Mammogram Retrieval System: Aggregating Image Classifiers for Enhanced Breast Cancer Diagnosis

Cátia Roriz
a21270590@isec.pt
Polytechnic University of Coimbra, Rua da Misericórdia
Lagar dos Cortiços, S. Martinho do Bispo, 3045-093
Coimbra, Portugal

Inês Moreira
icm@ess.ipp.pt
Centro Hospitalar Universitário de São João
Escola Superior de Saúde do Politécnico do Porto
Cintesis-FMUP, Portugal

Verónica Vasconcelos
veronica@isec.pt
Polytechnic University of Coimbra, Rua da Misericórdia
Lagar dos Cortiços, S. Martinho do Bispo, 3045-093
Coimbra, Portugal

Inês Domingues
inesdomingues@gmail.com
Polytechnic University of Coimbra, Rua da Misericórdia,
Lagar dos Cortiços, S. Martinho do Bispo, 3045-093
Coimbra, Portugal
IPO-Porto Research Center – Portuguese Institute of
Oncology of Porto, Portugal

ABSTRACT

Breast cancer remains a significant global health concern. This study presents an image retrieval system to aid specialists in the analysis of mammogram images. The system employs individual classifiers for eight dimensions: laterality, view, breast density, BI-RADS classification, masses, calcifications, distortions, and asymmetries. Four pre-trained networks, ResNet50, VGG16, InceptionV3, and InceptionResNetV2, were used to train these classifiers. The retrieval model combines these classifiers through a weighted sum. Four weight assignment strategies were explored, ranging from equal weights to weights based on empirical, literature-based, and specialist-informed considerations. Results are illustrated using the INBreast database, which comprises 410 images. Besides the native annotations, ground truth to validate retrieval models had to be acquired. Classification accuracy is as high as 100% for some of the dimensions. Results also demonstrate the effectiveness of the proposed weighted-sum approach, with variations in weight assignments impacting model performance.

CCS CONCEPTS

• **Information systems** → **Retrieval tasks and goals**; • **Computing methodologies** → **Machine learning approaches**.

KEYWORDS

Mammogram Retrieval System, Breast Cancer Diagnosis, Image Classification, Medical Imaging, Deep Learning

ACM Reference Format:
Cátia Roriz, Inês Moreira, Verónica Vasconcelos, and Inês Domingues. 2024. Mammogram Retrieval System: Aggregating Image Classifiers for Enhanced Breast Cancer Diagnosis. In *2024 6th International Conference on Intelligent Medicine and Image Processing (IMIP 2024), April 26–29, 2024, Bali, Indonesia.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3669828.3669829

IMIP 2024, April 26–29, 2024, Bali, Indonesia

ACM ISBN 979-8-4007-1003-2/24/04
https://doi.org/10.1145/3669828.3669829

1 INTRODUCTION

Breast cancer is a malignant disease that affects thousands of individuals worldwide, primarily women. By the year 2040 [4], 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 [4] 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 [4]. Although various approaches have been developed, much more effort has to be made to successfully fight this disease.

In Portugal [6], 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 [6]. Given the recommended two-year screening interval for women aged 50 to 69, taking Portugal as an example, a substantial number of mammograms require analysis each day [5].

The ultimate 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.

The main contributions of the present work include:

- 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 into four different concrete retrieval models that combine the above mentioned eight classifiers.

Furthermore, this document includes an illustration of ground truth acquisition for validation of the retrieval models, not previously existing as part of INBreast [17][1]. A thorough evaluation of both the classifiers and the retrieval models is also provided.

The structure of this document is as follows. A comprehensive review of the state of the art is given in Section 2. The methods are given in Section 3, with particular emphasis on the classifiers, retrieval model, database and respective ground truth, and experimental setting. Results are presented in Section 4, where information on four ways of deriving weights is also included. The document finishes in Section 5 with some conclusions and directions for future work.

2 RELATED WORK

In order to find the most relevant previous work, a search was made in the search engine "*Google Scholar*" with the phrase "intitle: Retrieval + intitle: mammogram", obtaining a set of 64 results. Adding documents citing the aforementioned works and excluding works published more than five years ago results in 106 documents. Nine were duplicates, and 91 were deemed irrelevant after screening the titles and abstracts, resulting in six studies, briefly described below.

