 4, 5, and 6, the metrics “Loss” and “Accuracy” refer to the training set, while “Val_accuracy” and “Val_loss” are results obtained in the validation set. In Figure 7 the metrics “Loss” and “Accuracy” refer to the test set.

Based on the obtained results (Figure 3), it can be concluded that the InceptionV3 pre-trained network achieved the highest accuracy of 96% among the models used for classifying mammograms based on the BI-RADS dimension. For the breast density dimension (ACR), the network that obtained the best performance was InceptionResNetV2 with an accuracy of 98%.

Regarding the results obtained by the models that classify the mammograms according to the existence or not of masses (Figure 4), it is possible to say that the pre-trained network InceptionV3 was the one that obtained the best results with an accuracy of 100%. The network that achieved the greatest results for the dimension calcification was InceptionV3, which also had a 100% accuracy rate.

From the results obtained by the models that classify the mammograms according to the view (Figure 5), it is possible to deduce that the pre-trained network InceptionResNetV2 was the one that obtained the best results with an accuracy of 100%. With a 100% accuracy rate, InceptionV3 achieved the best performance for the laterality dimension.

	BI-RADS				ACR			
	Loss	Accuracy	Val_loss	Val_accuracy	Loss	Accuracy	Val_loss	Val_accuracy
	RESNET50				RESNET50			
treinable = false	1.4491	0.5041	1.4147	0.4945	1.3538	0.3415	1.2747	0.3556
treinable = true	0.5730	0.8211	1.5515	0.4945	0.3987	0.9024	1.3151	0.3556
	VGG16				VGG16			
treinable = false	1.4504	0.4553	1.4198	0.4945	1.3646	0.3496	1.2698	0.3556
treinable = true	1.4498	0.5122	1.4100	0.4945	1.3348	0.3008	1.2512	0.3556
	InceptionResNetV2				InceptionResNetV2			
treinable = false	1.4226	0.4797	1.4109	0.4945	1.3215	0.3171	1.2650	0.3556
treinable = true	0.3047	0.9024	1.5363	0.4945	0.0897	0.9837	1.4237	0.3778
	InceptionV3				InceptionV3			
treinable = false	1.3807	0.4797	1.3740	0.4945	1.3523	0.3496	1.2738	0.3333
treinable = true	0.1242	0.9593	1.5639	0.2088	0.0884	0.9593	1.6155	0.3111

Fig. 3. BI-RADS and breast density results.

	Mass				Calcifications			
	Loss	Accuracy	Val_loss	Val_accuracy	Loss	Accuracy	Val_loss	Val_accuracy
	RESNET50				RESNET50			
treinable = false	0.6069	0.7236	0.6392	0.6703	0.6096	0.7398	0.5567	0.7692
treinable = true	0.6663	0.8130	0.8439	0.6703	0.0665	0.9837	0.5566	0.7692
	VGG16				VGG16			
treinable = false	0.6201	0.7154	0.6353	0.6703	0.6164	0.7236	0.5840	0.7692
treinable = true	0.6067	0.7236	0.6368	0.6703	0.6288	0.7398	0.5489	0.7692
	InceptionResNetV2				InceptionResNetV2			
treinable = false	0.5896	0.7317	0.6707	0.6703	0.6199	0.7317	0.5418	0.7692
treinable = true	0.0619	0.9756	0.6443	0.6703	0.0208	1.0000	0.6229	0.7692
	InceptionV3				InceptionV3			
treinable = false	0.5597	0.7317	0.6465	0.6703	0.5604	0.7398	0.5219	0.7692
treinable = true	0.0135	1.0000	0.8536	0.6703	0.0139	1.0000	0.5798	0.7692

Fig. 4. Mass and calcifications results.

	View				Laterality			
	Loss	Accuracy	Val_loss	Val_accuracy	Loss	Accuracy	Val_loss	Val_accuracy
	RESNET50				RESNET50			
treinable = false	0.7305	0.4146	0.6898	0.6889	0.7069	0.5203	0.6918	0.4945
treinable = true	0.0580	0.9756	0.7659	0.4778	0.0027	1.0000	0.8133	0.5275
	VGG16				VGG16			
treinable = false	0.7242	0.5122	0.7005	0.4778	0.7806	0.4309	0.6942	0.4725
treinable = true	0.6945	0.4959	0.6339	0.4778	0.6929	0.5122	0.6926	0.5275
	InceptionResNetV2				InceptionResNetV2			
treinable = false	0.7070	0.4878	0.6884	0.5667	0.6911	0.5447	0.6933	0.4725
treinable = true	0.0074	1.0000	0.6876	0.5222	0.0081	1.0000	0.7826	0.4725
	InceptionV3				InceptionV3			
treinable = false	0.6418	0.6423	0.6244	0.8000	0.4107	0.8862	0.3775	0.9560
treinable = true	0.0079	1.0000	0.6493	0.6222	0.0047	1.0000	0.6349	0.5934

Fig. 5. View and laterality results.

With regard to the classification based on dimensional asymmetries, it is verified that the pre-trained ResNet50 network is able to obtain the best results, with an accuracy of 100% (Figure 6). Among the tested networks, InceptionV3 demonstrated superior performance in handling dimensional distortions, achieving a remarkable accuracy of 100%.

	Asymmetries				Distortions			
	Loss	Accuracy	Val_loss	Val_accuracy	Loss	Accuracy	Val_loss	Val_accuracy
	RESNET50				RESNET50			
treinable = false	0.2417	0.9431	0.1806	0.9560	0.0674	0.9919	0.1194	0.9780
treinable = true	0.0110	1.0000	0.2132	0.9560	0.0123	0.9919	0.1549	0.9780
	VGG16				VGG16			
treinable = false	0.2356	0.9350	0.1819	0.9560	0.0569	0.9919	0.1307	0.9780
treinable = true	0.2322	0.9431	0.2287	0.9560	0.0574	0.9919	0.1217	0.9780
	InceptionResNetV2				InceptionResNetV2			
treinable = false	0.2677	0.9431	0.1800	0.9560	0.0558	0.9919	0.1160	0.9780
treinable = true	0.0429	0.9837	0.4145	0.9560	0.0078	1.0000	0.1907	0.9780
	InceptionV3				InceptionV3			
treinable = false	0.2507	0.9431	0.1836	0.9560	0.0802	0.9919	0.1262	0.9780
treinable = true	0.0112	1.0000	0.1743	0.9560	0.0047	1.0000	0.1339	0.9780

Fig. 6. Asymmetries and distortions results.

Figure 7 illustrates the test set's outcomes. For the BiRads dimension, the networks ResNet50 and VGG16 were able to obtain the best results with an accuracy of 66%. For the breast density dimension (represented by ACR in Figure 7) it is the VGG16 network that obtains the best result with an accuracy of 41%. For the dimension asymmetries, the network that obtained the best result was VGG16 with an accuracy of 100%, and for the dimension distortions, it is ResNet5 that obtained the best performance, with an accuracy of 100%. For the dimension masses, the network ResNet50 obtained the best result with 81% accuracy. For the dimension calcification and view the inceptionResNetV2 network is the one that obtains the best performance with an accuracy of 81% and 94%, respectively. Finally, for the dimension laterality, it is the VGG16 network that obtains the best result with an accuracy of 44%. Overall, the test results were reasonable, having the networks behaved differently in relation to each dimension.

BI-RADS		ACR		Asymmetries		Distortions	
Loss	Accuracy	Loss	Accuracy	Loss	Accuracy	Loss	Accuracy
RESNET50		RESNET50		RESNET50		RESNET50	
1.5804	0.6562	1.5725	0.3750	0.0690	1.0000	0.0011	1.0000
VGG16		VGG16		VGG16		VGG16	
1.5006	0.6562	1.4786	0.4062	0.3195	1.0000	0.0998	1.0000
InceptionResNetV2		InceptionResNetV2		InceptionResNetV2		InceptionResNetV2	
2.6745	0.6562	2.3488	0.2812	7.0141	1.0000	3.1124	1.0000
InceptionV3		InceptionV3		InceptionV3		InceptionV3	
1.8360	0.5000	4.0283	0.2812	4.4703	1.0000	1.0803	1.0000
Masses		**Calcification**		**View**		**Laterality**	
Loss	Accuracy	Loss	Accuracy	Loss	Accuracy	Loss	Accuracy
RESNET50		RESNET50		RESNET50		RESNET50	
0.5750	0.8125	0.7292	0.7812	0.9862	0.5312	0.7816	0.4375
VGG16		VGG16		VGG16		VGG16	
0.7831	0.2187	0.6802	0.5937	1.1091	0.5312	0.6956	0.4062
0.6334	0.8125	0.5382	0.7812	0.9625	0.4687	0.7080	0.4375
InceptionResNetV2		InceptionResNetV2		InceptionResNetV2		InceptionResNetV2	
0.8398	0.6875	0.8050	0.8125	0.2266	0.9375	1.0682	0.5625
InceptionV3		InceptionV3		InceptionV3		InceptionV3	
1.7035	0.8125	0.9828	0.7812	1.5004	0.4687	3.6191	0.5000

Fig. 7. Results obtained in the test set.

