Algorithm for the classification of lesions in digital breast tomosynthesis

Karol Juliana Galvis Florez and Mirna Lorena González Santamaría

Degree Work to opt for the degree of Electronic Engineer

Director

Said Pertuz

PhD in Computer Science

Universidad Industrial de Santander

Facultad de Ingenierías Fisicomecánicas

Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones

Bucaramanga

2022

Acknowledgments

This project would not have been possible without the support of many people. Many thanks to my adviser, Angie Nicole Hernández Durán, who read our numerous revisions and helped make some sense of the confusion. Also thanks to our director Said Pertuz, who offered guidance and support during the Bachelor's Thesis. Thanks to the Universidad Industrial de Santander for providing us with the financial means to finalize the last semesters and the tools to complete this work. And finally, thanks to our parents, and numerous friends who endured this long process with us, always offering support and love, especially our friend Gabriela Galvis Gómez who participated in a part of the process and giving us unconditional support since she met us.

Contents

List of Figures

List of Tables

Resumen

Título: Algoritmo para la clasificación de lesiones en tomosíntesis digital de mama [*]

Autor: Karol Juliana Galvis Florez y Mirna Lorena González Santamaría [**]

Palabras Clave: DBT, clasificación, lesiones mamarias, algoritmo.

Descripción: En los últimos años se han desarrollado diferentes herramientas para ayudar a la detección temprana del cáncer de mama a partir de modalidades de imagen como la mamografía, la resonancia magnética, la tomosíntesis, entre otras. La tomosíntesis digital de mama (DBT, por sus siglas en inglés) es una nueva modalidad de imagen que ha demostrado reducir la tasa de rellamadas falsas en comparación con la mamografía. Sin embargo, las investigaciones sobre el análisis computarizado de imágenes DBT son escasas debido a que la mamografía es la modalidad más utilizada y el foco de atención de la comunidad investigadora hasta el momento. Debido a su ventaja sobre la mamografía, es muy importante explorar el uso de DBT para la clasificación automática de lesiones mamarias, como un esfuerzo para contribuir con este tema de investigación. Este trabajo se centra en la clasificación de lesiones en imágenes DBT de la base de datos Breast Imaging Archive, diseñando e implementando un algoritmo de aprendizaje automático para esta tarea. En primer lugar, se realizó un preprocesamiento de las imágenes para mejorar el proceso posterior de extracción de características. Luego, se entrenó un modelo lineal generalizado con regresión por pasos, que emplea la regresión logística como clasificador, y usa regresión paso a paso, *forward* y *backward*, para seleccionar las características más relevantes para el modelo final. Finalmente, la validación se realizó mediante *leave-one-out cross-validation*. El algoritmo desarrollado tiene un rendimiento aceptable de AUC = 0,651 (IC 95%: 0,537 - 0,766), que podría mejorarse en futuras investigaciones agregando más anotaciones de diferentes pacientes al conjunto de datos e incorporando otras etapas, como la fase de detección del cáncer de mama. Además, este trabajo introdujo una nueva forma de clasificar las lesiones utilizando bases de datos públicas con anotaciones de imágenes DBT limitadas.

---

[*] Trabajo de Grado

[**] Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y telecomunicaciones. Director: Said Pertuz, PhD in Computer Science.

Abstract

Title: Algorithm for the classification of lesions in digital breast tomosynthesis [*]

Author: Karol Juliana Galvis Florez and Mirna Lorena González Santamaría [**]

Keywords: DBT, classification, breast lesions, algorithm.

Description: In recent years, different tools have been developed to help early detection of breast cancer from imaging modalities such as mammography, magnetic resonance imaging, tomosynthesis, among others. Digital Breast Tomosynthesis (DBT) is a new imaging modality that has been demonstrated to reduce false recall rates compared to mammography. However, researches on the computerized analysis of DBT images is scarce due to the fact that mammography are the most widely used modality, and the focus of the research community at the moment. Due to its advantage over mammography, it is very important to explore the use of DBT for the automatic classification of breast lesions, as an effort to contribute to this research topic. This work focuses on the classification of lesions in DBT images from the Breast Imaging Archive database, designing and implementing a machine learning algorithm for this task. First, a pre-processing of the images was carried out to improve the subsequent process of feature extraction. Then, a generalized linear model with stepwise regression was trained, which employs logistic regression as classifier, and uses forward and backward stepwise regression to select the most relevant features for the final model. Finally, validation was performed using leave-one-out cross-validation. The developed algorithm has an acceptable performance of AUC = 0.651 (95% CI: 0.537 - 0.766), which could be improved in future research by adding more annotations from different patients to the dataset and incorporating other stages, such as breast cancer detection phase. Additionally, this work introduced a new way to classify lesions using public databases with limited DBT image annotations.

---

[*] Bachelor's Thesis

[**] Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y telecomunicaciones. Director: Said Pertuz, PhD in Computer Science.

Introduction

Breast cancer has surpassed lung cancer as the cancer with the highest incidence in 2020, with an estimated 2.3 million new cases, representing 11.7% of all cancer cases. In the same year, it was fifth leading cause of cancer mortality worldwide, with 685,000 deaths. Among women, breast cancer accounts for 1 in 4 cancer cases and for 1 in 6 cancer deaths, ranking first for incidence in the vast majority of countries, and for mortality in 110 countries (Sung et al., 2021). Cancer arises through a series of somatic DNA alterations that culminate in unrestrained cell proliferation. They can appear as a consequence of random errors in replication or exposure to carcinogens (such as radiation) and can be exacerbated by defects in DNA repair processes. Many cancers appear sporadically but in some cases they occur in families that carry a germline mutation in a cancer gene. (Bunz and Vogelstein, 2022). The unrestrained cell proliferation often forms a mass of tissue called a lump, growth, or tumor (Gowri and Amudha, 2014). Methods for the automatic detection of such lesions using medical images have been widely studied, but they face several challenges regarding the nature of the data used, such as low contrast and high noise generated by the acquisition process. This makes the analysis difficult, specially on images obtained from denser breasts.