The approach proposed in [18] involves preprocessing the input dataset by removing noise and enhancing contrast, followed by the removal of pectoral muscles using Single Sided Edge Marking. Feature extraction is performed, incorporating Grey Level Co-occurrence Matrix (GLCM) features, Gabor features, and Local Binary Pattern features, which are then classified into benign, malignant, and normal classes. The study introduces a classifier named Adaptive Neuro Fuzzy Inference System (ANFIS), optimised using Improved Particle Swarm Optimization (IPSO). The similarity between trained feature distance vectors and the input query image is assessed using Euclidean Distance metric for retrieval. Experimental results demonstrate the system's performance using metrics such as sensitivity, specificity, accuracy, precision, recall, F-measure, and Matthews Correlation Coefficient (MCC).

The model proposed in [10] aims to optimise the weight function of a hybrid pattern for image retrieval. The hybrid pattern combines Local Vector Pattern (LVP) and Local Derivative Pattern (LDP), and the weight optimisation is performed using a meta-heuristic algorithm called Improved Local Leader-based Spider Monkey Optimisation algorithm (ILLU-SMO). The hybrid pattern extraction process includes image pre-processing, feature extraction, and weight optimisation. The experimental results are quantified using precision, recall, and F1-score.

The paper [19] presents the Artificial Bee Colony-Modified Adaptive Neuro-Fuzzy Inference System (ABC-MANFIS). The proposed system employs a Modified Weiner Filter (MWF) for preprocessing, followed by pectoral muscle removal and region of interest (ROI) extraction. Twelve features, including contrast, correlation, variance, and others, are then extracted from the ROI. The ABC-MANFIS classifier is used for categorising images into normal, malignant, or benign classes. The classification scores are calculated using Principal Component Analysis (PCA), and image retrieval is performed based on the Euclidean distance between query and database image scores. The experimental results, conducted on a mini-MIAS database, show the effectiveness of the proposed system in terms of specificity, sensitivity, accuracy, precision, recall, F-measure, and other performance metrics, with the ABC-MANFIS classifier outperforming other classifiers like ANFIS, Nearest Neighbour, Naive Bayes (NB), and Support Vector Machines (SVM).

Jouirou, Baazaoui, and Barhoumi [14] introduce a four-view Content-Based Mammogram Retrieval (CBMR) method. The proposed method evolves from Single View (SV)-CBMR to Multi-View (MV)-CBMR, considering four views (Left Mediolateral Oblique, Left Craniocaudal, Right Mediolateral Oblique, and Right Craniocaudal). The method employs dynamic similarity assessment using Random Forest (RF) for distance prediction and a late fusion module based on shared information for combining results from different views. Experimental results are presented using DDSM (Digital Database for Screening Mammography).

The system in [22] uses a two-phase approach, where textual features are initially exploited and the search is then refined using visual features. The dataset used for experimentation is the IRMA version of the DDSM mammography database. The low-level visual features are extracted using various methods, including Haralick's GLCM and Law's Texture Energy Measures. High-level semantic features are obtained using a biomedical ontology of mammograms and controlled vocabulary. The retrieval process involves a three-tier architecture, including a Data Layer, Semantic Layer, and Visual Layer. The system allows users to input queries based on semantic annotations or upload mammogram images for similarity searches. Relevance Feedback (RF) is incorporated to learn user preferences and improve search results.

A method for mammogram image retrieval using Scale-Invariant Feature Transform (SIFT) and Crow Search Optimisation-based Deep Belief Network (CSO-DBN) is proposed in [21]. The proposed approach combines noise removal with Kalman filter, feature extraction using SIFT, and optimisation through CSO-DBN. The experimental shows an accuracy rate of 93.44%, outperforming SVM, NB, Butterfly Optimisation Algorithm (BOA), and Cat Swarm Optimisation (CSO).

Several noteworthy observations emerge from the review. Firstly, none of the examined studies utilise the INBreast database, which is employed in the current work. Additionally, the reviewed methods rely on traditional feature extraction techniques. In contrast, the present study distinguishes itself through the incorporation of deep learning classifiers. For a comprehensive exploration of Convolutional Neural Networks (CNNs) in mammography, including tasks like lesion localization, detection, risk assessment, image retrieval, and classification, refer to the literature review provided in [1].