4.2 Retrieval models

Four versions of the weights were attempted:

- In the model r_{const}, all of the weights are equal.
- In the model r_{emp}, the weight values were empirically defined as: $w_{lat} = 0.000$, $w_{view} = 0.000$, $w_{density} = 0.167$, $w_{birads} = 0.167$, $w_{mass} = 0.167$, $w_{calcs} = 0.167$, $w_{dist} = 0.167$, $w_{assy} = 0.167$.
- The model r_{lit} was based on existing studies. For w_{lat}, the value 0.018 was set based on findings from [1], which suggest a more aggressive biology in the left breast. The value 0.000 was assigned to w_{view} due to a lack of references. For $w_{density}$, the value 0.172 was chosen as mammographic breast density, as indicated by [2], is a significant reproducible risk factor for breast cancer. w_{birads} was also set to 0.172 based on [22], which highlights the standardised terminology and reduced inter-observer variability provided by BI-RADS. According to [10], assigning a weight of 0.172 to w_{mass} allows for accurate and rapid diagnosis and prognosis by directly visualising mass characteristics. w_{calcs} was assigned a value of 0.172, as suggested by [8], indicating the effectiveness of calcification identification in detecting breast cancer. As architectural distortion represents 6% of abnormalities in screening mammography, as stated in [9], a weight of 0.172 was assigned to w_{dist}. Finally, according to [17], pairing opposite breasts to examine asymmetry improves lesion detection, leading to a value of 0.122 for w_{assy}.
- The creation of the model r_{exp} involved information gathered from an specialist. Therefore, the following weights were defined: $w_{lat} = 0.143$, $w_{view} = 0.140$, $w_{density} = 0.098$, $w_{birads} = 0.097$, $w_{mass} = 0.132$, $w_{calcs} = 0.118$, $w_{dist} = 0.142$, $w_{assy} = 0.130$.

The weights for each one of the four versions are summarised in Table 2.

Table 2. Summary of weights in different models

Model	w_{lat}	w_{view}	$w_{density}$	w_{birads}	w_{mass}	w_{calcs}	w_{dist}	w_{assy}
r_{const}	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
r_{emp}	0.000	0.000	0.167	0.167	0.167	0.167	0.167	0.167
r_{lit}	0.018	0.000	0.172	0.172	0.172	0.172	0.172	0.122
r_{exp}	0.143	0.140	0.098	0.097	0.132	0.118	0.142	0.130

For an illustration of the retrieval results, the probe image shown in Figure 8 was used, as well as the ground truth for each dimension.

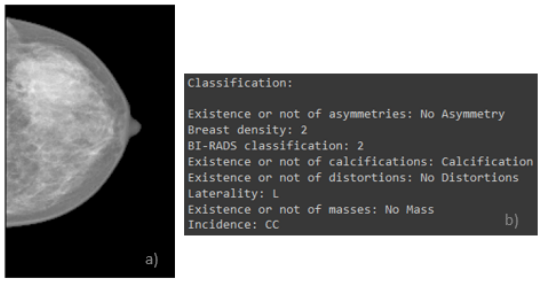


Fig. 8. Probe image along with the ground truth (ground truth is not used during retrieval)

In Figure 9 it is possible to visualise the top-5 retrieved mammograms by the model r_{const}. In this first approach, one would expect the dimensions laterality, view, mass, calcification and distortions to be the same as in the probe image, since in the distance calculation the weights have the same value, the same importance. It is clear that the use of a different distance metric (left for future work) would improve these results.

Fig. 9. Top-5 images retrieved by r_{const}

In Figure 10 it is possible to visualise the top-5 retrieved mammograms by the model r_{emp}. In this approach, since the weights that make up the distance formula have been defined empirically, it is possible to realise that laterality and view do not have the same importance.

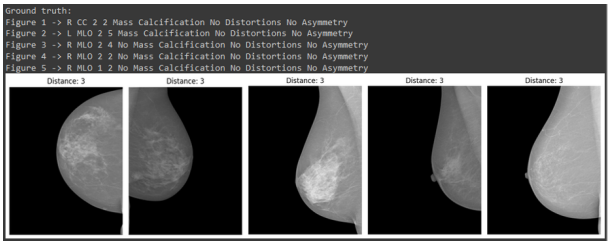


Fig. 10. Top-5 images retrieved by r_{emp}

In Figure 11 it is possible to visualise the top-5 retrieved mammograms by the model r_{lit}. In this approach, the weights were defined based on the literature, so the five retrieved mammograms are the result of this research.

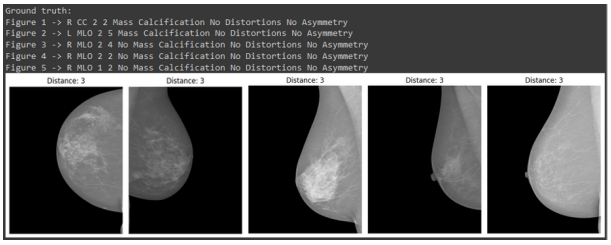


Fig. 11. Top-5 images retrieved by r_{lit}

Finally, in Figure 12 it is possible to visualise the top-5 retrieved mammograms by the model r_{lit}. In this approach, which is based on expert opinion, the five mammograms displayed are expected to be the most relevant when given the probe image in Figure 8. The specialist's information is somewhat constrained because the work was not done under ideal circumstances, which is one of this approach's shortcomings. For instance, using a mammography reading screen would make it easier to understand each component of the mammograms.

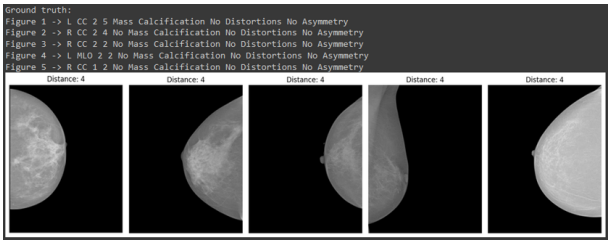


Fig. 12. Top-5 images retrieved by r_{exp}

5 Conclusions and directions for Future Work

Breast cancer is a widespread illness that affects a large number of people, with women being more at risk for developing it. Despite advancements in medical knowledge and treatments, there is still a critical need for extensive research in this field. This urgency arises from the fact that breast cancer continues to claim the lives of thousands of individuals each year. By investing in further research efforts, we can strive to improve prevention, early detection, and treatment strategies, ultimately reducing the devastating impact of this disease on countless lives. The present project is related to the aforementioned problem since early detection of breast cancer is essential for a favourable prognosis.

A retrieval model has been developed in which an image is an input into the system and a set of relevant images with known diagnoses and their respective histories are returned. Towards this end, multiple models were created, one for each dimension including breast density, asymmetries, BI-RADS classification, presence of calcifications, distortions, laterality (right or left breast), presence of masses, and image incidence (Cranio-Caudal or Medio-Lateral Oblique). Each model was trained with four pre-trained networks: ResNet50, VGG16, InceptionV3 and InceptionResNetV2, in order to understand which one performed best. To create the final model retrieval model, we proposed the use of a weighted sum of the aforementioned models. Four versions were then attempted, a first version where all the dimensions are equally important, a second version where each dimension was empirically weighted, a third where the weights were defined according to the literature, and a final one where the values of the weights were defined by a specialist.

As for the results obtained by the individual classification models in the test phase, according to each dimension, the best accuracy achieved was 66% for BI-RADS classification, 41% for breast density, 98% for the presence of masses, 81% for the presence of calcifications, 94% for view, 44% for laterality, 100% for asymmetries, and 100% for distortions. The final retrieval models behaved differently since the weights attributed to each dimension were defined in different ways, i.e. each

dimension had a specific importance. A quantitative evaluation of these approaches is left for future work, but we expect the model r_{exp} to achieve the best results.

Several future directions are foreseen. The main ones are focused on the retrieval model. We are, for instance, planning to experiment different distance metrics. Another line to pursue is to automatically derive the weights of each dimension in the final weighted sum model. To achieve this objective, we can employ artificial intelligence techniques such as the training of neuronal networks or the use of evolutionary algorithms such as genetic algorithms, or Particle Swarm Optimisation type algorithms. To do so, we will need to acquire ground truth to train these models. To do so, we will need to collect ground truth in order to train these models. To collect this information, an application is being created that will be used by an expert [27, 28].

References

1. Abdou, Y., Gupta, M., Asaoka, M., Attwood, K., Mateusz, O., Gandhi, S., Takabe, K.: Left sided breast cancer is associated with aggressive biology and worse outcomes than right sided breast cancer. Scientific Reports **12**(1), 13377 (2022)
2. Anandarajah, A., Chen, Y., Colditz, G.A., Hardi, A., Stoll, C., Jiang, S.: Studies of parenchymal texture added to mammographic breast density and risk of breast cancer: a systematic review of the methods used in the literature. Breast Cancer Research **24**(1), 101 (2022)
3. Arnold, M., Morgan, E., Rumgay, H., Mafra, A., Singh, D., Laversanne, M., Vignat, J., Gralow, J.R., Cardoso, F., Siesling, S., Soerjomataram, I.: Current and future burden of breast cancer: Global statistics for 2020 and 2040. The Breast **66**, 15–23 (2022)
4. Berber, T.: Integration of content-based image retrieval and database management system: A case study with digital mammography. Thesis, DEÜ Fen Bilimleri Enstitüsü (2013)
5. Bessa, S., Domingues, I., Cardosos, J.S., Passarinho, P., Cardoso, P., Rodrigues, V., Lage, F.: Normal breast identification in screening mammography: a study on 18 000 images. In: IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 325–330 (2014)
6. Bulu, H.: Ontology-based medical image annotation and retrieval. Thesis, DEÜ Fen Bilimleri Enstitüsü (2013)
7. Cancro, L.P.C.o.: Cancro da Mama : Liga Portuguesa Contra o Cancro, `http://www.ligacontracancro.pt/cancro-da-mama/`, access date: 2023-01-31
8. Chaudhury, S., Rakhra, M., Memon, N., Sau, K., Ayana, M.T.: Breast Cancer Calcifications: Identification Using a Novel Segmentation Approach. Computational and Mathematical Methods in Medicine **2021**, e9905808 (2021)
9. Chen, X., Zhang, Y., Zhou, J., Wang, X., Liu, X., Nie, K., Lin, X., He, W., Su, M.Y., Cao, G., Wang, M.: Diagnosis of architectural distortion on digital breast tomosynthesis using radiomics and deep learning. Frontiers in Oncology **12** (2022)
10. Cuypers, E., Claes, B.S.R., Biemans, R., Lieuwes, N.G., Glunde, K., Dubois, L., Heeren, R.M.A.: 'On the Spot' Digital Pathology of Breast Cancer Based on Single-Cell Mass Spectrometry Imaging. Analytical Chemistry **94**(16), 6180–6190 (2022)
11. Deserno, T., Soiron, M., Oliveira, J., Araujo, A.: Towards Computer-Aided Diagnostics of Screening Mammography Using Content-Based Image Retrieval. In: 24th SIBGRAPI Conference on Graphics, Patterns and Images. pp. 211–219 (2011)
12. Deserno, T.M., Soiron, M., Oliveira, J.E.E.d., Araújo, A.d.A.: Computer-aided diagnostics of screening mammography using content-based image retrieval. In: Medical Imaging: Computer-Aided Diagnosis. vol. 8315, pp. 647–655. SPIE (2012)
13. Domingues, I., Sales, E., Cardoso, J., Pereira, W.: Inbreast-database masses characterization. In: XXIII Congresso Brasileiro de Engenharia Biomédica (CBEB) (2012)

