



Segmentation of Masses on Mammograms Using Data Augmentation and Deep Learning

Felipe André Zeiser¹ · Cristiano André da Costa¹ · Tiago Zonta^{1,2} · Nuno M. C. Marques³ · Adriana Vial Roehe⁴ · Marcelo Moreno⁵ · Rodrigo da Rosa Righi¹

Published online: 23 March 2020

© Society for Imaging Informatics in Medicine 2020

Abstract

The diagnosis of breast cancer in early stage is essential for successful treatment. Detection can be performed in several ways, the most common being through mammograms. The projections acquired by this type of examination are directly affected by the composition of the breast, which density can be similar to the suspicious masses, being a challenge the identification of malignant lesions. In this article, we propose a computer-aided detection (CAD) system to aid in the diagnosis of masses in digitized mammograms using a model based in the U-Net, allowing specialists to monitor the lesion over time. Unlike most of the studies, we propose the use of an entire base of digitized mammograms using normal, benign, and malignant cases. Our research is divided into four stages: (1) pre-processing, with the removal of irrelevant information, enhancement of the contrast of 7989 images of the Digital Database for Screening Mammography (DDSM), and obtaining regions of interest. (2) Data augmentation, with horizontal mirroring, zooming, and resizing of images; (3) training, with tests of six-based U-Net models, with different characteristics; (4) testing, evaluating four metrics, accuracy, sensitivity, specificity, and Dice Index. The tested models obtained different results regarding the assessed parameters. The best model achieved a sensitivity of 92.32%, specificity of 80.47%, accuracy of 85.95% Dice Index of 79.39%, and AUC of 86.40%. Even using a full base without case selection bias, the results obtained demonstrate that the use of a complete database can provide knowledge to the CAD expert.

Keywords Breast cancer · Segmentation · Computer-aided detection · Fully convolutional network · Deep learning · U-Net

Introduction

Breast cancer is the most deadly disease in women, accounting for more than half a million deaths per year [3]. It is the second most incident type of cancer in the world [3]. Even in

the face of advances in treatment, early diagnosis is a crucial point for defining the patient outcome.

Currently, the primary technique used in screening programs for breast cancer is the mammogram [8]. Traditionally, two projections per breast are obtained, one

✉ Cristiano André da Costa
cac@unisinos.br

Felipe André Zeiser
felipezeiser@edu.unisinos.br

Tiago Zonta
tiago.zonta@unoesc.edu.br

Nuno M. C. Marques
nmm@fct.unl.pt

Adriana Vial Roehe
adrianar@ufcspa.edu.br

Marcelo Moreno
marcelo.moreno@uffs.edu.br

Rodrigo da Rosa Righi
rrighi@unisinos.br

¹ Software Innovation Laboratory – SOFTWARELAB, Applied Computing Graduate Program, Universidade do Vale do Rio dos Sinos – Unisinos, Av. Unisinos 950, São Leopoldo 93022-000, Brazil

² Ciências da Vida em Pesquisa, Universidade do Oeste de Santa Catarina, Chapecó, Brazil

³ Departamento de Informática, Universidade Nova de Lisboa, Almada, Portugal

⁴ Departamento de Patologia e Medicina Legal, Universidade Federal de Ciências da Saúde de, Porto Alegre, Brazil

⁵ Estudos Biológicos e Clínicos em Patologias Humanas, Universidade Federal da Fronteira Sul, Chapecó, Brazil

craniocaudal (CC) and another mediolateral oblique (MLO). Nonetheless, mass projection on a mammogram is directly affected by its shape and location, the similarity of density between healthy and disordered tissues, and by the technical and instrumental quality [4]. Besides, the mammogram analysis requires the professionals' time and a high degree of attention to analyzing many cases [17].

These factors motivated, especially in the last decade, the development of computational systems to aid in the interpretation of medical images, known as computer-aided detection (CAD). The use of these dual reading systems directly contributes to the increased sensitivity of the disease detection [9]. The current literature proposes the use of CAD systems, with the application of support vector machines (SVM) and micro-genetic algorithm (uGA) for the classification of masses and non-masses [16]. Another alternative is the use of a cascade of Convolutional Neural Network (CNN) to perform the classification in regions of mass or not [8]. As in Al-masni et al. [2] and Li et al. [12], CNN sets are used for segmentation and classification, respectively, in benign or malignant. Nevertheless, most of the time, the authors make a selection of images in the methodologies, however not generally detailing the process used.

The main contribution of this article is a new CAD system methodology to identify and segment mass in mammograms. The method allows the monitoring of the lesion over time, its format, and its characteristics. We apply the contrast enhancement and noise removal on the scanned mammograms. In this process, we evaluate the use of data augmentation and the selection of non-overlapping regions of interest (ROI) based on the identification of masses in the projections. We employed the U-Net model with some improvements to avoid overfitting, given the low amount of images available. Another contribution is the complete use of Digital Database for Screening Mammography (DDSM), both normal, benign, and malignant cases. Related work uses only a portion of the publicly available databases and sometimes does not clarify how the cases were selected. These facts directly impact on the number of cases analyzed. Our method allows future work to make a direct comparison with the proposed method, without manual selection or pseudo-random selection bias.

The article is divided into five main sections. “[Related Work](#)” presents the related work. “[Materials and Methods](#)” describes the materials and methods. In “[Results and Discussion](#),” we present the results and discussion. Finally, “[Conclusion and Future Work](#)” highlights the main conclusions and directions for future work.