3 METHODS

The methods applied in each iteration of the development of this work consist of three main steps (see Figure 1). The first step involves building several models using four different pre-trained networks, which allow the classification of mammograms according to the eight different dimensions (Section 3.1). 70% of the dataset

[1]InBreast Retrieval Ground Truth is available upon request to inesdomingues@gmail.com

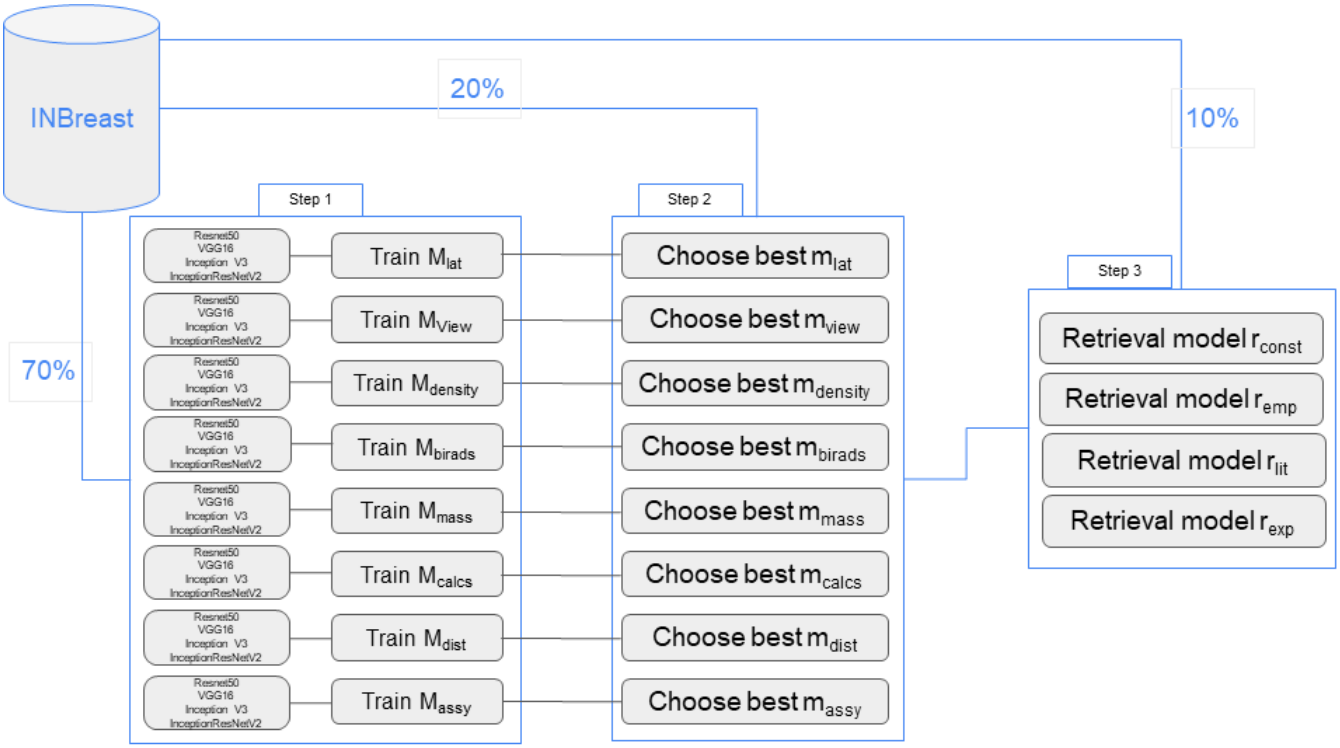


Figure 1: Overview of the training methodology

InBreast was used to train these models. In the second step, the model that classified mammograms most accurately in the 20% validation set, 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 retrieval model is created by combining the best individual classifiers using a weighted sum (Section 3.2). Evaluation of the retrieval models was performed in the final set, containing 10% of the data. This section also includes the description of the used database, the acquisition of the ground truth information (Section 3.3) and the experimental setting (Section 3.4).