14. Domingues, I., Cardoso, J.S.: Using Bayesian surprise to detect calcifications in mammogram images. In: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 1091–1094 (2014)
15. Domingues, I., Cardoso, J.S., Amaral, I., Moreira, I., Passarinho, P., Santa Comba, J., Correia, R., Cardoso, M.J.: Pectoral muscle detection in mammograms based on the shortest path with endpoints learnt by SVMs. In: Annual International Conference of the IEEE Engineering in Medicine and Biology. pp. 3158–3161 (2010)
16. El-Naqa, I., Yang, Y., Galatsanos, N., Nishikawa, R., Wernick, M.: A similarity learning approach to content-based image retrieval: application to digital mammography. IEEE Transactions on Medical Imaging **23**(10), 1233–1244 (2004)
17. Guan, Y., Wang, X., Li, H., Zhang, Z., Chen, X., Siddiqui, O., Nehring, S., Huang, X.: Detecting Asymmetric Patterns and Localizing Cancers on Mammograms. Patterns **1**(7), 100106 (2020)
18. Jiang, M., Zhang, S., Li, H., Metaxas, D.N.: Computer-Aided Diagnosis of Mammographic Masses Using Scalable Image Retrieval. IEEE Transactions on Biomedical Engineering **62**(2), 783–792 (2015)
19. Jing, H., Yang, Y.: Image retrieval for computer-aided diagnosis of breast cancer. In: IEEE Southwest Symposium on Image Analysis & Interpretation (SSIAI). pp. 9–12 (2010)
20. Kiruthika, K., Vijayan, D., R., L.: Retrieval driven classification for mammographic masses. In: International Conference on Communication and Signal Processing (ICCSP). pp. 0725–0729 (2019)
21. Magny, S.J., Shikhman, R., Keppke, A.L.: Breast imaging reporting and data system. In: StatPearls. StatPearls Publishing (2023), http://www.ncbi.nlm.nih.gov/books/NBK459169/
22. Meng, L., Zhao, X., Guo, J., Lu, L., Cheng, M., Xing, Q., Shang, H., Zhang, B., Chen, Y., Zhang, P., Zhang, X.: Improved Differential Diagnosis Based on BI-RADS Descriptors and Apparent Diffusion Coefficient for Breast Lesions: A Multiparametric MRI Analysis as Compared to Kaiser Score. Academic Radiology (2023)
23. Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S.: Inbreast: toward a full-field digital mammographic database. Academic radiology **19**(2), 236–248 (2012)
24. Nishikawa, R.M., Wei, L., Yang, Y.: Retrieval-driven microcalcification classification for breast cancer diagnosis. In: 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro. pp. 1260–1263 (2007)
25. Oliveira, J., Araújo, A., Deserno, T.: Content-based image retrieval applied to BI-RADS tissue classification in screening mammography. World Journal of Radiology **3**(1), 24–31 (2011)
26. Rayen, S.J., Subhashini, R.: An Efficient Mammogram Image Retrieval System Using an Optimized Classifier. Neural Processing Letters **53**(4), 2467–2484 (2021)
27. Roriz, C., Vasconcelos, V., Domingues, I.: Low-code application for ground truth data acquisition in mammogram retrieval. In: 29th edition of the Portuguese Conference on Pattern Recognition (RECPAD) (2023 (to appear))
28. Roriz, C.: Mammogram retrieval. Master's thesis, Instituto Superior de Engenharia de Coimbra (Portugal) (2023 (to appear))
29. Tang, J., Rangayyan, R.M., Xu, J., Naqa, I.E., Yang, Y.: Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances. IEEE Transactions on Information Technology in Biomedicine **13**(2), 236–251 (2009)
30. Tsochatzidis, L., Zagoris, K., Savelonas, M., Papamarkos, N., Pratikakis, I., Arikidis, N., Costaridou, L.: Microcalcification oriented content-based mammogram retrieval for breast cancer diagnosis. In: IEEE International Conference on Imaging Systems and Techniques (IST). pp. 257–262 (2014)
31. Zheng, B.: Computer-Aided Diagnosis in Mammography Using Content-Based Image Retrieval Approaches: Current Status and Future Perspectives. Algorithms **2**(2), 828–849 (2009)

Annex C - RecPad Extended Abstract

Low-Code Application for Ground Truth Data Acquisition in Mammogram Retrieval

Cátia Roriz[1]
a21270590@isec.pt
Verónica Vasconcelos[1]
veronica@isec.pt
Inês C. Moreira[2]
icm@ess.ipp.pt
Inês Domingues[13]
ines.domingues@isec.pt

[1] Instituto Politécnico de Coimbra, Instituto Superior de Engenharia, Rua Pedro Nunes Quinta da Nora, 3030-199 Coimbra, Portugal
[2] Centro Hospitalar Universitário de São João Escola Superior de Saúde do Politécnico do Porto Cintesis-FMUP
[3] Centro de Investigação do Instituto Português de Oncologia do Porto (CI-IPOP): Grupo de Física Médica, Radiobiologia e Protecção Radiológica

Abstract

Breast cancer is a significant global health concern, affecting thousands of individuals, primarily women, with estimated cases expected to climb by 2040. This paper describes the creation of an application to collect ground truth data to aid engineers in the development of a mammography retrieval system. The application is built upon OutSystems, a low-code application platform. Key features of the application include allowing experts to view probe images and associate them with relevant images from the database. Additionally, the platform allows image filtering based on eigth mammogram dimensions. While the ultimate goal is to create a system for medical specialists, the current platform represents a step in the process, facilitating the acquisition of ground truth.

1 Introduction

Breast cancer is a malignant disease that affects thousands of individuals worldwide, primarily women. By the year 2040 [1], a significant increase in the number of breast cancer cases is estimated, with more than 3 million new cases annually, representing a 40% increase. In addition, the number of deaths related to this disease is expected to increase by 50%, exceeding 1 million deaths per year. Although various approaches have been developed to deal with this problem, much more effort has to be done to fight this disease.

The main objective of this work is to develop an initial version of a platform that will serve as a tool to gather data that will facilitate the development of a mammogram retrieval system [4]. The OutSystems platform was used to develop this application [2].

The remainder of this document is organised as follows. This section presents an introductory reflection on the subject. Section 2 provides a brief overview of existing software. Section 3 describes the application requirements and use cases, while Section 5 presents the performed implementation. Finally, Section 6 addresses the final considerations and future work.

2 Available Mammogram Retrieval Software

The retrieval of relevant mammograms is the ultimate goal of the software that will be created. Table 1 shows some competing software. In contrast to the software provided in Table 1, the created software thus far allows experts to submit ground truth data for the later building of retrieval models [4]. To the best of our knowledge, no tools are currently available for this purpose.

Table 1: Competing Software

	PenFetch [3]	**Mammogram Image Retrieval System [5]**
Capabilities	Automates fetching previous medical studies from PACS[1]	Users can send RF[2] to improve system understanding of information needs
Methods	Uses DICOM[3] worklists to identify patients and images, requesting transfers based on predefined rules	Uses text-based methods for initial search and augments outcomes with visual aids
Advantages	Streamlines workflow by sending pertinent images to radiologists' stations ahead of time, minimising delays	Applications include computer-aided diagnosis, medical education, and research
Disadvantages	Dependency on DICOM Worklists. Limited to PACS environments	Specific focus on certain domains may limit broader applicability

[1] PACS: Picture Archiving and Communication System
[2] RF: Relevance Feedback
[3] DICOM: Digital Imaging and Communications in Medicine

3 Requirements analysis

To be able to develop the application, an initial list of requirements was identified:

ID01- View probe image: The software must allow the visualisation of the probe image that will later be analysed and associated with a relevant image from the database.

ID02- View images in the database: The software shall allow the visualisation of the various images of the database, images that are tagged according to the several dimensions of the mammograms.

ID03- Filter images based on dimensions: The software should allow the filtering of the images in the database according to their dimensions such as breast density, existence or not of asymmetries, Breast Imaging-Reporting and Data System (BI-RADS), existence or not of calcifications, existence or not of distortions, laterality (right or left breast), existence or not of masses and image incidence (Cranio-Caudal (CC) or Medio-Lateral Oblique (MLO)).

ID04- Select an image from the database: For each probe image, the software must allow the selection of one image from the database, the one considered by the expert as the most relevant.

ID05- Visualise registrations: The software must allow the visualisation of all the registrations made in the application.

During the development of the software, several improvements were suggested by the expert. These include:

ID06- The probe image cannot be associated with it itself:: The software must ensure that the probe image does not appear in the database images to be selected.

ID07- The probe image must always be visible: The software should allow the probe image to be visible at all times, i.e. it should always remain visible while selecting the image from the database.

ID08- Zoom for Enhanced Viewing: The software should allow zooming in on images, both probe images and database images.