Related Work

CAD systems based on aid in the detection of mammograms masses are designed to increase the performance of specialists

by serving as double-reading systems [8] and contributing to the reduction of the number of false positive or false negative results [17]. There are numerous methods of mass segmentation in mammograms. A summary of the most relevant methods is presented in Table 4. The evaluation metrics presented are the most frequently used in the literature [14]. Despite the available data, it is considered an unresolved problem [6], mainly due to the small number of images used in the studies, mass variability, and computational limitations.

In this way, obtaining a consistent dataset and labeled by specialists in the medical field is one of the main challenges in the development of a CAD. In the analysis of the literature has identified the use of digitized mammograms databases: DDSM,¹ INBreast,² Breast Cancer Digital Repository (BCDR),³ Mammographic Image Analysis Society (MIAS),⁴ and private image datasets. However, the amount of images provided by the bases is still insufficient for the generalization of the problem, due to the biological variability and size of the masses [1]. To solve these problems, the literature proposes the use of data augmentation and transfer learning techniques, which indicate an improvement in the performance of deep learning models [2, 18]. In this context, the data augmentation technique is efficient mainly for the acceleration of convergence and to avoid overfitting of the models [1].

One of the main advantages of using deep learning techniques when compared to manual resource extraction techniques is the ability to learn a set of high-level attributes and provide high accuracy even in complex problems [1]. Correspondingly, Dhungel et al. [7] proposed a cascade of deep learning methods with a Deep Belief Network to select suspicious regions, a CNN to keep the correct candidates, and a random forest (RF) to reduce false positives. In 2017, Dhungel et al. [6] developed a complete method with detection, segmentation, and classification of the masses, using the masses found in the previous work to segment them employing deep structured learning followed by a conditional random field (CRF). At the end of the processing, the authors performed a classification of the regions in benign and malignant using a CNN.

With the use of data augmentation, Li et al. [12] were able to improve the results for the classification of mammograms into benign and malignant. For the classification, the authors used a CNN adapted from the DenseNet model. The authors employed data augmentation techniques, applied rotations of 90°, 180°, and 270°, vertical and horizontal mirroring, and a scalar reduction of 80%. In the end, the method achieved an accuracy of 94.55%.

¹ <http://www.eng.usf.edu/cvprg/Mammography/Database.html>

² http://medicalresearch.inescporto.pt/breastresearch/index.php/Get_INbreast_Database

³ <https://bcdr.eu/information/about>

⁴ <https://www.repository.cam.ac.uk/handle/1810/250394>

In the same way, Al-masni et al. [2] developed a simultaneous detection and classification CAD system based on the You Only Look Once (YOLO) architecture. The YOLO is used for locating potential masses and classifying them as benign and malignant. One of the differentials presented by the authors is the ability to detect masses in the pectoralis muscle or in the vicinity of dense breast tissues, considered challenging regions for detection. In the work of Al-antari et al. [1], the YOLO architecture is used only in the detection of masses, processing the ROI obtained in the previous step in a Full Resolution Convolutional Network (FrCN). This network consists of a decoder-encoder network, with the removal of the max pool and subsampling layers, preserving the spatial resolution of the input image. The modification in the structure of FrCN aims to avoid the loss of any information during the generation of resource maps. The output of the resource maps is fed into a softmax classifier to obtain the probability that each pixel is a mass. Finally, the classification of the masses into benign and malignant is performed in a simplified version of AlexNet.

Methodologies using the extraction of manual resources present good results in certain studies. In de Nazaré Silva et al. [13], a contrast enhancement method with a low-pass filter and the use of a Wavelet transform is proposed. In their work, mammograms are segmented using the technique known as Quality Threshold. The classification is performed using SVM, based on resources extracted with the Haralick descriptor and a correlogram function.

In Sampaio et al. [16], the authors suggest a method adapted to density for the identification of the masses, dividing the breasts into two groups: dense and not dense. For this purpose, a variance filter is applied to the mammograms of the DDSM database, extracting internal and mean-variance characteristics. These are then grouped into vectors optimized by a uGA and processed in a SVM, in charge of the classification of breast density. In the next step, they perform the segmentation of suspicious regions using an uGA to create a texture proximity mask. The resulting ROI have their textures analyzed by combining clustering density-based spatial clustering of applications with noise (DBSCAN), phylogenetic trees, local binary standards, and SVM. Finally, an uGA was used to select the ROI that generate the best model and that maximize the classification of masses and non-masses used in the SVM.

Another similar approach was proposed by Diniz et al. [8], classifying the breasts in the same subgroups of Sampaio et al. [16], but with the use of a CNN architecture. After this classification, the segmentation of asymmetric regions between the two breasts was performed using similarity indexes. These asymmetric regions were processed in another CNN for the classification in masses and not masses. A method of detection and classification of suspicious areas is proposed by Chakraborty et al. [5] with the use of an iterative approach of

region growth by a threshold controlled based on the mean value of intensity and with values obtained in the analyzed region itself. After identifying possible candidates, they performed a false positive reduction and extracted characteristics to be classified with a random forest.