3.1 Classifiers

Four pre-trained networks were tested, ResNet50, VGG16, InceptionV3, and InceptionResNetV2. These are selected for having been identified as widely used and accepted in the computer vision community. In this way, four classifiers were created for each dimension (laterality, view, breast density, BI-RADS, masses, calcifications, distortions, and asymmetries), each with a different pre-trained network. In total, 32 models were created (four architectures × eight dimensions). Illustrative examples of the eight dimensions are given in Figure 2.

Initially, the preprocessing of the images is performed, where the images are resized to 224 × 224 pixels and converted from RGB (Red, Green, Blue) to BGR (Blue, Green, Red). Next, each colour channel is zero-centred concerning the ImageNet dataset, without scaling. Two approaches are adopted for training:

- 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.

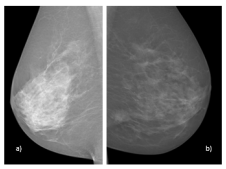
3.2 Retrieval Model

The retrieval model leverages the individual models described in Section 3.1. It is a weighted sum and has the following generic equation:

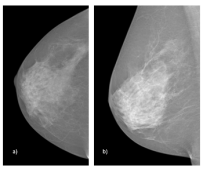
$$d^{probe,i} = w_{lat} * d^{probe,i}_{lat} + w_{view} * d^{probe,i}_{view} + w_{density} * d^{probe,i}_{density} + w_{birads} * d^{probe,i}_{birads} + w_{mass} * d^{probe,i}_{mass} + w_{calcs} * d^{probe,i}_{calcs} + w_{dist} * d^{probe,i}_{dist} + w_{assy} * d^{probe,i}_{assy},$$

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

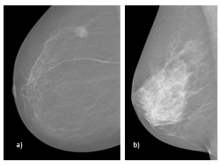
where w_{dim} is the weight attributed for each dimension (laterality, CC or MLO view, breast density, BI-RADS class, existence or



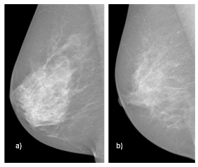
(a) Lateralities: left breast and Right breast



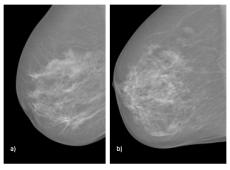
(b) Views: Craniocaudal (CC) and Mediolateral Oblique (MLO)



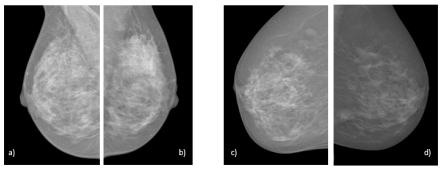
(c) Masses: with masses and without masses



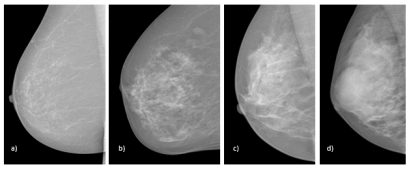
(d) Calcifications: with calcification and without calcification



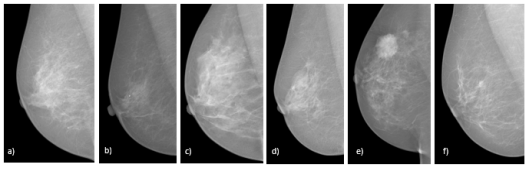
(e) Distortions: with distortions and without distortions



(f) Asymmetry: on the left with asymmetry and on the right without asymmetry



(g) Density: ACR breast density a to d



(h) BI-RADS: 1 to 6

Figure 2: Illustration of the eight mammogram dimensions

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 3.3) for the image i and $\tilde{V}_{dim}^{probe}$ is the result of the classifier (see Section 3.1) 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, as translated by $i = 1 : 410 \backslash \{probe\}$ in the weighted sum equation.

Section 4.2 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 an expert.

3.3 Database

The acquisition of the INBreast database took place at the Breast Centre of the Centro Hospitalar São João, Porto, Portugal authorised by the Hospital's Ethics Committee and the National Data Protection Commission [12, 17]. The images in the database have a pixel size of 70 µm (microns) and a contrast resolution of 14 bits, with an image matrix size of 3328 x 4084 or 2560 x 3328 pixels. These images were stored in the Digital Imaging and Communications in Medicine (DICOM) format, ensuring the removal of all confidential medical information from the DICOM files. The maintenance of correspondence between images of the same patient is achieved through randomly generated patient IDs.