ID09- Navigation to Next Probe Image: The software should allow to skip the current probe image and move on to the next probe image while still having access to the prior probe image.

ID10- Filter selection confirmation: The software should include a functionality to generate a warning, allowing for verification of whether the filter selection has been finalized or not.

The requirements are pictured in the Use Case Diagram of Figure 1.

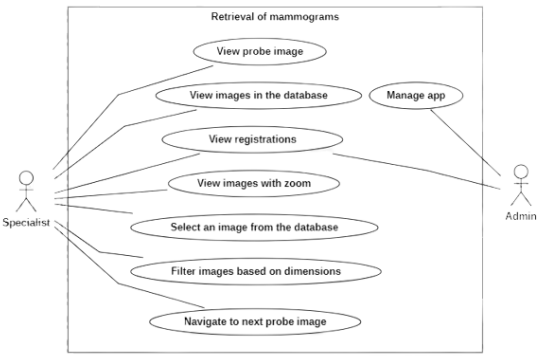


Figure 1: Use case diagram

4 Database

The data model, shown in Figure 2, was created to represent the information that is necessary for the application to meet its objectives.

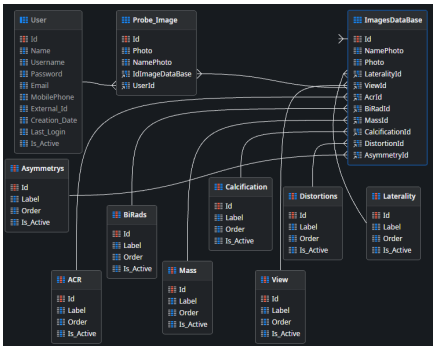


Figure 2: Database

The Database was built with three entities (Probe_Image, ImagesDataBase and User) and eight static entities (Distortions, ACR, BiRads, Mass, Calcification, Asymmetrys, View and Laterality). The Probe_Image entity represents all the probe images that will be associated with an image in the database. The ImagesDataBase entity represents all the images of the database, where only one of these, in each interaction, will be selected as the relevant image. The User entity is a default entity of the OutSystems platform and represents the registration of users on the platform. Static entities are entities that have a predefined set of possible values and in this context represent the dimensions of the mammograms. Static entities have a set of values ("Records") as shown in Figure 3.

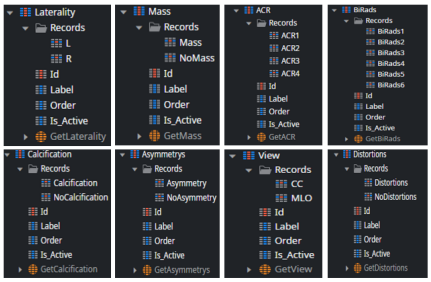


Figure 3: Static Entities Records

5 Implementation

The application starts with a login page (Figure 4) where the expert can fill in the form with their credentials to access the software.

Figure 4: Login page

If the provided software access credentials are valid, the user will be directed to the homepage. There, the user only needs to click the "Start" button, which will then take them to a new page (Figure 5). This newly accessed page enables image retrieval, the specialist associates the probe image (highlighted in blue in Figure 5) with the most relevant image on the right (highlighted in orange in Figure 5). To assist the expert in their selection process, the images on the right (highlighted in orange in Figure 5) can be filtered using available filters (highlighted in green in Figure 5). These filters allow the selection of images with specific dimensions of the mammogram (as seen in Figure 3) exclusively. After conducting a thorough analysis, the expert can select the most relevant image by selecting the respective checkbox, followed by clicking the "Save" button to save the record in the database.

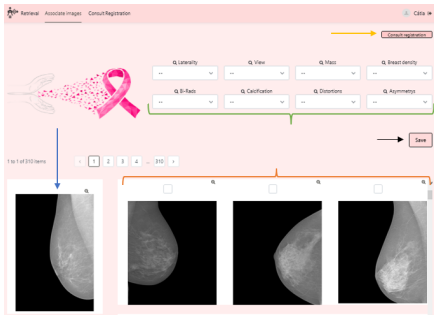


Figure 5: Mammogram retrieval page

After the association procedure is done, the expert is led to a new page (see Figure 6), where all records made can be viewed in detail. The probe image and the image selected by the expert as most relevant, along with their dimensions, are displayed on this page.

Figure 6: Registry example

6 Conclusions and directions for Future Work

Breast cancer is a widespread illness that affects a large number of individuals, with women being at a higher risk of developing it. Despite advancements in medical knowledge and treatment options, there remains a crucial need for ongoing research in this field. This urgency arises from the fact that breast cancer continues to claim the lives of thousands of individuals each year. By investing in further research efforts, we can strive to improve prevention, early detection, and treatment strategies, ultimately reducing the devastating impact of this disease on countless lives. The present project is related to the aforementioned concerns since early detection of breast cancer is essential in achieving a favourable prognosis. A software tool was developed to gather data, thereby facilitating the creation of a mammography retrieval system [4]. This software was developed according to initial requirements, and subsequently refined, to have a final quality version to facilitate the specialist's work. The disadvantages of the software are that it is not constantly fed with new mammogram images and has not yet been tested in a clinical environment. In terms of future prospects, the primary focus centres on software enhancement, specifically the creation of a version suitable for clinical application.

References

[1] M. Arnold, E. Morgan, H. Rumgay, A. Mafra, D. Singh, M. Laversanne, J. Vignat, JR Gralow, F. Cardoso, S. Siesling, and I Soerjomataram. Current and future burden of breast cancer: Global statistics for 2020 and 2040. *The Breast*, 66:15–23, 2022.

[2] OutSystem. Low-code de alto desempenho para desenvolvimento de aplicativos, 2018. URL `https://www.outsystems.com/pt-br/`. Access date: 2023-07-27.

[3] PenRad. PenFetch Automatic Smart Image Transfer and More. Eliminate Waste., 2008. URL `http://penrad.comcastbiz.net/pdf/PENFETCH_2%20Page.pdf`. Access date: 2023-08-07.

[4] Cátia Roriz. Retrieval de mamografias. Master's thesis, Polytechnic University of Coimbra, 2023 (to be submitted).

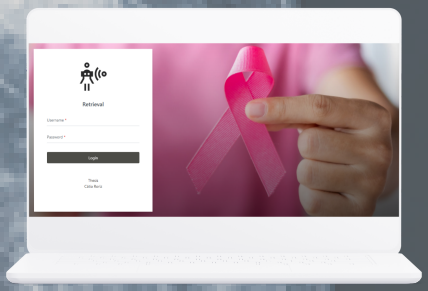
[5] V. Sharma. Mammogram Image Retrieval System Using Texture and Semantic Features. *J of Physics: Conf Ser*, 2267(1):012071, 2022.

Annex D - RecPad poster

Low-Code Application for Ground Truth Data Acquisition in Mammogram Retrieval

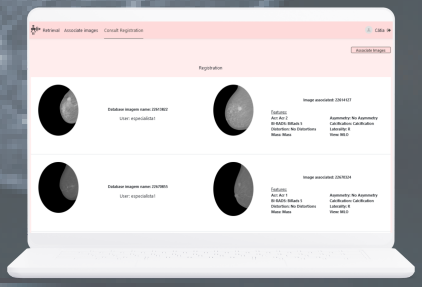
CÁTIA RORIZ, VERÓNICA VASCONCELOS, INÊS C. MOREIRA, INÊS DOMINGUES

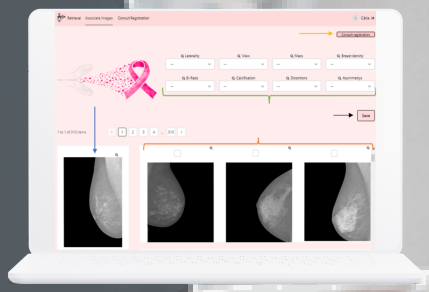
Breast cancer is a significant global health concern, affecting thousands of individuals, primarily women, with estimated cases expected to clim by 2040.



This paper describes the creation of an application to collect ground truth data to aid engineers in the development of a mammography retrieval system. The application is built upon OutSystems, a low-code application platform.

Key features of the application include allowing experts to view probe images and associate them with relevant images from the database. Additionally, the platform allows image filtering based on eigth mammogram dimensions.





While the ultimate goal is to create a system for medical specialists, the current platform represents a step in the process, facilitating the acquisition of ground truth.

isec Politécnico de Coimbra RECPAD 2023 i2a Politécnico de Coimbra

Annex E - Industrial Seminars paper

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Comparação de algoritmos de classificação para o diagnóstico de cancro da mama

Cátia Roriz[1]
[1]Instituto Politécnico de Coimbra, 3045-093 Coimbra, Portugal
Autor correspondente: Cátia Roriz (e-mail: a21270590@isec.pt).

RESUMO O cancro da mama é um dos maiores problemas na sociedade em geral, não só porque é muito comum, mas também porque tem repercussões na condição física, emocional e social. A mortalidade por cancro da mama diminuiu consideravelmente, atribuindo este fenómeno ao rastreio, o que permite um diagnóstico cada vez mais precoce, o que justifica uma investigação intensa para melhorar a vida dos doentes. Com base nisto, propõe-se apresentar uma análise comparativa do desempenho dos diferentes modelos de classificação com base em estudos já realizados, que permitam uma melhor classificação do cancro da mama, permitindo assim um melhor diagnóstico do cancro da mama.

Palavras-chave: Cancro da mama, Diagnóstico, Modelos de Classificação, Mortalidade, Rastreio.