Materials and Methods

This work follows the methodology presented in Fig. 1. The employed method can be divided into four main parts: pre-processing, data augmentation, training, and testing. In pre-processing, we perform the removal of irrelevant information, contrast enhancement, and performed the resizing of DDSM images. To perform the data augmentation, we applied a horizontal mirroring and zoomed in the images, extracting a set of ROI. Finally, we performed the training and tests under six U-Net adapted models.

For training, validation, and testing, we used a computer with an Intel Core i7-9700K processor, with 32 GB of RAM, and NVIDIA Titan XP graphic card with 12 GB of memory.

Materials Description

We chose to use the DDSM because it is the most widely used dataset in the literature [14]. It has approximately 2500 cases subdivided into three significant volumes, normal, benign, and malignant, collected at Massachusetts General Hospital, Wake Forest University, and the University of Washington at St. Louis Medical School [11].

Each case has two projections for each breast (CC and MLO), as well as patient information (study date, breast density, age, number of lesions, type of path, among others) and the image (scan date, image extension, file name, pixels per line, bits per pixel, among others). At this point, we had the collaboration of two specialists who followed the research and made some notes. In the first analysis of the available images, for example, pointed out possible difficulties for the segmentation of the masses, due to the higher density of the mammary parenchyma in young patients. Moreover, the specialists reviewed the ground truth of each case before the use in the methodology.

Data Pre-processing

We selected all cases of the DDSM base (normal, benign, and malignant), except cases with calcifications, since our objective is the detection and segmentation of masses. When the specialist detects a mass only in one of the projections, either CC or MLO, we eliminated the projection without a mass to avoid noise in the dataset. By using normal cases in our methodology, we can simulate a real environment for breast cancer diagnosis. And also, because of the simple use of cases with

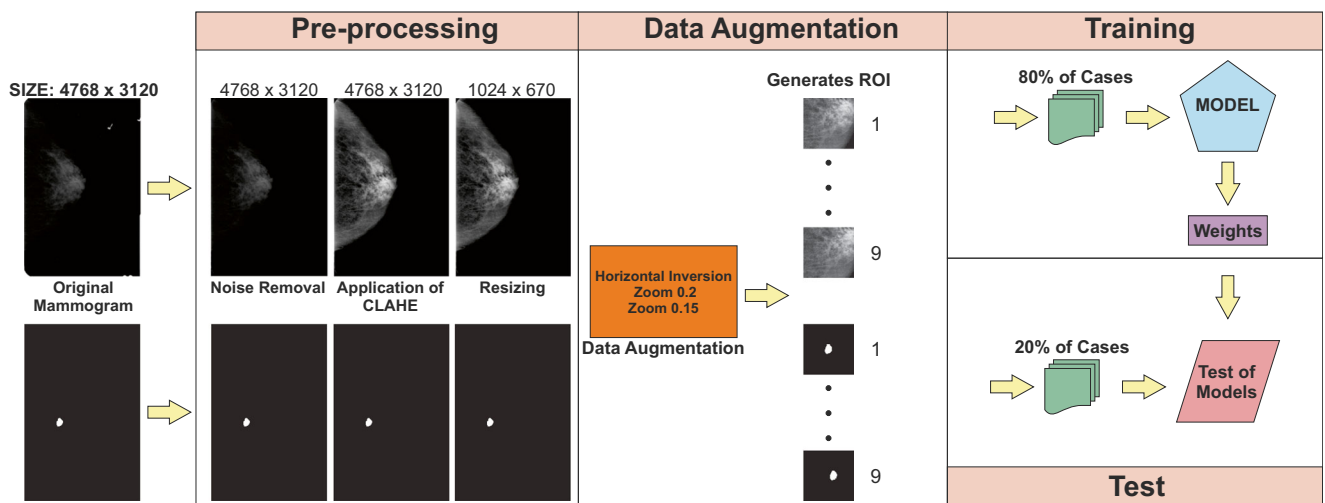


Fig. 1 The proposed diagram for the CAD system methodology, based on a U-Net model for mass detection and segmentation. From left to right: original mammogram, pre-processing phase (noise removal, application

of CLAHE, and resizing), data augmentation (horizontal inversion with zoom 0.2/0.15 and generated ROI), training (80% of cases), and test (20% of cases)

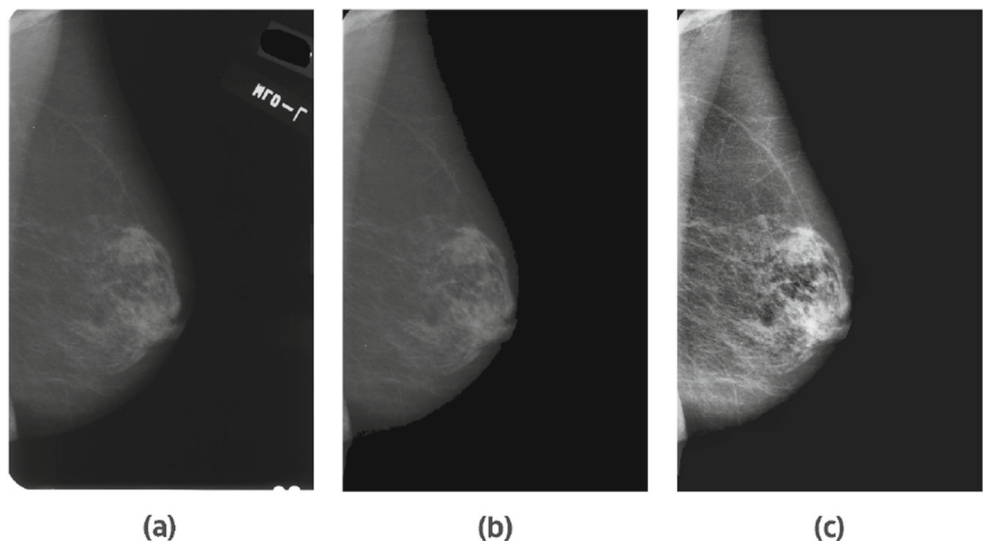
mass presence, the results can be overestimated. Our methodology for using DDSM base images is perfectly replicable using only the information provided by the base. Also, for the normal cases, where ground truth was not present, only the characteristics of the case, we managed a black image for these cases. After the selection, the images were resized to a height of 1024 pixels, with a reduction of the width proportional to the height. The reduction was necessary because the images of the DDSM did not have a standard size. The height of 1024 allows a reduction in the computational processing time and does not significantly affect the results [8, 16].