The INBreast database comprises full-field digital mammography (FFDM) images, covering screening, diagnostic, and follow-up cases. Diagnosis is established when screening reveals signs of abnormality, while follow-up images involve cases where cancer was previously detected and treated. The collection includes a total of 115 cases, with 90 having two images (MLO and CC) of each breast, and the remaining 25 cases representing women who underwent a mastectomy, with two images of only one breast included, resulting in a total of 410 images. Furthermore, eight of the 91 cases with two images per breast also have images acquired at different times (follow-up).

A distinctive feature of this dataset is its annotations, which were carried out by an expert in the field and subsequently validated by a second expert between April 2010 and December 2010. These annotations are stored in XML format and were performed using OsiriX, an open-source PACS workstation running on a Macintosh platform. The annotated dimensions include Asymmetry, Calcification, Cluster, Masses, Distortion, Spiculated region, and Pectoral muscle. Additional information provided in the database includes the patient's age at the time of image acquisition, family history, ACR breast density annotation, and BI-RADS classification.

While the above information was used to train and evaluate the performance of the individual classifiers described in Section 3.1, INBreast does not provide ground truth information to evaluate retrieval models. To achieve this goal, a dedicated software [20] was developed and information was provided by a specialist (the same specialist that provided the INBreast annotations).

In total, 102 records were made in the database, i.e. 102 images were analysed out of the 410 that existed. Figure 3 shows an example of ground truth obtained by the expert.

Given the eight (8) dimensions for the mammograms (vertical axis), the graph in Figure 4 shows the count by input (line) of equal dimensions between the probe image and the image chosen by the specialist. This count can vary between 0 and 8 equal dimensions. The overall average distribution is also visible in a dashed red line.

An overview of the distribution of the averages of equal values for each dimension that the expert has chosen is shown in Figure 5. 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.

3.4 Experimental Setting

This study was conducted using Python, with Tensorflow, version 2.11.0. The programming environment for importing the libraries was Google Colab with the NVIDIA T4 Tensor Core GPU.

Data was divided into 70% for training, 20% for validation, and 10% for testing. Having in mind that there may be more than one image per patient, and 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 RESULTS

This section is subdivided into two parts. Section 4.1 present the evaluation of the individual classifiers, while Section 4.2 contains the formal evaluation of the retrieval models.

4.1 Individual classifiers

Table 1 displays the metrics corresponding to both the training set and the 20% validation set. The "Training" column denotes the employed transfer learning methodology. When set to *false*, all layers, except the last one, are frozen; conversely, when set to *true*, all layers undergo training. Inception V3 emerges as the most frequently chosen architecture.

Concerning the outcomes on the test set, where "Training" was configured as *True*, as presented in Table 2, the lowest losses were observed for the Asymmetries dimension, followed by Laterality, BI-RADS, Distortions, View, Calcifications, Density, and Mass. In terms of accuracy, perfect scores were attained by the Asymmetries and Distortions dimensions, succeeded by Mass, Calcifications, View, and Laterality. Multi-class problems such as Density (four ACR classes) and BI-RADS (six classes) exhibit relatively lower performances, as anticipated.

We would like to note that the development of the individual classifiers is not the ultimate goal of this work. Any other architecture or image processing technique can be used [11, 15].

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 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 [2], 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 [3], mammographic breast density is a strong reproducible risk factor for breast cancer. For w_{birads}, the value 0.172 is set because according to [16] 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 [9] 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 [7] 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 [8] 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 [13], 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 (see Section 3.3). The weights correspond to the normalised ratio between the number of selected images that have each dimension in common with the probe image. 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 3.

For an illustration of the retrieval results, the probe image shown in Figure 6 was used, where the ground truth for each dimension is also shown.

Quantitative retrieval results are summarised in the box-plot of Figure 7. To create this box-plot, all of the images, in turn, were considered probe image. Next, the position that the retrieval model placed the image chosen by the specialist (see Section 3.3) is recorded. This list of positions was used to create the box-plot.