I. INTRODUÇÃO

O cancro de mama é um dos problemas de saúde mais comuns entre as mulheres e uma das causas principais de morte no mundo na atualidade. O diagnóstico e a deteção precoce são um fator essencial na deteção do cancro de mama. Quanto mais cedo o cancro da mama for detetado, maiores serão as probabilidades de um paciente receber tratamento. Existe um grande esforço por parte dos médicos para determinar o diagnóstico de cancro da mama, se este é benigno ou maligno. Existem inúmeros fatores que dão origem a esta doença, e não são fáceis de identificar. Os resultados são difíceis de obter, mesmo para médicos especializados, levando a um aumento do uso de *machine learning* e inteligência artificial como ferramentas de diagnóstico nos últimos anos. Portanto, muitos métodos de deteção precoce estão a ser investigados e utilizados para combater o cancro da mama. No entanto, dada a magnitude do problema, os esforços ainda estão a ser feitos. Classificar com precisão os vários tipos de cancro da mama é uma tarefa fundamental. As técnicas baseadas em inteligência artificial e *machine learning* podem reduzir erros e economizar tempo. Atualmente a mamografia é ainda um dos métodos mais utilizados e um dos mais tradicionais na identificação de cancro da mama, pois é capaz de registar imagens da mama com o objetivo de identificar ou não a presença de massas que permitam detetar a doença. Este tipo de exame requer experiência e uma atenção muito grande por parte dos especialistas de forma a fazer um diagnóstico certeiro.
Em Portugal, são detetados anualmente cerca de 7.000 novos casos de cancro da mama e 1.800 mulheres morrem da doença [1].

O resto do presente documento está organizado da seguinte forma. A seção I apresenta uma reflexão introdutória ao tema a abordar. A seção II faz uma descrição de algoritmos de classificação. A seção III apresenta uma descrição dos artigos que apresentam ideias semelhantes sobre o tema apresentado. A seção IV aborda a metodologia a adotar na construção do estudo. A seção V apresenta um breve resumo sobre cada um dos artigos analisados e os respetivos algoritmos de classificação. A seção VI faz uma análise crítica aos resultados obtidos pela literatura. Finalmente a seção VII apresenta algumas reflexões sobre o tema apresentado e trabalho futuro.

II. ALGORITMOS

Esta seção apresenta uma descrição mais detalhada dos principais algoritmos mais utilizados na classificação do cancro da mama.
A classificação [2] trata-se da atribuição de uma categoria diferente de acordo com alguns parâmetros, a uma determinada amostra de um problema. Em seguida a categoria é prevista para os dados.
O algoritmo *Support vector machine* (SVM) é um dos algoritmos de classificação mais conhecidos e um dos mais utilizados no diagnóstico de cancro da mama, uma vez que apresenta na grande maioria das vezes um bom desempenho. Trata-se de um algoritmo [3] que tem como objetivo, num espaço N-dimensional (N que é o número de características), encontrar um hiperplano que classifica os pontos de dados, hiperplano este que se trata de uma linha. Os *support vector* [4] são as coordenadas de cada tipo de dados, ou seja, SVM pode ser considerado a fonteira entre as duas classes de dados utilizando o hiperplano. Quanto às vantagens [4] deste algoritmo, trata-se de um modelo generalizado pois tem

menos risco de *overfitting* e ainda conseguem gerar resultados de classificação precisos devido à sua robustez. Quanto às desvantagens [4], trata-se do facto de o tempo de treino ser mais demorado ao utilizar um conjunto de dados relativamente grande e ainda o facto de ser difícil de interpretar o modelo e o seu impacto.
O algoritmo *K-nearest neighbors*[5], também conhecido por KNN ou k-nn, trata-se de um algoritmo que na maioria das vezes é utilizado para classificação, embora também possa ser utilizado para regressão. O objetivo deste algoritmo passa por identificar os pontos vizinhos mais próximos do ponto de consulta, para que se possa atribuir uma determinada classe a esse ponto. O valor k define a quantidade de vizinhos que serão necessários para classificar um determinado ponto, e a escolha do valor de k irá depender da quantidade de dados de estrada. Quanto às vantagens deste algoritmo [6], trata-se de um algoritmo que é muito fácil de implementar e ainda é considerado um algoritmo *lazy learning* pois não requer qualquer treino antes de fazer as previsões, ou seja, torna-se muito mais rápido que o SVM que requer treino. Quanto às desvantagens do algoritmo [6], passam pelo alto custo na previsão de grandes conjuntos de dados e ainda na presença de variáveis categórica presentam uma baixa *perfomance*.
O algoritmo *K-means* [7] é um algoritmo de *clustering*, que dentro do conjunto de dados, agrupa os objetos semelhantes entre si em clusters de modo que sejam distintos dos outros clusters, diminuindo a distâncias entre os pontos. Pode ser considerado como a descoberta de clusters que não tem um rótulo. Quanto as vantagens deste algoritmo [8] , passam pela atribuição dos objetos automaticamente a um grupo e ainda a localização do grupo no espaço pode variar, o que faz com que sejam estabelecidas condições iniciais de dependência. Quanto às desvantagens [8], é considerado um algoritmo que inicialmente tem de se decidir quantos grupos são necessários e ainda o facto de os objetos terem de pertencer obrigatoriamente a um grupo.

III. ESTADO DA ARTE

Esta seção apresenta trabalhos de investigação que recorrem a algoritmos de classificação com o objetivo de fazer o diagnóstico de cancro da mama.
Com vista a encontrar os melhores trabalhos de investigação relacionados com o diagnóstico de cancro da mama e por sua vez fazer uma análise dos respetivos algoritmos que melhor identificação o mesmo, foi feita uma pesquisa no *google scholar* com a pesquisa *"Diagnosis of Breast Cancer with Mammography Retrieval"* obtendo um total de 28 600 resultados. Este resultado foi filtrado, apenas foram selecionados os trabalhos a partir do ano 2000, obtendo um total de 18 000 resultados. A pesquisa foi feita em intervalos de tempo de forma a diminuir o número de resultados, e dos respetivos trabalhos foram apenas selecionados os 10 artigos com maior relevância, entre os anos 2004 e 2021.

No artigo [9], os autores tinham como objetivo implementar uma nova abordagem de *content-based image retrieval* (CBIR) de imagens de mamografias, em que a semelhança destas era aprendida partindo de exemplos fornecidos por observadores humanos. O *dataset* utilizado é do *Department of Radiology at The University of Chicago* e é constituído por 76 mamografias com *clustered microcalcifications* (MC's). O desempenho do sistema de recuperação foi avaliado em duas fases, uma fase de classificação e uma fase de regressão. A fase de classificação funciona como uma fase de triagem das imagens, em que o classificador binário elimina as imagens que não tem qualquer semelhança com a imagem de consulta. A fase de regressão faz a comparação das imagens anteriores com a imagem de consulta. Foram utilizadas diferentes estruturas, sendo elas: para a primeira fase um Fisher discriminant linear e para a segunda fase um SVM (Fisher-SVM); para a primeira fase um SVM linear e para a segunda fase um SVM (SVM-SVM); para a primeira fase um SVM linear com a função objetivo modificada e para a segunda fase um SVM (MSVM-SVM); para a primeira fase um SVM linear com a função objetiva modificada e para a segunda fase um General Regression Neural Network (GRNN) (MSVM-GRNN). Os resultados experimentais mostraram que MSVM-SVM obteve o melhor resultado com aproximadamente um de *precision*.
O artigo [9] apresenta um grande potencial uma vez que está publicado numa revista de referência, além disso é interessante na mediada em que as semelhanças das imagens são aprendidas a partir de exemplos fornecidos por observadores humanos. Um dos pontos fracos deste estudo é o facto de a abordagem desenvolvida não ter sido testada em ambiente clínico, uma vez que poderia vir a ter muito potencial.

No artigo [10], os autores tinham como objetivo desenvolver e avaliar um sistema CBIR para mamografias. Investigaram a forma como é que os anteriores casos semelhantes de recuperação poderiam ser utilizados para melhorar o desempenho de um classificador numérico. O dataset utilizado é do *Department of Radiology at the University of Chicago* e é constituído por 200 imagens de mamografias, das quais 58 imagens são de cancro da mama benigno e 46 malignos. Neste estudo foi utilizado o classificador SVM e um *Adaptive* SVM(Ada-SVM), e é feita uma comparação entre estes quanto à utilização do *discriminant adaptive nearest neighbors* (DANN), que foi utilizada para fazer a recuperação das imagens semelhantes. Por fim, ficou comprovado que o Ada-SVM obteve a melhor *accuracy* com o 0.7925 (Ada-SVM com N=9).
O artigo [10] é bastante interessante uma vez que levou a um erro de generalização reduzido. Um dos pontos fracos é o facto de a abordagem não ter passado por avalições clínicas, o que seria uma mais-valia de forma a testar esta abordagem.

No artigo [11], os autores fazem uma revisão geral da literatura, das técnicas mais recentemente desenvolvidas pelos sistemas *computer-aided diagnostics*(CAD), como deteção de massas, calcificações, assimetria bilateral, distorção arquitetónica bem como recuperação de imagens.

No entanto, o desempenho destes sistemas CAD ainda precisa de melhorias significativas para poder satisfazer as aplicações clínicas. Das inúmeras literaturas mencionadas, os algoritmos de predominam são SVM e KNN.
O artigo [11] é interessante uma vez que faz uma revisão geral até o ano de 2009, daquilo que já foi desenvolvido no âmbito da detenção do cancro da mama. Contudo os resultados ficaram um pouco aquém uma vez que os estudos já realizados até ao ano de 2009 não são muito eficazes, o que é necessário continuar a fazer esforços e melhorias de modo a satisfazer o ambiente clínico.