One of the imaging modalities used to detect anomalies in the breast is *digital breast tomosynthesis* (DBT). DBT is a pseudo-three-dimensional image, whereby it is possible to visualize the lesion, if there is one, in the slices closest to it. It is an advanced form of mammography that consists on taking several images of the breast using low doses of x-rays at different angles (Hous-

sami et al., 2017), the individual images are then reconstructed into a series of thin, high-resolution slices typically 1 mm thick (Smith, 2012). DBT achieves a clearer objectification of the edges of the lesions and architectural distortions (Joaquín José Mosquera Osés, 2012), which is of great help in lesion classification. In addition, DBT has been shown to increase detection rate, improve lesion characterization, tumor staging, radiologist's performance, workflow and decrease callback rate (Rocha García and Mera Fernández, 2019). Despite all its advantages, the use of DBT for learning algorithms, computerized detection and lesion classification to date, we have only found one work directly related to DBT (Ricciardi et al., 2021). A detailed revision of the state of the art will be presented in section 2.

The aim of this project is to develop an algorithm for the classification of lesions in DBT. For our experiments, we use images from the public dataset released by (Buda et al., 2020) in the Cancer Imaging Archive[1]. This database includes 101 patients of which 39 have cancer and 62 have a benign lesion; This corresponds to a total of 224 digital breast tomosynthesis images with the four standard views: craniocaudal and mediolateral, of each breast.

Due to the fact that we work with clinical information, the experimental design is restricted to the use of public databases in order to comply with current regulations and preserve the anonymity of the patient. Also, considering the large amount of images and their high resolution, restrictions must be incorporated that limit the computation time and the use of computational resources, thus allowing the execution of the algorithm on desktop computers.

---

[1] https://www.cancerimagingarchive.net/

1. Objectives

1.1. General objective

To design an algorithm for the classification of lesions in Digital Breast Tomosynthesis images.

1.2. Specific objectives

1. To select and implement feature extraction algorithms for the analysis of DBT images.

2. To design an image processing pipeline for the classification of DBT images from previously extracted features.

3. To evaluate the performance of the algorithm and analyze the results obtained.

2. Literature review

Previous research has shown that the classification of breast lesions is a challenging task, the majority of the methods have focused on classifying mammography images using machine learning and deep learning algorithms. Based on a review of the literature, in this section we describe the most relevant stages for the analysis of DBT and mammography images: image pre-processing, features extraction and employed models.

2.1. Image pre-processing

When researching related works, we found that most of them use a first stage that corresponds to the pre-processing of the images, which is useful to improve their quality and make the feature extraction phase more reliable. The pre-processing stage can involve several processes, such as filtering, used to reduce impulsive or salt and pepper noise, the opening filter to eliminate objects smaller than the structuring element, and the closing filter, used to fill in gaps in the image without greatly altering the area of the objects. A common pre-processing method is region of interest (ROI) segmentation, where single ROI methods can be used, in which areas such as the whole breast or the retroareolar region are selected (Africano et al., 2020); and iterative ROI segmentation based on the active contour, whose advantage is that it preserves the information, the edge and the shape of the mass region (K. Yuvaraj, 2022). Other pre-processing methods are used to improve the contrast of the image by assigning a new intensity to each pixel according to an adaptive transfer function (Yu and Bajaj, 2004). Margin clipping is also often used to remove image regions that may affect feature extraction, or for isolated region removal, to remove objects

that are not part of the subject of interest. Finally, the histogram equalization is usually applied to obtain a uniform distribution of the image, the binarization to reduce the image to a gray scale of two unique values, the normalization process to change the range of intensity values of the pixels and the resizing to standardize image sizes to a given resolution.

2.2. Feature extraction

The feature extraction from an image is of vital importance because it contains information about it, including the lesion if there is one, and it is what is ultimately used to train the classification model (Thawkar and Ingolikar, 2017). The features extracted from medical images are classified into three types: intensity, shape, and texture. Intensity features include average gray level, average contrast, smoothness, skewness, uniformity and entropy (Thawkar and Ingolikar, 2017). The shape features are based on the shape of the detected mass such as perimeter, area, major axis, minor axis, thinness ratio, eccentricity, dispersion, compactness, circularity, roundness, elongation, and shape index (Mohamed and Salem, 2018). Finally, for the texture features, the difference between the gray levels in the image is analyzed.

For lesions feature extraction, the following types of features are used: first-order statistics, such as, mean, standard deviation, skewness, kurtosis, entropy, maximum and minimum intensity, all of these depend on a single pixel value; second order statistics, where, from the mass, we get the Gray Level Run Length Matrix (GLRLM) and the Gray Level Co-occurrence Matrix (GLCM) which is used to compute features like contrast, correlation, energy, homogeneity and cluster shade (Mohamed and Salem, 2018). Feature extraction also depends on the chosen model, deep learning models, for example, usually work with the entire image itself, and not directly with its features.

2.3. Review of classification models

We performed a literature review on the classification models that exist for both imaging modalities. Machine learning algorithms and deep learning algorithms are often used to create these models. The most widely used machine learning algorithms in this area belong to the category of supervised learning. Some of them are Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Logistic Regression (LR), Random Forest (RF). For deep learning algorithms there are Convolutional Neural Networks (CNN), Deep Convolutional Neural Networks (DCNN) and Artificial Neural Network (ANN).

For classification using only mammography, the following algorithms have been employed. ANN is used due to its versatility, power, and scalability. It is composed of hidden layers, making them ideal to approach large and highly complex tasks. SVM is well-qualified for performing task such as regression, linear or nonlinear classification in complex but small or medium sized data sets and anomalies detection. Fuzzy Support Vector Machines (FSVM) are linear classifiers based on statistical learning theory and have been developed for numeric prediction that share many of the properties encountered in the classification case: they produce a model that can usually be expressed in terms of a few support vectors and can be applied to nonlinear problems using kernel functions (Witten et al., 2017). Finally, CNN learns directly from data by extracting features and trains the network with a huge number of samples.