Figure 3: Ground Truth examples

Table 1: Classification results on the train and validation sets

Dimension	Chosen Architecture	Training	Train		Validation	
			Loss	Accuracy	Loss	Accuracy
Assymetries	ResNet50	False	0.1773	0.9652	0.1854	0.9560
		True	0.0657	0.9756	0.2130	0.9560
BI-RADS	InceptionV3	False	1.3728	0.5345	1.3837	0.4945
		True	0.1394	0.9690	1.5889	0.4835
Calcifications	InceptionV3	False	0.5318	0.7561	0.5209	0.7692
		True	0.2188	0.9338	0.7231	0.3297
Density	InceptionResnetV2	False	1.3694	0.3241	1.3096	0.3043
		True	0.1458	0.9759	0.1974	0.3913
Distortions	InceptionV3	False	0.0344	0.9965	0.1602	0.9780
		True	0.0050	1.0000	0.1902	0.9780
Laterality	InceptionV3	False	0.1635	0.9652	0.1301	0.9778
		True	0.0019	1.0000	0.8202	0.5889
Mass	InceptionV3	False	0.5526	0.7561	0.6605	0.6703
		True	0.0049	1.0000	0.9285	0.6593
View	InceptionResnetV2	False	0.6892	0.5157	0.6864	0.5444
		True	0.0093	0.9965	0.6945	0.5222

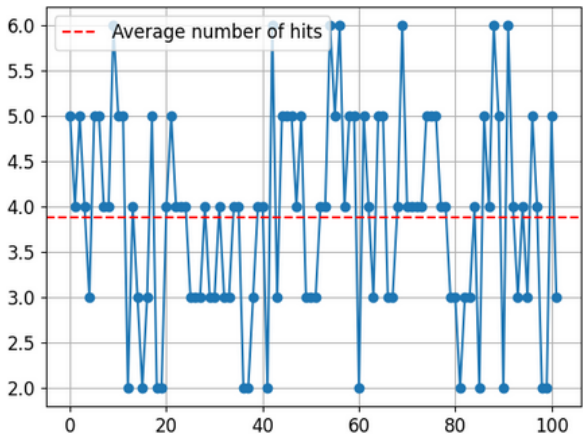


Figure 4: Number of equal values per line

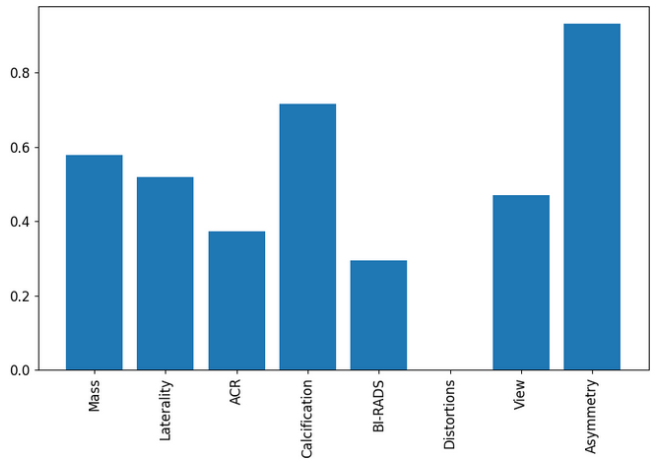


Figure 5: Average Equal Values per dimensions

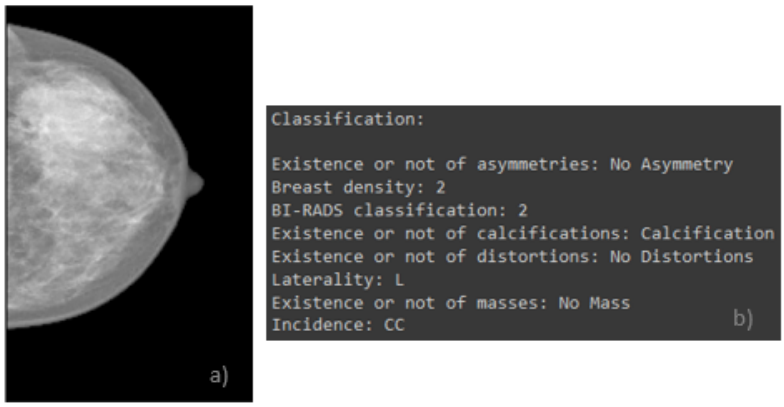


Figure 6: Probe image example along with the ground truth (ground truth is not used during retrieval)