To remove information irrelevant to the identification of masses, we performed segmentation of mammograms using the proposed algorithm of Felzenszwalb and Huttenlocher [10] (Fig. 2b), which is based on the representation of the image using an undirected graph $G = (V,$

$A)$, with vertices $(v_i, v_j \in A)$ to be segmented and the edges $(v_i, v_j \in A)$ corresponding to the pair of neighboring vertices. Each edge $(v_i, v_j) \in A$ has a non-negative weight $w(v_i, v_j)$ associated, being the measure of dissimilarities between the two neighboring elements v_i e v_j . In images, the elements of V are the pixels, and the weight w of an edge A is a measure of dissimilarity, such as the difference in color or texture of the image. For the segmentation, the algorithm adopts a greedy process, joining vertices in the same connected component with low weights and joining edges that connect vertices of different connected components with larger weights [10].

Moreover, given the tenuous differences between healthy tissues and mammogram patients, the contrast enhancement technique is used in the literature for reducing this problem [8, 16]; this concern was also pointed out by the specialist

Fig. 2 Mammogram of case 1329: **a** right breast MLO original projection; **b** removing the noise with the targeting **c** contrast enhancement by application of CLAHE



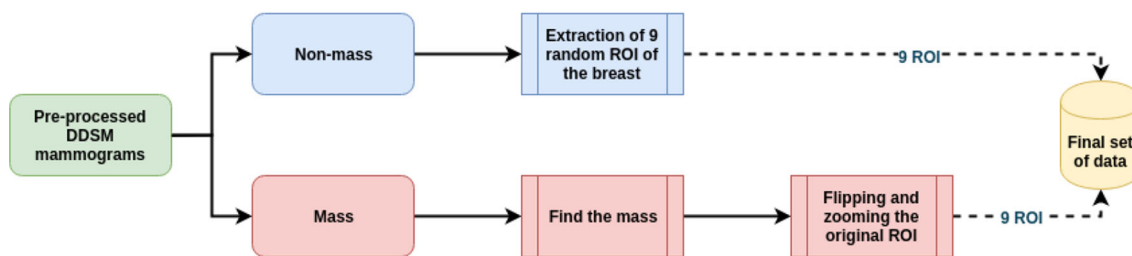


Fig. 3 Flow of data augmentation and extraction of ROI for cases with mass and non-mass

researchers. To deal with these differences, we applied the contrast-limited adaptive histogram equalization (CLAHE) (Fig. 2c), which subdivides the image into sub-areas using interpolation between the edges, and to avoid increasing noise, we use a gray-level threshold, redistributing the pixels above this threshold in the image. The CLAHE can be defined by:

$$p = [p_{\max} - p_{\min}] * G(f) + p_{\min} \quad (1)$$

where p is the new gray-level value of the pixel, the values p_{\max} and p_{\min} are the pixels with the highest and lowest values in the neighborhood and $G(f)$ corresponds to the cumulative distribution function [19].

Data Augmentation

First, we randomly divided the cases into three sets: training (70%), validation (10%), and testing (20%) [2]. The masses, in most cases, are proportionally smaller regions of the breast tissue, making the training data set unbalanced, complicating the training process. Therefore, to take advantage of the examples available in DDSM and increase the accuracy of the model, we use techniques of data augmentation, such as horizontal inversion and image zoom, extracting nine regions of interest (ROI) of 256×256 pixels per mammogram image. The masses were located based on the ground truth of the DDSM base and endorsed by a radiologist, so a window of 256×256 was positioned to ensure that the complete frame of the mass. In cases where the mass has larger dimensions, as many windows have been used as necessary to frame the mass and ensure that none overlap another. For the cases with no mass present, we randomly extracted 9 ROI from the images. Figure 3 shows the process flow. This process also ensures that the data is balanced.

After this processing, we obtained two data sets, one without data augmentation and another one based on the methodology described in “Data Augmentation,” generating 48,659 ROI. Table 1 shows the number of samples used in each of the sets.