For the task of classification and detection in DBT, DCNN is helpful, a type of artificial neural network composed of convolutional layers and fully connected layers within a deep archi-

tecture. During training, a DCNN learns patterns through the kernels in each convolutional layer (Ricciardi et al., 2021).

Few researchers have studied the lesion classification task comparing both mammography and tomosynthesis, where it can be observed that the performance of tomosynthesis tends to be better (Gui et al., 2019). In classification for mammography and DBT the models that are used in the literature up to this day are RF, which is very powerful and capable of fitting complex data sets, looking for the best feature among a random subset of features, and KNN that is a simple non-parametric learning method based on instances that does not require a learning phase. The proper functioning of the method depends on the choice of a number of parameters such as the parameter k, which represents the number of neighbors chosen to allocate the class to the new point, and the distance used (Mohamed and Salem, 2018). Logistic Regression is used to classify into classes and estimate the probability, between 0 and 1, of an event belonging to one class. The model performs well when the data set is linearly separable, it is also easy to implement and interpret.

Table 1 presents a summary of the previous researches on lesion classification, where the modality can be seen as DBT, MAMMO for mammography and digitized MAMMO; The classification model used, if pre-processing was applied, which type was it, features extraction, the size of the database and the accuracy obtained.

Table 1

Summary of relevant research

Researches	Modality	Model	Pre-processing	Features extraction	Database	Performance
(Kathale and Thorat, 2020)	MAMMO	RF	Median Filtering and Resizing	Mean, Entropy and GLCM	$mini-MIAS$[2]: 24 images benign and malign.	ACC: 95.3%
(Mohamed and Salem, 2018)	MAMMO	ANN, SVM and KNN		Shape, texture, GLCM, GLRLM	$DDSM$[3]: 250 images (116 benign and 134 malign)	ACC: 98.9%, ACC: 97.7%, ACC: 96.6%
(Bai and Qian, 2008)	MAMMO	FSVM	Median filter, cropping margins, eliminating isolated regions, equalizing.	Media, varianza, Bias, kurtosis	$WBCD$[4]: 683 images	ACC: 86.13%
(Sun et al., 2020)	MAMMO	CNN and self-paced learning		cross-entropy loss function	$DDSM$[3]: 1670 images (892 benign and 778 malign)	ACC: 76.46%
(Ricciardi et al., 2021)	DBT	DCNN			$Hospital_1$[5]: 4962 slices with 137 lesions (33 benign and 104 malign), $Hospital_2$[6]: 3024 slices (1127 benign and 1897 malign)	ACC: 94.0%
(Gui et al., 2019)	MAMMO and DBT	KNN, RF and LR	Only contours of the tumors were maintained (ROI).	Fractal features	Peking University Shenzhen Hospital: MAMMO: 395 images (77 benign and 318 malign). DBT: 1476 images (270 benign and 1206 malign).	MAMMO: AUC: 0.788, AUC: 0.861, AUC: 0.846 DBT: AUC: 0.956, AUC: 0.982, AUC: 0.967
(Deep Deb et al., 2020)	MAMMO	DCNN	ROI Acquisition,Data Augmentation, Image Thresholding, Gaussian filter, contrast limited adaptive histogram equalization (CLAHE) and Zero-mean normaliztion	Six different DCNN	$CBIS-DDSM$[7]: 1295 training dataset images (673 benign and 622 malign) and 445 test dataset images (298 benign and 147 malign)	AUC: 0.76
(Amrane et al., 2018)	Digitized MAMMO	KNN and NB			BCD[8]: 683 samples(444 benign and 239 malign)	ACC: 97.51%, ACC: 96.19%
(Ara et al., 2021)	Digitized MAMMO	SVM, DT, NB, KNN, RF and LR		Radius, Texture, Perimeter, Area, Smoothness, Compactness, Concavity, Concavity points, Symmetry and Fractal dimension features	$WBCD$[4]: 357 benign and 212 malign	ACC: 96.%, ACC: 95.1%, ACC: 92.3%, ACC: 95.8%, ACC: 96.5%, ACC: 94.4%

Note.

[2]Mammographic Image Analysis Society (MIAS).

[3]Digital Database for Screening Mammography (DDSM).

[4]Wisconsin Breast Cancer Dataset (WBCD).

[5]Azienda Ospedaliera Cardarelli ($Hospital_1$).

[6]Azienda Ospedaliera Universitaria San Giovanni di Dio Ruggi d'Aragona ($Hospital_2$).

[7]Curated Breast Imaging Subset of DDSM (CBIS-DDSM).

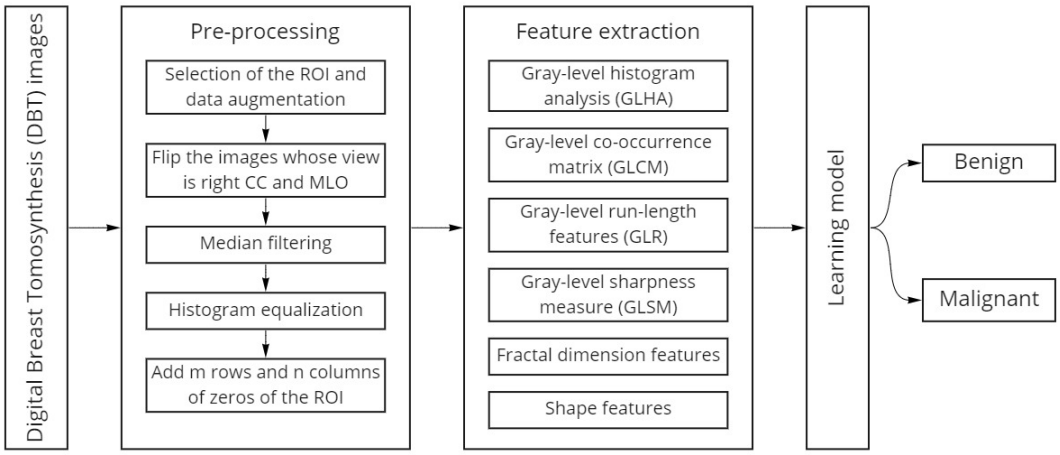
[8]Breast Cancer Dataset (BCD).