Table 2: Classification results on the test set

Dimension	Chosen Architecture	Loss	Accuracy
Assymetries	ResNet50	0.0209	1.0000
BI-RADS	InceptionV3	1.4696	0.0625
Calcifications	InceptionV3	3.7794	0.7812
Density	InceptionResnetV2	6.3784	0.2500
Distortions	InceptionV3	1.7830	1.0000
Laterality	InceptionV3	0.3020	0.6588
Mass	InceptionV3	7.7452	0.8125
View	InceptionResnetV2	2.3246	0.7688

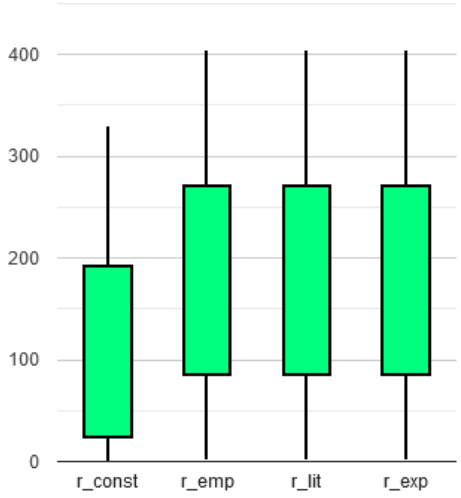


Figure 7: Retrieval evaluation

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.

Upon further exploration of the results, it could also 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. This further enhances the performance of r_{const} when compared with the other models.

5 CONCLUSION

Breast cancer is a widespread illness that affects a large number of people, with women being more at risk of 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 problem mentioned above since early detection of breast cancer is essential for a favourable prognosis.

A retrieval model has been thought 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 classification models were created, one for each dimension, including breast density, asymmetries, BI-RADS classification, the 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, 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.

From the review of the state of the art, we observed that none of the studies under examination make use of the INBreast database, which is a key dataset in the present study. Furthermore, the methods scrutinised in the review predominantly rely on traditional

Table 3: Summary of weights in different retrieval 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

feature extraction techniques. In contrast, the current study stands out by incorporating deep learning classifiers.

As for the results obtained by the individual classification models in the test phase, according to each dimension, the best accuracy achieved is 100% both for Asymmetry and Distortion dimensions. Mass identification achieved 81%, the presence of calcifications 78%, CC or MLO views 77%, and Right or Left 66%. Lower accuracies of 23% and 6% were attained for the multi-class problems of breast density and BI-RADS classification. 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. In the present preliminary experiments, all weights uniformly achieved the best performances.

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. Other architectures will also be experimented with. For the retrieval models, it is planned, for example, to test 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.

REFERENCES

[1] Dina Abdelhafiz, Clifford Yang, Reda Ammar, and Sheida Nabavi. 2019. Deep convolutional neural networks for mammography: advances, challenges and applications. *BMC bioinformatics* 20 (2019), 1–20.

[2] Yara Abdou, Medhavi Gupta, Mariko Asaoka, Kristopher Attwood, Opyrchal Mateusz, Shipra Gandhi, and Kazuaki Takabe. 2022. Left sided breast cancer is associated with aggressive biology and worse outcomes than right sided breast cancer. *Scientific Reports* 12, 1 (2022), 13377.

[3] Akila Anandarajah, Yongzhen Chen, Graham A. Colditz, Angela Hardi, Carolyn Stoll, and Shu Jiang. 2022. 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 (2022), 101.

[4] Melina Arnold, Eileen Morgan, Harriet Rumgay, Allini Mafra, Deependra Singh, Mathieu Laversanne, Jerome Vignat, Julie R. Gralow, Fatima Cardoso, Sabine Siesling, and Isabelle Soerjomataram. 2022. Current and future burden of breast cancer: Global statistics for 2020 and 2040. *The Breast* 66 (2022), 15–23.

[5] Silvia Bessa, Inês Domingues, Jaime S Cardosos, Pedro Passarinho, Pedro Cardoso, Vitor Rodrigues, and Fernando Lage. 2014. Normal breast identification in screening mammography: a study on 18 000 images. In *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 325–330.

[6] Liga Portuguesa Contra o Cancro. [n. d.]. Cancro da Mama : Liga Portuguesa Contra o Cancro. http://www.ligacontracancro.pt/cancro-da-mama/ Access date: 2023-01-31.