Experimental Model Parameter Settings

In this work, we use models based on the U-Net architecture [15] (Fig. 4), developed to provide better segmentation in medical images. Its name is due to the symmetrical shape of the

model, which can be divided into two parts: downsampling and upsampling. Downsampling extracted features with two 3×3 convolutions each with a ReLU, and gradually reduces the spatial size of the image with a 2×2 max pooling operation, doubling the number of feature maps at each layer. While in upsampling, the image is reconstructed with a 2×2 deconvolution and the concatenation of layer maps of the same level of downsampling and two 3×3 convolutions each followed by a ReLU. In the end, a 1×1 convolutional layer is used to map each vector feature to the number of classes in the dataset. Another essential element is the depth of the network, which initially has four levels plus the base of the model. Its main advantage is the combination of contextual information, allowing the use of images with different dimensions and demonstrates efficiency in the use of the data augmentation [15]. We tested six models derived from U-Net with the characteristics presented in Table 2. We added the original architecture with a padding in each convolution, avoiding the loss of information during the feature extraction process. In addition, at each level after the first convolution, we add a dropout of 0.3 to decrease overfitting possibilities.

We use some assumptions for all models, being sixteen feature maps on the first layer, one-pixel padding on each layer, batch size of eight, a learning rate of $1e-5$, Adam for weight optimization, Dice Index as loss function, and 140 epochs.

Results and Discussion

In this section, we present the generated models in the training and validation phase, the results for the evaluation metrics for each model, and two test cases.

Table 1 Number of samples of the dataset

Set	Type	Images	ROI
Training	Mass	2928	17,881
	Non-mass	2664	16,278
Validation	Mass	418	2552
	Non-mass	381	2321
Test	Mass	844	5087
	Non-mass	754	4549

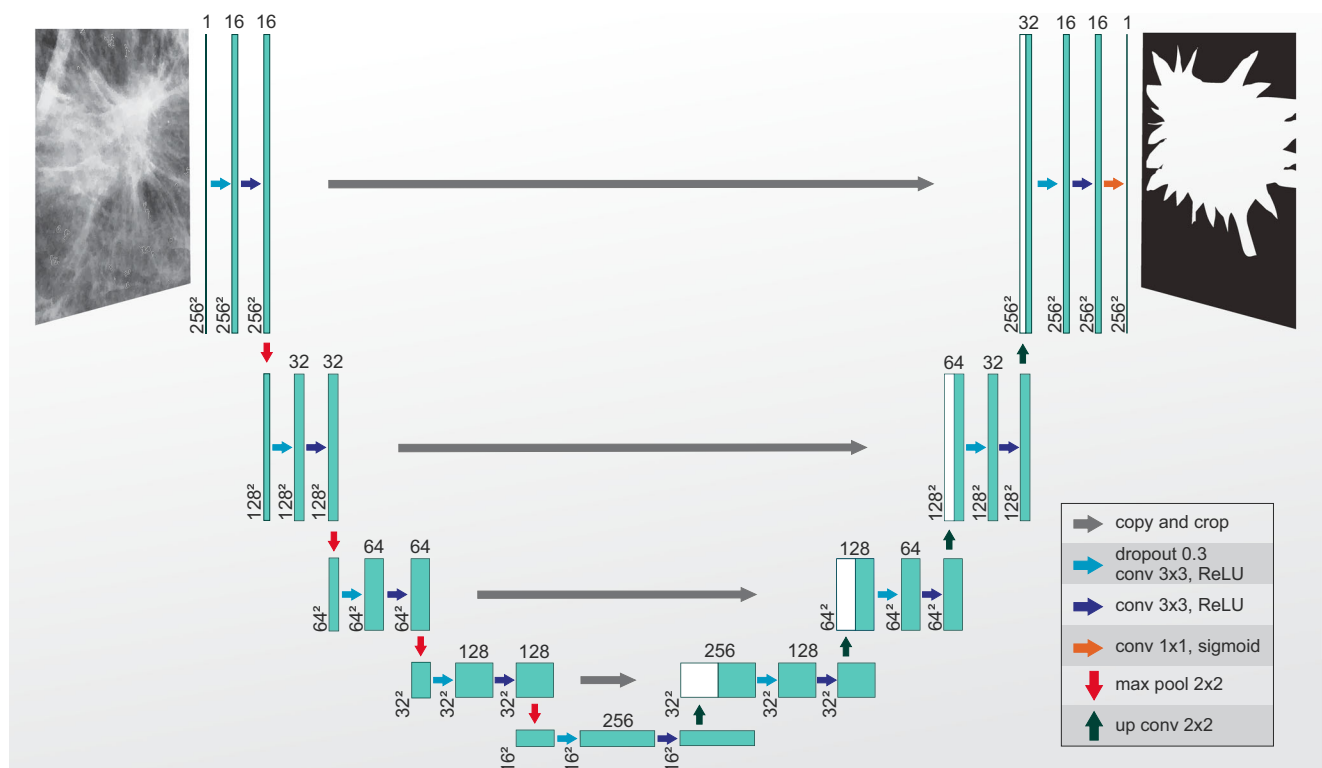


Fig. 4 Proposal of a deep learning model for the segmentation of the masses, based on U-Net architecture

Training and Validation of the Models

To monitor the evolution of the models, we use the 1–Dice Index; this index, as mentioned previously, can evaluate how much the images are similar to each other. In Fig. 5 are the indices obtained for each of the 140 times of the models tested for the CAD system definition.

For model A (Fig. 5), we can see that the network remained relatively stable, not showing a significant evolution over the 140 epochs. For models B and C, both with greater complexity of feature maps, they show good variation approximately until the epoch 40, after both are overfitting. On the other hand, model D tends to overfit, but without being able to present any learning characteristics. While for the models E and F, in the first one, we have a more unstable network, if we take into account the validation sub-set, and in the second a

more constant learning characteristic, but without overfitting characteristics, as in previous models.