3. Materials and methods

This section describes the dataste used, the classification algorithm, the complementary steps of the process and the methodology, as seen in Figure 1, which describes the steps that were carried out. Among these is the description of the pre-processing performed on the images, the features used and the machine learning model.

Figure 1

Methodology diagram



3.1. Dataset

The database used in this work was found thanks to the research of (Buda et al., 2020). The Cancer Imaging Archive[1] (TCIA) data is organized as “collections”; typically these are pa-

[1] https://www.cancerimagingarchive.net/

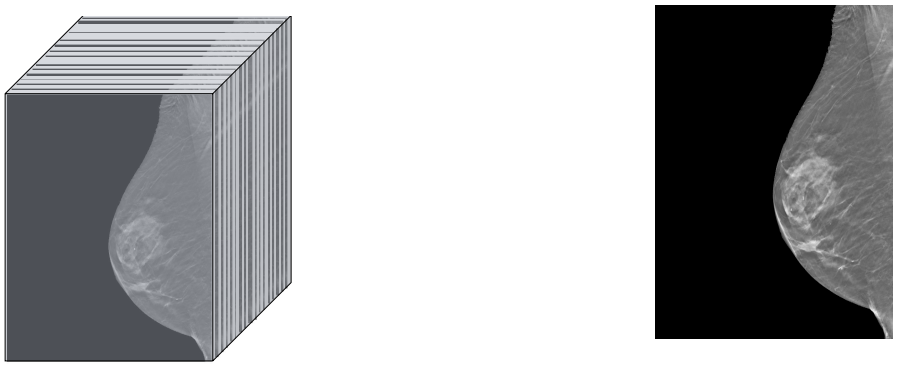
tient cohorts related by a common disease (e.g. breast cancer), image modality or type (MRI, CT, DBT, etc) or research focus. For this research we used the public collection Breast Cancer Screening-DBT which contains 22032 reconstructed DBT volumes belonging to 5610 studies from 5060 patients. The data for this collection is provided by Duke University Hospital/Duke University and contains four components: (1) DICOM images, (2) a spreadsheet indicating which class each case belongs to, (3) annotation boxes, and (4) Image localization within the database for patients/studies/views. For each patient, there is at least one craniocaudal (CC) and mediolateral oblique (MLO) views available for the left or right breast. While in mammography, a view (whether CC or MLO) refers to a single image, in DBT it refers to a stack of reconstructed slices taken for the same breast (Figure 2(a)). Every two-dimensional slice in a DBT volume will be referred as image (Figure 2(b)). The radiology reports where made by two radiologists (18 and 25 years of experience) at Duke University, who annotated the studies using a rectangular box enclosing a biopsied tumor only in the central slice. These studies had either mass or architectural distortion which resulted in a biopsy with a benign or cancer finding. The dates of these studies lie between August 26, 2014 and January 29, 2018.

Of the 5060 patients, only 101 patients have the radiologist's annotations that are necessary for the algorithm development. These are specified in the file Training set Phase 2 - Boxes, that indicates the patient ID, view, volume slices, slice where the center of the lesion is located, width, height and class of the lesion. For this reason, the dataset used only includes 101 patients, of which 39 have cancer and 62 have a benign lesion. Initially, we selected the annotated slice or slices of the patient, this correspond to a total of 224 DBT images. We performed a visual inspection of

all the 224 images where the lesion was located to check if there was any damage. We found that it was not possible to visualize the lesion in 12 images of the benign class and 2 images of the cancer class, this images were excluded and the final dataset had 95 patients with a total of 210 DBT images of which 125 have cancer and 85 have a benign lesion.

Figure 2

(a) DBT volume, (b) DBT slice



(a) (b)

Given the relatively small number of images due to limited number of radiologist's annotations, new data was added from the original database. The strategy for data augmentation was to include a certain number of slices before and after the central slice.

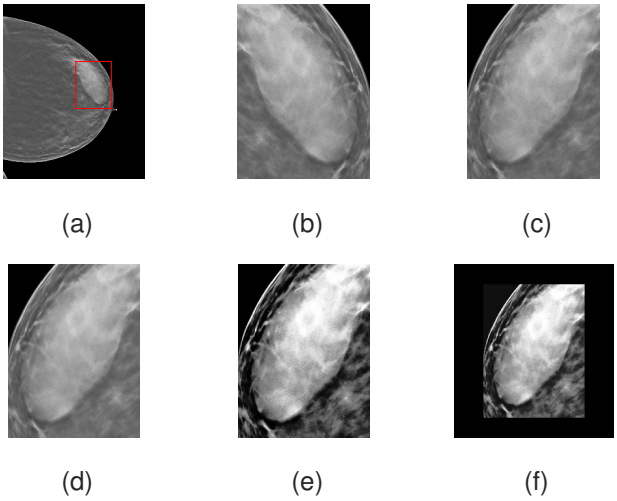
3.2. Image pre-processing

The initial step in this phase was selecting the annotated area inside the image as ROI. The next step was to orient all the lesions to the right by doing a flip in the images whose view is left CC and MLO. Then, median filtering was performed to filter out the noise of the grayscale ROI. Following that, we carried out histogram equalization to improve the contrast. Finally, we noticed

that the lesions have significant differences in the size, which we consider as a relevant feature, whereby the size of the ROI was standardized by adding m rows of zeros to the top and bottom of the image and n columns of zeros to the right and left of the image, the previous steps are shown in Figure 3.