No artigo [12], os autores têm como objetivo desenvolver um método baseado em recuperação de imagem com o intuito de desenvolverem classificadores para casos no CAD. O *dataset* utilizado é o DDSM da *University of South Florida* que é constituído por 589 imagens de mamografias, das quais 331 são cancro da mama benignos e 258 malignos. Começam por fazer a recuperação das imagens obtendo um conjunto de imagens com características semelhantes ao caso de diagnóstico. Posteriormente estes casos são utilizados para otimizar um classificador, com o objetivo de aumentar a *accuracy* de classificação do caso de cancro da mama. Foram utilizados, um simples classificador linear e um classificador SVM não-linear. O classificador SVM não-linear obteve o melhor resultado com uma *accuracy* de 0,8.
O artigo [12] é bastante realista na medida em que as características das imagens são extraídas a partir de um algoritmo que quantifica as lesões, em vez de essa deteção ser feita de forma manual.

No artigo [13], os autores tinham como objetivo desenvolver um sistema *content-based image retrieval* (CBIR), para classificar o tecido mamário. O *dataset* de referência (IRMA) utilizado é constituído por 10000 mamografias provenientes de vários conjuntos de dados, DDSM *database*, MIAS *database*, LLNL *database* e RWTH *dataset*. Todo o processo de extração e seleção de features foi realizado pelo *Singular value decomposition* (SVD), e a semelhança entre casos é feita pelo SVM, categorizando o tecido mamário utilizando a escala *Breast Imaging Reporting and Data System* (BI-RADS). Quanto aos resultados das experiências, foi obtida uma *precision* de 82,14% utilizando 25 de SVD, podendo concluir que o SVD aliado ao SVM para o desenvolvimento do sistema CBIR pode ser uma mais-valia no diagnóstico.
O artigo [13] é interessante na medida que é utilizada a escala de BI-RADS para fazer a caracterização do tecido mamário, uma vez que é uma escala de referência. Um dos pontos fracos é o facto de a abordagem desenvolvida não combinar a densidade mamária, a visão e a lesão como um padrão de forma a fazer a recuperação pois seria uma mais-valia de modo a perceber melhor o comportamento do algoritmo.

No artigo [14], os autores tinham como objetivo introduzir um método de avaliação de risco de malignidade de cancro da mama em mamografias com *microcalcifications* (MCs), e realizar o CBIR das mamografias. O dataset utilizado trata-se do DDSM constituído por 87 mamografias com MCs. O método de avaliação teve em conta a morfologia das MCs definida na escala BI-RADS, e utilizou características de forma e textura como *input* no classificador SVM com vista a realizar a recuperação das imagens. Dado os resultados experimentais foi possível verificar que a recuperação foi feita de forma eficaz pelo esquema CBIR com uma *precision média* entre 0,57 e 0,60 para as várias classes de MCs.
O artigo [14] é interessante na medida que é desenvolvida uma abordagem que é feita uma avaliação do risco de malignidade e é avaliada com base nas classificações BI-RADS, sendo uma escala de referência. Um dos pontos fracos desta abordagem é o baixo desempenho do SVM, pois se as características do input fossem personalizadas, poder-se-ia obter eventualmente um melhor desempenho de recuperação para cada BI_RADS.

No artigo [15], os autores tinham como objetivo desenvolver um método para fazer a segmentação da região *parenchymal* a partir do tamanho do peito dos pacientes. Os dados utilizados provêm da *Mammography Image Analysis Society* (Mini-MIAS) *database* que é constituída por 322 mamografias. Quanto ao classificador utilizado foi ao SVM, que funcionou como um modelador das características, classificando o tecido mamário como, *glandular*, *fatty* e *dense*. Posteriormente foi utilizado o KNN para recuperar as primeiras k imagens, mais próximas da imagem de consulta. Quanto aos resultados da experiência, foi possível verificar que o método proposto foi eficaz e pode ser utilizado na recuperação de mamografias uma vez que foi obtida uma *average precision* de 86.15%. Este método foi desenvolvido utilizado o *software* MATLAB R2012a (Version 7.14).
O artigo [15] é bastante promissor, uma vez que poderia ser utilizado para adaptar os parâmetros de classificação e recuperação de lesões mamárias.

No artigo [16], os autores tinham como objetivo desenvolver um novo método de diagnóstico e recuperação de massas mamárias, uma vez que os métodos existentes ficam muito aquém na fase de recuperação. Quanto ao conjunto de dados foi retirado da *digital database for screening mammography*, *extraindo* 11553 *region of interest* (ROIs). Para determinar se a consulta ROI tem alguma massa são utilizadas as ROIs recuperadas, processo este que é efetuado pelo *scale-invariant feature transform* (SIFT) que faz a extração e pesquisa das características dos ROIs mamográficos. O algoritmo utilizado nesta experiência foi o *k-means*, que obteve uma *accuracy* de 90.8% com k=5.
O artigo [16] é bastante interessante na medida que faz a recuperação de massas mamárias a partir de consultas ROI, tornando-se uma boa abordagem diferenciada das demais já existentes, além disso está publicado numa revista conceituada.

No artigo [17], os autores tinham como objetivo desenvolver um sistema CAD para poder servir de apoio ao diagnóstico

de cancro da mama. Este sistema é baseado na recuperação de imagens de mamografias semelhantes com base na patologia respetiva, e ainda utiliza os casos recuperados para melhorar a performance do classificador. Quanto aos dados utilizados foi a DDSM *database*, constituído por 4300 imagens de cancro da mama. O sistema de apoio à decisão faz uma exploração das mais valias dos descritores de características locais em relação aos descritores de características globais. Para representar uma imagem quanto às características globais, foi necessário um único descritor, já para descrever uma imagem localmente foram necessários vários descritores. Foram utilizados vários descritores, *Scale invariant feature transform* (SIFT), *Speeded Up Robust Feature* (SURF) e *Local Binary Pattern* (LBP). Quanto ao classificador utilizado foi o KNN ponderado. Quanto aos resultados, o LBP ultrapassou o SIFT e SURF, com uma *precision* de aproximadamente 0,9.
O artigo [17] comprova que os CADs utilizando CBIR melhoram a tomada de decisão, além disso os resultados da recuperação são promissores quando se faz a comparação com técnicas já existentes.

No artigo [18], os autores tinham como objetivo utilizar um classificado otimizado de modo a desenvolver um método para recuperação de mamografias. Os dados utilizados provêm da Mini-MIAS *database* extraindo 216 imagens, das quais 51 são malignas, 63 benignas e 102 são imagens normais. Quanto aos classificadores utilizados foram o MANFIS otimizado, otimizado pelo algoritmo *Artificial Bee Colony* (ABC), o classificado ANFIS, NN, *Naives Bayes*(NB) e SVM. Quanto aos resultados experimentais o melhor resultado foi obtido perlo ABC-MANFIS com uma *acuraccy* de 0.98148. Este método foi desenvolvido utilizado o software MATLAB.
O artigo [18] é interessante e apresentou-se com uma *accuracy* bastante elevada, contudo, não foi testado em ambiente clínico

IV. Metodologia

Esta seção apresenta a metodologia adotada.
Os métodos aplicados em cada iteração do desenvolvimento deste trabalho consistem em quatro etapas. A primeira etapa passa por analisar os objetivos dos artigos percebendo se todos eles vão ao encontro do mesmo objetivo, a deteção do cancro da mama. A segunda etapa passa por analisar os conjuntos de dados de maneira a perceber se os dados utilizados são o mais similar possível. A terceira etapa destina-se a perceber quais os algoritmos que obtiveram os melhores resultados na deteção de cancro da mama. Por fim a última etapa passa por analisar os resultados obtidos pelos algoritmos de classificação.
Analisando apenas os algoritmos e os restivemos resultados obtidos, não seria possível ter um termo de comparação entre algoritmos, portanto os objetivos dos artigos da literatura e os respetivos conjuntos de dados tem de ser analisados também.
Com esta metodologia é possível perceber qual ou quais os algoritmos que melhor podem fazer o diagnóstico de cancro da mama.

V. RESUMO DA LITERATURA

Nesta seção é apresentado na Tabela 1 um breve resumo sobre cada um dos artigos analisados da literatura, quanto ao conjunto de dados adotado, algoritmos utilizados e ainda os respetivos resultados.

Tabela 1- Resumo da literatura

Artigo	Conjunto de dados	Algoritmos	Resultados
A similarity learning approach to content-based image retrieval: application to digital mammography [9]	*Department of Radiology at The University of Chicago*: 76 imagens de mamografias	Fisher-SVM SVM-SVM MSVM-SVM MSVM-GRNN	MSVM-SVM *Precision* $\simeq 1$
Retrieval-Driven Microcalcification Classification For Breast Cancer Diagnosis [10]	*Department of Radiology at the University of Chicago*: 200 imagens de mamografias 58 imagens de cancro da mama benigno e 46 malignos	SVM e *Adaptive* SVM	*Adaptative SVM* *Accuracy:* 0.7925
Computer-Aided Detection and Diagnosis of Breast Cancer with	Mamografias	SVM e KNN	Sem resultados, uma vez que se trata de uma revisão da literatura

Mammography: Recent Advances [11]			
Image retrieval for computer-aided diagnosis of breast cancer [12]	DDSM da *University of South Florida:* 589 imagens de mamografias, 331 são cancro da mama benignos e 258 malignos.	Classificador linear e classificador SVM não-linear.	*SVM não-linear Accuracy* :0,8
Content-based image retrieval applied to BI-RADS tissue classification in screening mammography [13]	IRMA (DDSM *database*, MIAS *database*, LLNL *database* e RWTH *dataset)*: 10000 imagens de mamografias	SVM	*Precision:*82,14%
Microcalcification oriented content-based mammogram retrieval for breast cancer diagnosis [14]	DDSM: 87 imagens de mamografias	SVM	*Average precision:* entre 0,57 e 0,60
An intelligent content-based image retrieval system for mammogram image analysis [15]	Mini-MIAS *database*: 322 de imagens de mamografias	SVM	A*verage precision:* 86.15%
Computer-Aided Diagnosis of Mammographic Masses Using Scalable Image Retrieval [16]	*Digital database for screening mammography:* 11553 ROIs	*K-means*	*Accuracy:* 90.8% com k=5
Retrieval Driven Classification for Mammographic Masses [17]	DDSM *databas*e: 4300 imagens de mamografias	KNN	*Precision:*0.9
An Efficient Mammogram Image Retrieval System Using an Optimized Classifier [18]	Mini-MIAS *database:* 216 imagens de mamografias, 51 são cancro da mama maligno, 63 benigno e 102 imagens normais.	MANFIS otimizado pelo algoritmo ABC, ANFIS, NN, NB e SVM.	ABC-MANFIS *Accuracy:* 0.98148

VI. DISCUSSÃO

Nesta seção é apresentada uma pequena discussão sobre os resultados apresentados no estado da arte, analisando quais as características dos demais artigos apresentados, com o objetivo de prever qual o algoritmo mais adequado para o diagnóstico de cancro da mama.