Test of Models

The results obtained in the evaluation were founded on the subsets of tests for the best time of each model of the training phase. The choice was made automatically, based on the indexes of loss and accuracy. Table 3 presents the performance of the model. Figure 6 shows the ROC curves and AUC values. Here, all quantitative measures for mass segmentation were calculated per pixel of segmented maps with the same resolution as the original image (mammographic image input or ROI).

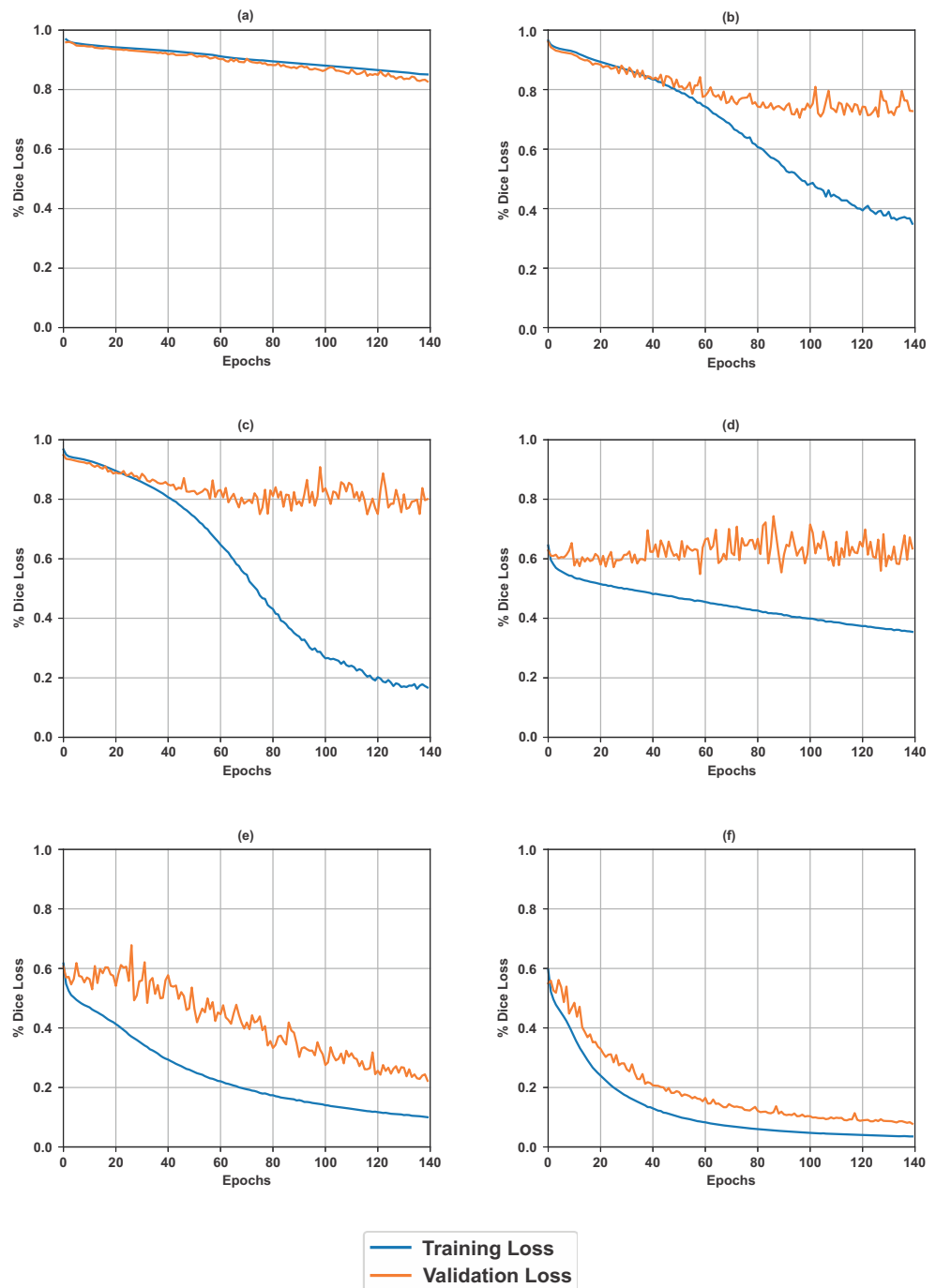
In Table 3, we can observe the metrics obtained by the models. Model A shows high sensitivity. However, not always a high sensitivity may also reflect a high specificity because of the prevalence of FP cases. This behavior can also be seen in the performance of the ROC curve (Fig. 6) and many times related work focus solemnly on one index, such as sensitivity or accuracy, which can lead to a high rate of unnecessary biopsies or the exclusion of malign masses. Despite that, this article considers five evaluation metrics as presented in Table 3.