[7] Sushovan Chaudhury, Manik Rakhra, Naz Memon, Kartik Sau, and Melkamu Teshome Ayana. 2021. Breast Cancer Calcifications: Identification Using a Novel Segmentation Approach. *Computational and Mathematical Methods in Medicine* 2021 (2021), e9905808.

[8] Xiao Chen, Yang Zhang, Jiahuan Zhou, Xiao Wang, Xinmiao Liu, Ke Nie, Xiaomin Lin, Wenwen He, Min-Ying Su, Guoquan Cao, and Meihao Wang. 2022. Diagnosis of architectural distortion on digital breast tomosynthesis using radiomics and deep learning. *Frontiers in Oncology* 12 (2022).

[9] Eva Cuypers, Britt S. R. Claes, Rianne Biemans, Natasja G. Lieuwes, Kristine Glunde, Ludwig Dubois, and Ron M. A. Heeren. 2022. 'On the Spot' Digital Pathology of Breast Cancer Based on Single-Cell Mass Spectrometry Imaging. *Analytical Chemistry* 94, 16 (2022), 6180–6190.

[10] Nagadevi Darapureddy, Nagaprakash Karatapu, and Tirumula Krishna Battula. 2021. Optimal weighted hybrid pattern for content based medical image retrieval using modified spider monkey optimization. *International Journal of Imaging Systems and Technology* 31, 2 (2021), 828–853.

[11] Inês Domingues and Jaime S. Cardoso. 2014. Using Bayesian surprise to detect calcifications in mammogram images. In *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 1091–1094. https://doi.org/10.1109/EMBC.2014.6943784

[12] Inês Domingues, Jaime S. Cardoso, Igor Amaral, Inês Moreira, Pedro Passarinho, João Santa Comba, Ricardo Correia, and Maria J. Cardoso. 2010. 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*. 3158–3161. https://doi.org/10.1109/IEMBS.2010.5627168

[13] Yuanfang Guan, Xueqing Wang, Hongyang Li, Zhenning Zhang, Xianghao Chen, Omer Siddiqui, Sara Nehring, and Xiuzhen Huang. 2020. Detecting Asymmetric Patterns and Localizing Cancers on Mammograms. *Patterns* 1, 7 (2020), 100106.

[14] Amira Jouirou, Abir Baâzaoui, and Walid Barhoumi. 2021. Shared Information-Based Late Fusion for Four Mammogram Views Retrieval using Data-driven Distance Selection.. In *VISIGRAPP (4: VISAPP)*. 144–155.

[15] Francisco Marques, Hugo Duarte, João Santos, Inês Domingues, José P Amorim, and Pedro H Abreu. 2019. An iterative oversampling approach for ordinal classification. In *34th ACM/SIGAPP Symposium on Applied Computing*. 771–774.

[16] Lingsong Meng, Xin Zhao, Jinxia Guo, Lin Lu, Meiying Cheng, Qingna Xing, Honglei Shang, Bohao Zhang, Yan Chen, Penghua Zhang, and Xiaoan Zhang. 2023. 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).

[17] Inês C. Moreira, Igor Amaral, Inês Domingues, António Cardoso, Maria João Cardoso, and Jaime S. Cardoso. 2012. INbreast: Toward a Full-field Digital Mammographic Database. *Academic Radiology* 19, 2 (2012), 236–248. https://doi.org/10.1016/j.acra.2011.09.014 Publisher: Elsevier.

[18] Sonia Jenifer Rayen and R Subhashini. 2019. Mammogram Image Retrieval using IPSO Optimized Anfis Classifier. *Int. J. Innov. Technol. Explor. Eng* 8, 9 (2019), 799–804.

[19] Sonia Jenifer Rayen and R Subhashini. 2021. An efficient mammogram image retrieval system using an optimized classifier. *Neural Processing Letters* 53 (2021), 2467–2484.

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

[21] A Selvi and S Thilagamani. 2023. Scale Invariant Feature Transform with Crow Optimization for Breast Cancer Detection. *Intelligent Automation & Soft Computing* 36, 3 (2023).

[22] Vipul Sharma. 2022. Mammogram Image Retrieval System Using Texture and Semantic Features. In *Journal of Physics: Conference Series*, Vol. 2267. IOP Publishing, 012071.