A. OBJETIVO DOS ARTIGOS

O diagnóstico precoce de cancro da mama é essencial, e muitos são os estudos feitos para este efeito. O artigo apresentado tem como objetivo identificar os melhores algoritmos que permitam a identificação ou deteção do cancro da mama com base em estudos já realizados, para isso, foram apresentados alguns artigos na literatura. Os objetivos da revisão da literatura passam por fazer o diagnóstico/deteção do cancro da mama. Com isto é possível perceber que todos os artigos vão ao encontro do objetivo pedido.

B. DATASETS UTILIZADOS

Os dados utilizados na maioria dos artigos da literatura, são imagens de cancro da mama(mamografias) fazendo variar apenas o número de instâncias de cada. O único artigo que varia o tipo de *dataset* é apenas um [16], que é constituído por consultas ROI que permitem determinar se a consulta ROI apresenta ou não alguma massa mamária. Esta variação no número de dados, faz com que haja diferença também no desempenho do algoritmo, fazer um treino com mais ou menos dados torna-se diferente.

C. *ALGORITMOS*

Quanto aos algoritmos utilizados na literatura, aquele que se tornou mais presente foi o SVM. Foram apresentados outros algoritmos como K-means e KNN.
Estes algoritmos permitiram avaliar a eficácia e eficiência na deteção de cancro da mama nos vários *datasets* da literatura. Por esse motivo, o que torna os artigos da literatura mais distintos são essencialmente a forma como os algoritmos são treinados e a forma como fazem a identificação do cancro da mama. Infelizmente, dos artigos analisados, não há informação sobre os tempos de treino de cada algoritmo e quais a condições de treino bem como o sítio onde o código foi desenvolvido, por esse motivo não é possível fazer uma comparação exaustiva entre os algoritmos apenas tendo acesso à respetiva *precision* ou *accuracy*.

D. *RESULTADOS*

Segundo os resultados apresentados na literatura é possível concluir que todos os estudos conseguiram identificar com sucesso o cancro da mama, como se pode verificar pela Figura 1.

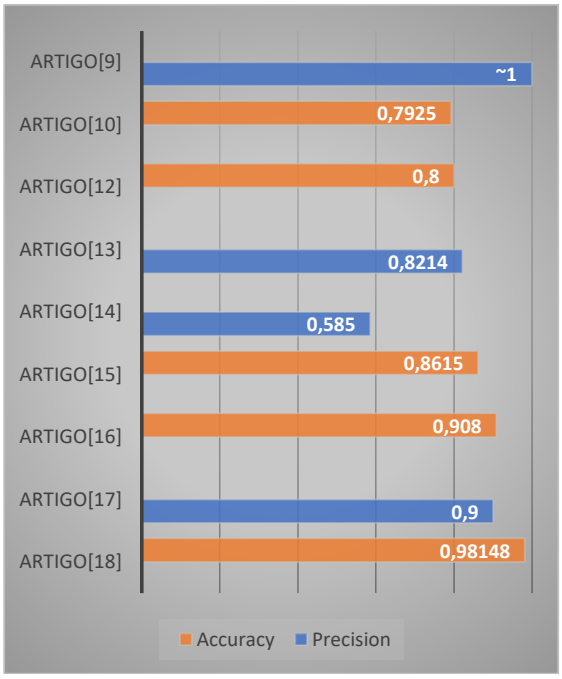


Figura 1- Resultados

Apesar de não serem utilizadas as mesmas métricas para todos os artigos da literatura, é possível observar que os resultados obtidos foram bastante bons no geral.
O algoritmo que apresenta melhor *performance* na revisão da literatura é o SVM, com uma *precision* de aproximadamente 1.

O SVM acaba por ser o algoritmo que mais vezes se apresenta como melhor resultado, sendo possível perceber que se torna num dos melhores algoritmos na deteção de cancro da mama.
Por fim, é possível deduzir que todos os algoritmos presentes na literatura, comprovaram que de uma forma geral conseguem identificar com alta precisão o cancro da mama.

VII. CONCLUSÃO E TRABALHO FUTURO

O diagnóstico precoce de cancro da mama é cada vez mais uma realidade, na medida que existem cada vez mais estudos, bons estudos, que permitem a prevenção e diagnóstico deste tipo de doença. Ao longo deste artigo foram abordados alguns dos estudos realizados para combater o cancro da mama, e foi feita uma análise dos algoritmos que melhor fazem a deteção desta doença. O algoritmo que se apresenta como o melhor a identificar cancro da mama foi o SVM, o que é possível concluir que este algoritmo é bastante bom para a identificação desta doença. A variação do número de dados dos *datasets* e o facto de não se ter acesso às condições onde o código foi desenvolvido e treinado, tornou-se uma das limitações no desenvolvimento do artigo de modo a ser feita uma boa análise. O trabalho futuro passa por analisar artigos que pensem de forma distinta como o artigo [16], diferenciada dos demais estudo já realizados para fazer o diagnóstico de cancro da mama.

REFERÊNCIAS

[1] «Cancro da Mama : Liga Portuguesa Contra o Cancro». https://www.ligacontracancro.pt/cancro-da-mama/ (acedido 9 de janeiro de 2023).
[2] «Aprendizagem por máquinas: Tipos de Algoritmos de Classificação | Southern Jordan». https://southernjordanguide.com/pt/classifica%c3%a7%c3%a3o-de-algoritmos-a-tomato-inspired-overview/ (acedido 9 de janeiro de 2023).
[3] zehClaudio, «Máquinas de Vetores de Suportes (SVM): Introdução aos algoritmos de AM», *Alvarez Soluções Digitais*, 23 de fevereiro de 2020. https://alvarezsolucoesdigitais.com/aprendizado-de-maquina/maquinas-de-vetores-de-suportes-svm-introducao-aos-algoritmos-de-aprendizado-de-maquina/ (acedido 9 de janeiro de 2023).
[4] «Algoritmo SVM | trabalhando com cenários | Prós e contras do SVM». https://pt.education-wiki.com/pt.education-wiki.com/4245111-svm-algorithm (acedido 11 de janeiro de 2023).
[5] «What is the k-nearest neighbors algorithm? | IBM». https://www.ibm.com/topics/knn (acedido 9 de janeiro de 2023).
[6] A. L. Vaz, «KNN —K-Nearest Neighbor, o que é?», *Data Hackers*, 10 de maio de 2021. https://medium.com/data-hackers/knn-k-nearest-

neighbor-o-que-%C3%A9-aeebe0f833eb (acedido 11 de janeiro de 2023).
[7] «What is K Means?», *NVIDIA Data Science Glossary*. https://www.nvidia.com/en-us/glossary/data-science/k-means/ (acedido 9 de janeiro de 2023).
[8] «K-Médias». https://web.tecnico.ulisboa.pt/ana.freitas/bioinformatics.ath.cx/bioinformatics.ath.cx/index651a.html?id=147 (acedido 11 de janeiro de 2023).
[9] I. El-Naqa, Y. Yang, N. P. Galatsanos, R. M. Nishikawa, e M. N. Wernick, «A similarity learning approach to content-based image retrieval: application to digital mammography», *IEEE Trans. Med. Imaging*, vol. 23, n.º 10, pp. 1233–1244, out. 2004, doi: 10.1109/TMI.2004.834601.
[10] R. M. Nishikawa, L. Wei, e Y. Yang, «RETRIEVAL-DRIVEN MICROCALCIFICATION CLASSIFICATION FOR BREAST CANCER DIAGNOSIS», em *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, abr. 2007, pp. 1260–1263. doi: 10.1109/ISBI.2007.357088.
[11] J. Tang, R. M. Rangayyan, J. Xu, I. E. Naqa, e Y. Yang, «Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances», *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, n.º 2, pp. 236–251, mar. 2009, doi: 10.1109/TITB.2008.2009441.
[12] H. Jing e Y. Yang, «Image retrieval for computer-aided diagnosis of breast cancer», apresentado na IEEE Southwest Symposium on Image Analysis & Interpretation (SSIAI), mai. 2010, pp. 9–12. doi: 10.1109/SSIAI.2010.5483930.
[13] J. E. E. de Oliveira, A. de Albuquerque Araújo, e T. M. Deserno, «Content-based image retrieval applied to BI-RADS tissue classification in screening mammography», *World J. Radiol.*, vol. 3, n.º 1, pp. 24–31, jan. 2011, doi: 10.4329/wjr.v3.i1.24.
[14] L. Tsochatzidis *et al.*, «Microcalcification oriented content-based mammogram retrieval for breast cancer diagnosis», em *2014 IEEE International Conference on Imaging Systems and Techniques (IST) Proceedings*, out. 2014, pp. 257–262. doi: 10.1109/IST.2014.6958484.
[15] K. Vaidehi e T. S. Subashini, «An intelligent content based image retrieval system for mammogram image analysis», *J. Eng. Sci. Technol.*, vol. 10, n.º 11, pp. 1453–1464, 2015.
[16] M. Jiang, S. Zhang, H. Li, e D. N. Metaxas, «Computer-Aided Diagnosis of Mammographic Masses Using Scalable Image Retrieval», *IEEE Trans. Biomed. Eng.*, vol. 62, n.º 2, pp. 783–792, fev. 2015, doi: 10.1109/TBME.2014.2365494.
[17] K. Kiruthika, D. Vijayan, e L. R., «Retrieval Driven Classification for Mammographic Masses», em *2019 International Conference on Communication and Signal Processing (ICCSP)*, abr. 2019, pp. 0725–0729. doi: 10.1109/ICCSP.2019.8698044.
[18] S. J. Rayen e R. Subhashini, «An Efficient Mammogram Image Retrieval System Using an Optimized Classifier», *Neural Process. Lett.*, vol. 53, n.º 4, pp. 2467–2484, ago. 2021, doi: 10.1007/s11063-020-10254-3.