All the models tested present a high level of sensitivity, which reinforces the correct detection of mass regions. Despite that, the models A to D did not present enough specificity and were not useful in the classification of non-massed cases. The best results were obtained in experiment F. This

Table 2 Tested configurations

Model	Depth	Augmentation	Time per epoch
A	3	No	≈ 120 s
B	4	No	≈ 165 s
C	5	No	≈ 240 s
D	3	Yes	≈ 315 s
E	4	Yes	≈ 390 s
F	5	Yes	≈ 640 s

Fig. 5 Plots of training and validation dice loss for each proposed model: **a** depth 3, **b** depth 4, **c** depth 5, **d** depth 3 and augmentation, **e** depth 4 and augmentation, and **f** depth 5 and augmentation



model employed a 5-depth network with data augmentation and got the best results in term of accuracy (85.95%), Dice Index (79.39%), and AUC (86.40%). Regarding sensibility (92.32%), although the obtained results were worse than model A, the much higher specificity (80.47%) shows that the number of feature maps was enough to differentiate among masses or non-masses. Another conclusion is that there is a tendency to approximate the evaluation metrics as we increase the depth of the network and also with the use of data augmentation. The improvement of the evaluation metrics is

mainly due to the similarity between diseased tissues and healthy, which requires a mathematical complexity and a larger data sample.

Comparison with Related Work

A direct comparison with related work is difficult, considering that they used only part of the dataset cases, selecting specific aspects or removing certain images. However, even with the use of a complete base, the methodology employed in this

Table 3 Evaluation metric results in test sets, showing the best results in italics. (*Sen* is the sensitivity, *Spec* is the specificity, *Acc* is the accuracy, and *AUC* is the area under the ROC curve)

Model	Sen %	Spec %	Acc %	Dice Index %	AUC %
A	96.69	14.09	29.73	10.90	55.39
B	75.20	52.60	56.88	19.54	63.90
C	73.41	72.90	70.26	28.43	73.16
D	84.27	66.97	74.97	39.98	75.62
E	86.15	84.38	85.20	61.16	85.26
F	92.32	80.47	85.95	79.39	86.40

article was able to obtain comparable results with the related works, as can be seen in Table 4. Sensitivity, which represents the detection capacity of the disease, is even higher compared to the work of [5, 8, 13, 16]. Regarding the specificity index, we obtained better results than the work of [6].

As for the accuracy, and the AUC, we can conclude that the use of a larger amount of data, without the bias of the choice of data, can directly affect the results. The lower value of the Dice Index compared to literature results may have been influenced by the use of digitized mammograms and not digital mammograms. However, the results are close to those presented in the analyzed literature and can transmit to the professional greater reliability of the CAD system by the number of cases investigated. Therefore, the proposed methodology, based on results and related work, indicates the possibility of use in the health area, as a screening test. In this way, this model can be used as a first step, eliminating the normal results and indicating a possible diagnose of abnormal exams for the CAD expert analysis.

Differently of some related works (as in [1]), we opted for not excluding FP cases. In our view, this exclusion can bias

Table 4 Comparison of the methodology with the related works. (*Type* are the case include in the study, *Img* is the sample size, *Sen* is the sensitivity, *Spec* is the specificity, and *Acc* is the Accuracy)

Article	Base	Type	Img	Sen %	Spec %	Acc %	Dice Index %	AUC%
[8]	DDSM/non-dense	Mass	1004	91.56	90.73	91.04	–	–
	DDSM/dense		1482	90.36	96.35	94.84	–	–
[2]	DDSM	Mass	600	100.0	94.00	99.70	–	96.45
[7]	DDSM-BCRP and INbreast	Mass	274	–	–	–	90.00	–
	INbreast		107	97.14	92.41	95.64	–	94.78
[1]	INbreast	Mass	410	98.00	70.00	90.00	–	91.00
[6]	INbreast	Mass/non-mass	1049	92.99	–	–	–	–
[16]	DDSM/non-dense	Mass	678	83.70	–	–	–	–
[12]	DDSM/dense private	Mass	2042	95.60	95.36	94.55	–	91.20
[5]	DDSM	Mass/non-mass	1300	85.00	–	77.89	–	–
[13]	DDSM	Mass	599	92.31	82.20	83.53	–	–
Proposed method	DDSM	Mass/non-mass	7989	92.32	80.47	85.95	79.39	86.40

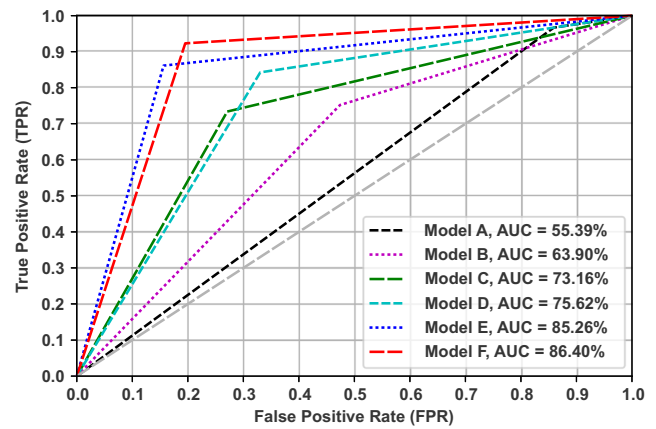


Fig. 6 The performance of mass segmentation in terms of ROC curves in test sets

the process, impacting the system’s final metric values. Using U-Net reduced the cost and time to perform mammogram segmentation in the test set compared to using manual resource extraction [13, 16]. Furthermore, with the use of higher ROI, we were able to capture information from the whole region of the lesion, in spite of using small areas, as in the work of Diniz et al. [8].