Figure 3

Shows a single slice for (a) Digital Breast Tomosynthesis image in the CC view (left), (b) cropped ROI, (c) flip to the right, (d) median filtering, (e) histogram equalization and (f) filling with zeros



3.3. Model

Due to the fact that one of the restrictions of our work is to ensure that the algorithm can be executed on desktop computers, we choose a model that worked with features instead of the whole image because when the full image is used to train the model, it works with a larger amount of data and takes longer to process, requiring more computational resources. Another limitation was the number of images of patients with annotations, so it was not possible to work with models that depended on the amount of data, such as Deep Learning models.

The model used for the classification of breast lesions in this work was a generalized linear model by stepwise regression, which is an effective mechanism for calculating probabilities between 0 and 1 for a binary classification problem. In a linear model, the sigmoid function (1) will yield a probability for the output (2) of a model trained with logistic regression, given the model's learned weights w_i, the bias b and the feature values x_i, the latter being the input of the model.

$$y^{'} = \frac{1}{1+e^{-z}} \tag{1}$$

$$z = b + w_1x_1 + w_2x_2 + ... + w_Nx_N \tag{2}$$

Logistic regression was used in this case because it allows binary classification, benign and malignant. It is also one of the most used supervised machine learning algorithms for classification tasks, like predicting if an e-mail is spam or not spam, or if a tumor is malignant or not malignant.

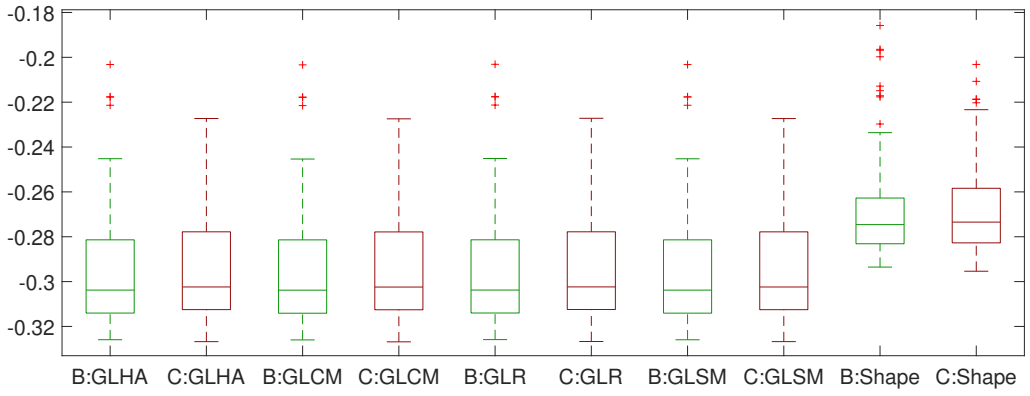
Usually, in the training of the model, there are features that are redundant or not so relevant for the classification, so a selection process must be used. Among these techniques is forward selection, which starts with an empty model and at each step forward will add a feature that will help improve the model. Another technique similar to forward selection is stepwise selection, which we used. It builds a regression model by adding or removing the predictor variables after each step in which a variable was added, all candidate variables in the model are checked to see if their significance has been reduced below the specified tolerance level. If a nonsignificant variable is found, it is removed from the model (Surhone et al., 2022).

3.4. Features selected

As mentioned, we train our classification model using extracted features. Since some of the classification algorithms for mammography use feature extraction, we decided to extract the same types of features from DBT images and analyze their performance. As it is shown in the research by Guangchao et al. (Gui et al., 2019), these two modalities have similarities considering that both manage their images in grayscale and use the same type of views, namely CC and MLO, so it is not far fetched to use the features used in mammography for DBT. Figure 4 shows box plots for the distributions of some of the features used: gray-level histogram analysis (GLHA), gray-level co-occurrence matrix (GLCM), gray-level run-length features (GLR), gray-level sharpness measure (GLSM), fractal dimension features and shape measurements. The features were extracted using the OpenBreast tool, wich is a Matlab-based open source software for the computerized, automatic analysis of mammography images for breast cancer risk assessment (Pertuz et al., 2019).

Figure 4

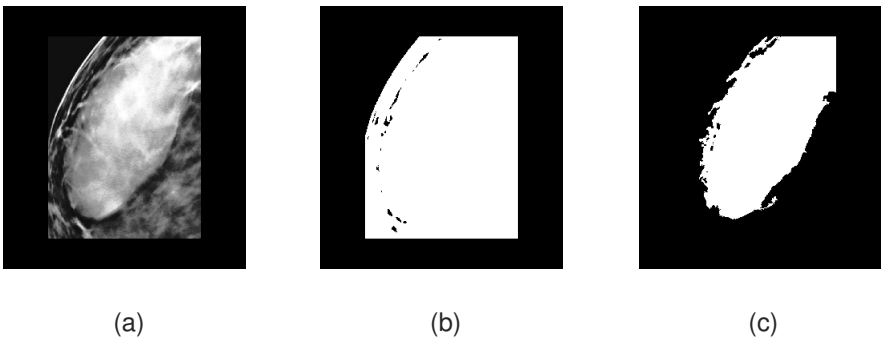
Box plot of the distributions of some of the selected features, for benign (B) and malignant (C) lesions



During feature extraction, two masks were used, in order to effectively exclude the background of the image, as present in Figure 5. The mask in Figure 5(b) was used for the features GLHA, GLCM, GLR, GLS and fractal dimension. The mask in Figure 5(c) was only used for the shape measurements.

Figure 5

Masks used for the extraction of features. (a) Pre-processed image. Segmented image using a mask for (b) gray-level and fractal features and (c) shape features



3.5. Performance measurements

For the evaluation of the machine learning algorithm performance, we split the data into training set and test set. The training set is used to train the model and the test is used to provide an evaluation of the final model in unseen data. One of the problems that might arise when training a model is over or underfitting. When the number of training samples is elevated, or the hyperparameters have been modify to produce a low error rate in training set, the model overlearns from the data, which is known as overfitting. This is not convenient because the goal of the model is to make good predictions in unseen data, and models that overfit on the training set do not tend to generalize well. On the other hand, when the number of training samples is too low, the training has a high

error and the model will be incapable of learn correctly from the data, this is underfitting. There are several Cross-validation techniques to split the data in different ways witht the aim to obtain a more reliable performance measurement, and avoid reporting only measurements that might be affected by the previous scenarios. We used Leave-one-out cross-validation (LOOCV), which is a special case of cross-validation where the number of folds equals the number of instances in the data set. Thus, the learning algorithm is applied once for each case, using all other instances as a training set and using the selected instance as a single-item test set (Sammut and Webb, 2010).