Annex F - Software operating instructions

In order to provide a better understanding of how to use the software, a practical guide to how it works is presented below.

Take the following steps:

1. Access the application;
2. Log in with your access credentials;
3. Click on "Start" to be redirected to the main page;
4. Analyse the probe image (left image) and use the magnifying glass if necessary to see all the details;
5. Select the filters relating to the characteristics that make up the mammograms;
6. Analyse the images on the right, already filtered by the characteristics;
7. Use the magnifying glass if you need to analyse the mammograms;
8. Select the most relevant mammogram.
9. Click on "Save" to save the record and wait for the information to be processed and saved in the database.
10. Repeat the same process until the probe images are finished.
11. Click on "Consult registration" to consult all the registrations made.

Annex G - Retrieval extended results

Probe Image	Choice	r_const	r_emp	r_lit	r_exp
22614127	22613822	38	149	149	38
22670324	22670855	60	162	162	60
50998322	30011530	122	206	206	122
20587080	22580393	142	142	142	142
50998981	20587664	11	125	125	11
50999121	50999273	109	263	263	109
50999327	20587836	4	9	9	4
51049628	20586986	111	111	111	111
53587427	53580979	274	274	274	274
22613848	50996056	85	63	63	85
20587612	22670620	126	328	328	126
30318067	30011824	45	45	45	45
50993670	22670511	165	165	165	165
50995872	24058738	49	190	190	49
50996137	24065761	119	195	195	119
50996945	24058738	190	190	190	190
50997080	24058738	49	190	190	49
50997107	22670124	153	153	153	153
50997651	20587226	58	1	1	58
22614568	20587810	167	315	315	167
22670177	24055877	37	37	37	37
22427682	20587294	2	4	4	2
22427751	22427705	14	128	128	14
22679008	20587054	0	0	0	0
51049053	53580804	68	270	270	68
51049276	22579754	11	11	11	11
24055328	22678495	30	30	30	30
24055382	22678449	14	29	29	14
24055483	24065707	315	360	360	315
53582304	20587080	1	1	1	1
53582395	50993670	67	51	51	67
53586805	24055654	109	183	183	109
24055958	50997678	38	80	80	38
53586960	50996972	295	239	239	295
53587508	50999175	174	261	261	174
24058738	50998634	302	376	376	302
20587664	24065761	73	195	195	73
22670147	20587664	11	125	125	11
50997796	22580732	335	142	142	335
50997823	22580732	356	310	310	356
50998177	22670177	18	18	18	18
22670703	50996325	89	233	233	89

50998258	30011647	25	52	52	25
50998440	22678495	30	30	30	30
50998580	51049276	93	93	93	93
24055355	50999459	218	384	384	218
53580831	53580638	97	97	97	97
24055877	20587690	15	126	126	15
53587014	24065707	315	360	360	315
53587454	22670488	159	159	159	159
24065434	22427864	5	10	10	5
24065461	24055654	132	183	183	132
50993616	50993670	51	51	51	51
50996881	22427682	6	13	13	6
50996972	24055600	181	181	181	181
50997597	50996945	69	69	69	69
20586960	20586908	0	310	310	0
22670620	22580192	15	305	305	15
50998607	53582304	265	389	389	265
51048891	30318067	46	46	46	46
53580611	24065380	300	342	342	300
53581379	50996325	89	233	233	89
53586751	22670147	154	154	154	154
53587131	22670832	11	21	21	11
53587599	51048945	45	91	91	45
24058686	50998634	204	376	376	204
22614097	50996854	50	68	68	50
50994589	50994535	61	61	61	61
50996083	50998440	87	87	87	87
50997434	22580576	15	15	15	15
50997488	22678980	15	29	29	15
50997515	22580492	15	19	19	15
24055464	20587346	7	5	5	7
53580804	22614266	32	155	155	32
22580706	22670673	205	330	330	205
24055904	22678495	30	30	30	30
53587104	20587226	1	1	1	1
24058660	50995789	145	224	224	145
53587744	24055931	23	38	38	23
24058712	20587346	5	5	5	5
24065407	53581237	125	278	278	125
22614074	22670278	122	324	324	122
27829215	53580611	62	98	98	62
50996827	50998440	87	87	87	87
22670673	20586960	1	116	116	1
22427728	50998440	87	87	87	87

50998204	30011700	53	53	53	53
53580665	22427864	5	10	10	5
27829188	22614127	42	151	151	42
50994535	20586960	227	311	311	227
50996854	22670832	25	161	161	25
50997742	53586869	299	299	299	299
53587572	51049276	46	93	93	46
53587690	27829215	60	203	203	60
50997461	20587784	12	8	8	12
22614150	51049276	64	93	93	64
24055931	20587320	1	2	2	2
53586778	20587320	1	2	2	2
53586896	22614266	118	155	155	155
50997769	22614405	26	26	26	26
50996800	53586724	102	102	102	102
53586987	20587346	5	5	5	5

Annex H - Version 2 of the extended results retrieval

Probe Image	Choice	r_const	r_emp	r_lit	r_exp
22614127	22613822	67	202	202	202
22670324	22670855	120	219	219	219
50998322	30011530	122	262	262	262
20587080	22580393	248	193	193	193
50998981	20587664	20	164	164	164
50999121	50999273	224	394	394	394
50999327	20587836	8	168	168	168
51049628	20586986	111	1	1	1
53587427	53580979	274	330	330	330
22613848	50996056	85	95	95	95
20587612	22670620	207	369	369	369
30318067	30011824	45	77	77	77
50993670	22670511	165	42	42	42
50995872	24058738	49	244	244	244
50996137	24065761	194	252	252	252
50996945	24058738	190	244	244	244
50997080	24058738	49	244	244	244
50997107	22670124	46	37	37	37
50997651	20587226	58	3	3	3
22614568	20587810	167	167	167	167
22670177	24055877	37	63	63	63
22427682	20587294	3	159	159	159
22427751	22427705	14	14	14	14
22679008	20587054	0	155	155	155
51049053	53580804	135	135	135	135
51049276	22579754	11	17	17	17
24055328	22678495	56	221	221	221
24055382	22678449	27	220	220	220
24055483	24065707	315	251	251	251
53582304	20587080	1	156	156	156
53582395	50993670	67	83	83	83
53586805	24055654	180	374	374	374
24055958	50997678	77	306	306	306
53586960	50996972	295	295	295	295
53587508	50999175	258	393	393	393
24058738	50998634	302	390	390	390
20587664	24065761	142	252	252	252
22670147	20587664	20	164	164	164
50997796	22580732	235	25	25	25
50997823	22580732	322	25	25	25
50998177	22670177	18	38	38	38
22670703	50996325	89	289	289	289

50998258	30011647	48	263	263	263
50998440	22678495	56	221	221	221
50998580	51049276	93	128	128	128
24055355	50999459	218	317	317	317
53580831	53580638	196	326	326	326
24055877	20587690	7	9	9	9
53587014	24065707	315	251	251	251
53587454	22670488	274	368	368	368
24065434	22427864	5	16	16	16
24065461	24055654	226	374	374	374
50993616	50993670	51	83	83	83
50996881	22427682	11	181	181	181
50996972	24055600	292	373	373	373
50997597	50996945	69	101	101	101
20586960	20586908	0	153	153	153
22670620	22580192	29	363	363	363
50998607	53582304	265	404	404	404
51048891	30318067	46	78	78	78
53580611	24065380	193	245	245	245
53581379	50996325	89	289	289	289
53586751	22670147	271	215	215	215
53587131	22670832	11	43	43	43
53587599	51048945	45	124	124	124
24058686	50998634	204	390	390	390
22614097	50996854	50	100	100	100
50994589	50994535	122	273	273	273
50996083	50998440	171	313	313	313
50997434	22580576	15	23	23	23
50997488	22678980	15	51	51	51
50997515	22580492	26	195	195	195
24055464	20587346	12	160	160	160
53580804	22614266	32	32	32	32
22580706	22670673	330	370	370	370
24055904	22678495	56	221	221	221
53587104	20587226	1	3	3	3
24058660	50995789	225	284	284	284
53587744	24055931	23	64	64	64
24058712	20587346	8	160	160	160
24065407	53581237	255	400	400	400
22614074	22670278	203	365	365	365
27829215	53580611	124	133	133	133
50996827	50998440	171	313	313	313
22670673	20586960	2	154	154	154
22427728	50998440	171	313	313	313

50998204	30011700	98	264	264	264
53580665	22427864	5	16	16	16
27829188	22614127	72	209	209	209
50994535	20586960	227	154	154	154
50996854	22670832	11	43	43	43
50997742	53586869	299	143	143	143
53587572	51049276	93	128	128	128
53587690	27829215	60	260	260	260
50997461	20587784	22	166	166	166
22614150	51049276	128	128	128	128
24055931	20587320	1	4	4	4
53586778	20587320	1	4	4	4
53586896	22614266	118	32	32	32
50997769	22614405	43	211	211	211
50996800	53586724	102	141	141	141
53586987	20587346	8	160	160	160

Instituto Superior de Engenharia
Politécnico de Coimbra