First, validation was performed at the patient level instead of the image level to prevent images of the same patient from being in both training and test sets. To measure the performance, we used the ROC curve, showing the diagnostic ability of a classification model for all thresholds; the area under the ROC curve (AUC), to measure a binary classifier given the probability that the model will categorize a positive sample higher than a negative; and the accuracy (3) regarding correct/incorrect classification where TP, FN, TN and FP are the number of true positives, false negatives, true negatives and false positives, respectively.

Since we are measuring estimations of the population parameters, we have to report them along with their confidence interval (CI). The CI indicates that we can say, with a certain level of confidence, usually 95%, that the population value we are estimating is located within this interval.

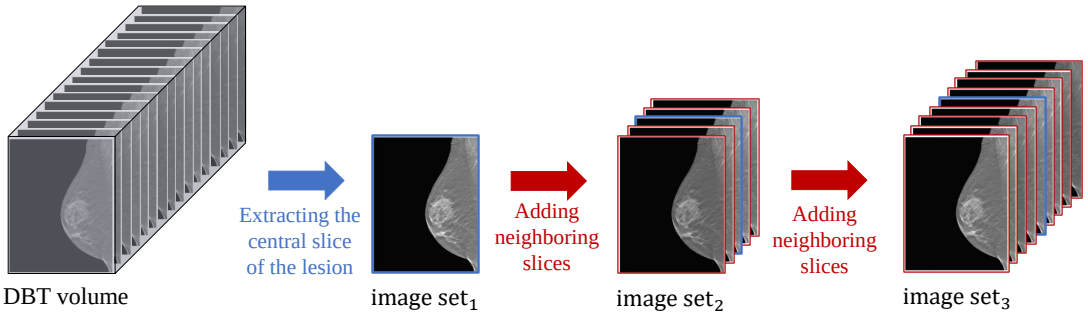
$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

4. Experiments and results

In this section we present the specific experimental designs we used, and the results obtained. For our experiments we used three image sets, extracted from the resulting database described in section 3.1. These are as follows: *image set*$_1$ corresponds to the images using the central slice, *image set*$_2$ contains the image of the central slice of the lesion and 4 neighboring slices, *image set*$_3$ contains the image of the central slice of the lesion and 8 neighboring slices, the image sets are illustrated in Figure 6. Table 2 presents the information on the number of images used for each image sets.

Figure 6

Augmented image sets



The performance of the proposed algorithm was executed in the following hardware environment: Windows 10, Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz 2.70 GHz and 8 GB RAM. The execution time of the training and testing was around 8 minutes for the *image set*$_1$, 27 minutes for the *image set*$_2$ and 37 minutes for the *image set*$_3$.

Table 2

Image sets

Image sets	Patients	Benign lesion	Malign lesion	Images
image set$_1$	95	125	85	210
image set$_2$	95	625	425	1050
image set$_3$	95	1125	765	1890

For each image of our image sets, we performed the pre-processing steps and extracted the features described in sections 3.2 and 3.4, respectively. Finally, we trained our model, as described in section 3.5. To measure the performance of the selected features in each image sets, we used ROC and AUC. The threshold shown in Table 3 was found by the operating point or cut-point, which tells us at what point the model is operating and what is the average threshold in this point.

Table 3

Performance measures

	AUC (95% CI)	Threshold	Accuracy [%]
image set$_1$	0.651 (0.537 - 0.766)	0.540	65.26
image set$_2$	0.667 (0.554 - 0.781)	0.512	63.16
image set$_3$	0.656 (0.542 - 0.771)	0.501	64.21

Figure 7 illustrates the information on Table 3. We can see there is major overlap between the CI, this indicates that there was not much difference using either of the 3 image sets.

Figure 7

AUC comparation for 1: image set$_1$, 2: image set$_2$ and 3: image set$_3$

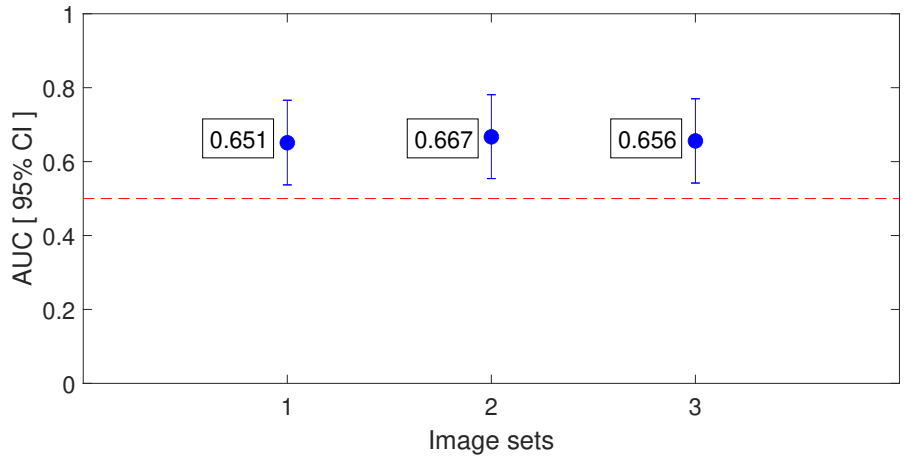


Figure 8 shows the confusion matrix for the 3 image sets, taken from their respective threshold shown in Table 3. It is also observed that the number of true negative values is greater than that of true positives.

Figure 8

Confusion matrix for (a) image set$_1$, (b) image set$_2$ and (c) image set$_3$

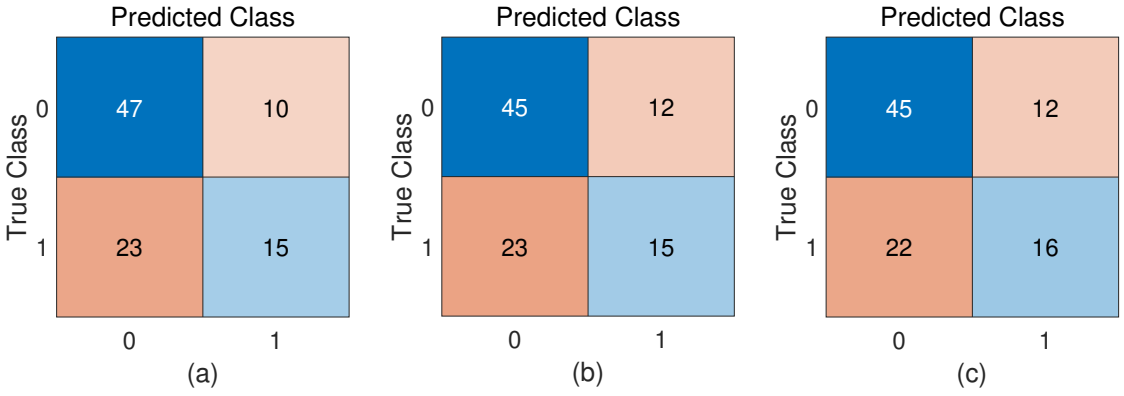
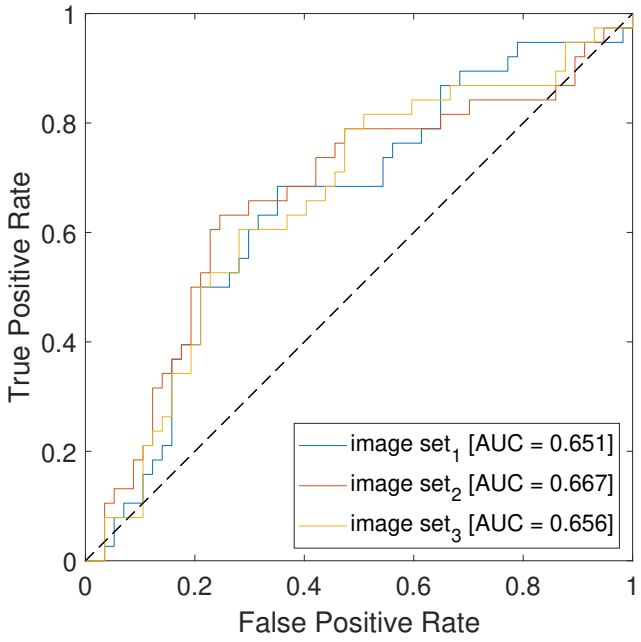


Figure 9 shows a comparison of the ROC curves of the 3 image sets. As can be seen, the

difference between the curves is not that significant but the *image set*$_2$ and *image set*$_3$ shows a small increase respect to the *image set*$_1$.

Figure 9

ROC curves for the image sets



5. Discussion and conclusion

An algorithm for the classification of lesions in DBT, benign or malignant, has been proposed, which included a pre-processing stage for ROI standardization, and the extractor o features commonly used in mammography. Data augmentation was included in two of the scenarios tested, using the neighboring slices. However, as shown in Figure 7 the significant data augmentation was not reflected in a significant improvement on performance, this may be due to the fact that the nearby slices do not provide much new information, in addition to the case that they are training by patients and not by images, so it is more essential to add more patients than images. This consistency is also noticeable by inspecting the confusion matrices, where the distribution of TP, TN, FP and FN remains similar across the image sets. The confusion matrices also tell us that the model classifies bening lesions more accurately, this might be to the fact that there are more benign lesions in our dataset. The opposite case is seen for cancer cases, the amount of malignant lesions classified correctly is inferior than the amount incorrectly classified. Using the threshold shown in Table 3 and the confusion matrix in Figure 8, the respective accuracy was found, which when compared with other investigations shows an acceptable performance, this is remarkable, taking into account the limitations in terms of image sets size and computational capacity. As future work, it would be useful to incorporate more radiologist's annotations to augment the image sets with information that is not redundant and improve the model. In addition, it could include a stage of data harmonization to improve the data through the use of machine learning algorithms and perform a better pre-processing. Regarding the previous work related to DBT in which a classifi-

cation of presence/absence of mass lesions was carried out using DCNN, with a different database, our work showed that it is also possible to carry out a classification using machine learning with computational resources such as presented in section 4. Also, pre-processing techniques were implemented, which have not been used for classification of lesions in DBT. Finally, a rotation like the one presented in (Ricciardi et al., 2021) could be performed on the image sets to analyze if this influences in the performance.

Deep learning models, which could be used for this classification task, have the downside of requiring large image sets and it is a problem because medical imaging data that is publicly available is limited. Taking into account the machine learning model used, Logistic regression, reached an acceptable performance in scenarios where image sets have reduced sizes with a maximum AUC of 0.667 (95% CI: 0.554 - 0.781) comparing it with previous works the performance is lower but it allows a good enough execution of the algorithm on desktop computers. Regarding the related works, the results are not far from those expected, taking into account the limitations and restrictions of the work. This algorithm can be a support tool for radiologists, thus helping to have an early diagnosis in case there is a lesion.

Bibliography

Africano, G., Arponen, O., Sassi, A., Karivaara-Mäkelä, M., Holli-Helenius, K., Rinta-Kiikka, I., Lääperi, A.-L., & Pertuz, S. (2020). A comparison of regions of interest in parenchymal analysis for breast cancer risk assessment. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 1136–1139. https://doi.org/10.1109/EMBC44109.2020.9176200

Amrane, M., Oukid, S., Gagaoua, I., & Ensarİ, T. (2018). Breast cancer classification using machine learning. *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, 1–4. https://doi.org/10.1109/EBBT.2018.8391453

Ara, S., Das, A., & Dey, A. (2021). Malignant and benign breast cancer classification using machine learning algorithms. *2021 International Conference on Artificial Intelligence (ICAI)*, 97–101. https://doi.org/10.1109/ICAI52203.2021.9445249

Bai, X.-l., & Qian, X. (2008). Medical image classification based on fuzzy support vector machines. *2008 International Conference on Intelligent Computation Technology and Automation (ICICTA)*, *2*, 145–149. https://doi.org/10.1109/ICICTA.2008.457

Buda, M., Saha, A., Walsh, R., Ghate, S., Li, N., Święcicki, A., Lo, J., & Mazurowski, M. (2020). *Detection of masses and architectural distortions in digital breast tomosynthesis: A publicly available dataset of 5,060 patients and a deep learning model*.

Bunz, F., & Vogelstein, B. (2022). Bases genéticas del cáncer. In J. Loscalzo, A. Fauci, D. Kasper, S. Hauser, D. Longo, & J. L. Jameson (Eds.), *Harrison. principios de medicina interna, 21e*. McGraw-Hill Education. accessmedicina.mhmedical.com/content.aspx?aid=1191732364

Deep Deb, S., Rahman, M. A., & Jha, R. K. (2020). Breast cancer detection and classification using global pooling. *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–5. https://doi.org/10.1109/ICCCNT49239.2020.9225375

Gowri, D. S., & Amudha, T. (2014). A review on mammogram image enhancement techniques for breast cancer detection. *2014 International Conference on Intelligent Computing Applications*, 47–51. https://doi.org/10.1109/ICICA.2014.19

Gui, G., Huang, R., Wu, Z., Chen, Y., Xie, Y., & Qin, W. (2019). A comparison of breast tumors classifications based on mammography and digital breast tomosynthesis: A fractal analysis study. *2019 International Conference on Medical Imaging Physics and Engineering (ICMIPE)*, 1–8. https://doi.org/10.1109/ICMIPE47306.2019.9098203

Houssami, N., Hunter, K., & Zackrisson, S. (2017). Overview of tomosynthesis (3D mammography) for breast cancer screening. *BREAST CANCER MANAGEMENT*, *6*(1), 179–186. https://doi.org/{10.2217/bmt-2016-0024}

Joaquín José Mosquera Osés, Á. I. L., Jose Ramón Varela Romero. (2012). Tomosíntesis. Un avance cualitativo en el diagnóstico de patología mamaria. *Revista de Senología y Patología Mamaria.*, *25*(4), 152–156. https://doi.org/10.1016/j.senol.2012.10.002

K. Yuvaraj, U. S. R. (2022). Hybrid active contour mammographic mass segmentation and classification. *Computer Systems Science and Engineering*, *40*(3), 823–834. https://doi.org/10.32604/csse.2022.018837

Kathale, P., & Thorat, S. (2020). Breast cancer detection and classification. *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, 1–5. https://doi.org/10.1109/ic-ETITE47903.2020.367

Mohamed, B. A., & Salem, N. M. (2018). Automatic classification of masses from digital mammograms. *2018 35th National Radio Science Conference (NRSC)*, 495–502. https://doi.org/10.1109/NRSC.2018.8354408

Pertuz, S., Torres, G. F., Tamimi, R., & Kamarainen, J. (2019). Open framework for mammography-based breast cancer risk assessment. *2019 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, 1–4. https://doi.org/10.1109/BHI.2019.8834599

Ricciardi, R., Mettivier, G., Staffa, M., Sarno, A., Acampora, G., Minelli, S., Santoro, A., Antignani, E., Orientale, A., Pilotti, I., Santangelo, V., D'Andria, P., & Russo, P. (2021). A deep learning classifier for digital breast tomosynthesis. *Physica Medica*, *83*, 184–193. https://doi.org/10.1016/j.ejmp.2021.03.021

Rocha García, A., & Mera Fernández, D. (2019). Tomosíntesis de la mama: Estado actual. *Radiología*, *61*(4), 274–285. https://doi.org/https://doi.org/10.1016/j.rx.2019.01.002

Sammut, C., & Webb, G. I. (Eds.). (2010). Leave-one-out cross-validation. In *Encyclopedia of machine learning* (pp. 600–601). Springer US. https://doi.org/10.1007/978-0-387-30164-8_469

Smith, A. (2012). The use of breast tomosynthesis in a clinical setting.

Sun, L., Wen, J., Wang, J., Zhao, Y., & Xu, Y. (2020). Classification of mammography based on semi-supervised learning. *2020 IEEE International Conference on Progress in Informatics and Computing (PIC)*, 104–111. https://doi.org/10.1109/PIC50277.2020.9350835

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/https://doi.org/10.3322/caac.21660

Surhone, L., Timpledon, M., & Marseken, S. (2022). *Ncss statistical software documentation | ncss software help*. NCSS. https://www.ncss.com/software/ncss/ncss-documentation/#Regression

Thawkar, S., & Ingolikar, R. (2017). Automatic detection and classification of masses in digital mammograms. *International Journal of Intelligent Engineering and Systems*, *10*, 65–74. https://doi.org/10.22266/ijies2017.0228.08

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). Chapter 7 - extending instance-based and linear models. In I. H. Witten, E. Frank, M. A. Hall, & C. J. Pal (Eds.), *Data mining (fourth edition)* (Fourth Edition, pp. 243–284). Morgan Kaufmann. https://doi.org/https://doi.org/10.1016/B978-0-12-804291-5.00007-6

Yu, Z., & Bajaj, C. (2004). A fast and adaptive method for image contrast enhancement. *2004 International Conference on Image Processing, 2004. ICIP '04.*, *2*, 1001–1004 Vol.2. https://doi.org/10.1109/ICIP.2004.1419470