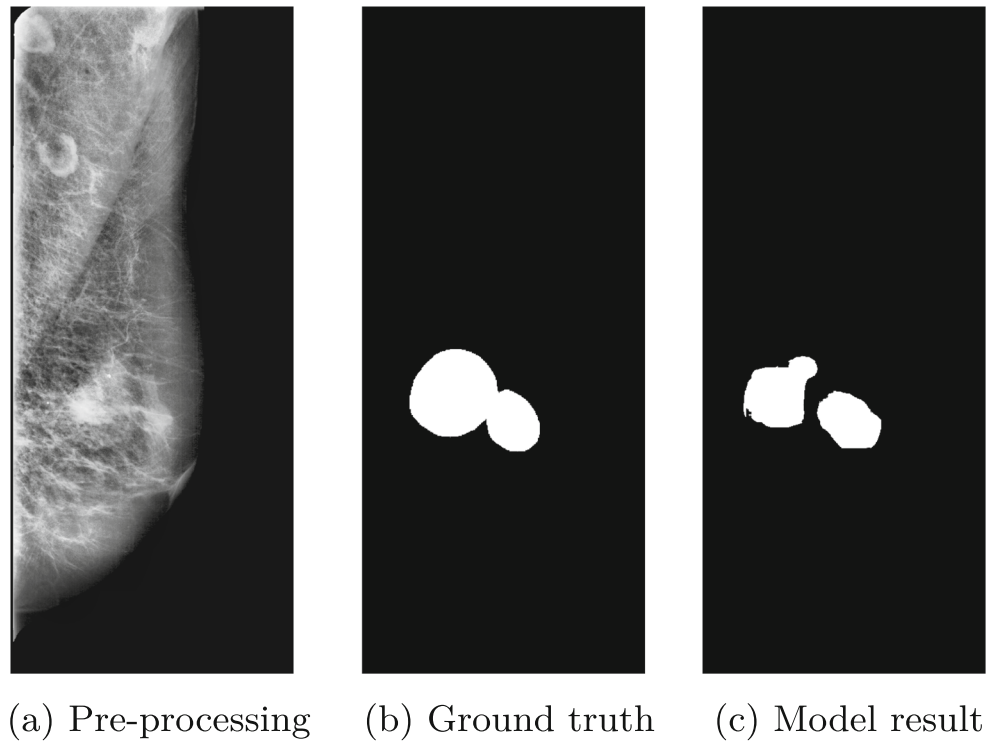
Case Studies

In this section, specific cases will be shown to demonstrate the application of the methodology in DDSM images.

True Positive Test Cases

The first example is patient A_1529_1. In this case, the methodology was able to detect the mass. Figure 7 a shows the

Fig. 7 Case study of patient mammography 1529, right MLO projection. **a** Pre-processing. **b** Ground truth. **c** Model result

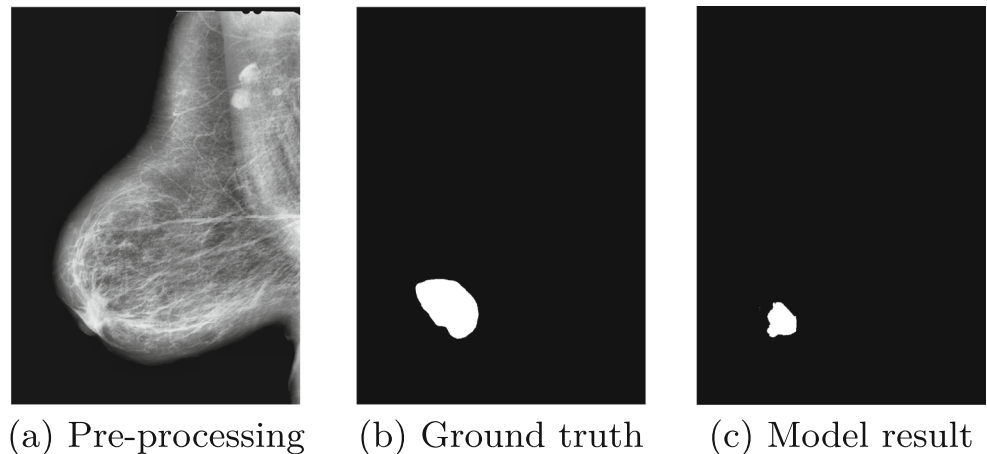


results after the pre-processing steps. Figure 7 b and c are demarcation of the specialist and the one produced by the model, respectively.

The second example is from patient A_1011_1 (Fig. 8). The mass detected by the proposed methodology is smaller than that indicated on the label provided by DDSM. However, the location of the model lesion is consistent with that of the specialist.

Finally, the third case of patient B_3048_1 (Fig. 9) is from a breast with a higher density compared to the previous examples. However, the model is able to detect the location of the lesion.

Fig. 8 Case study of patient mammography 1011, left MLO projection. **a** Pre-processing. **b** Ground truth. **c** Model result



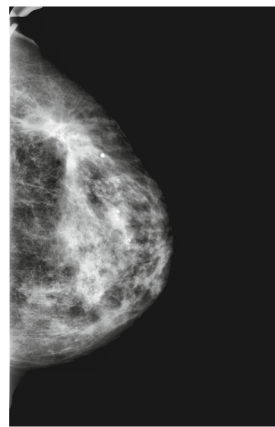
False Negative Test Case

In the case A_1617_1 (Fig. 10), the model was not able to detect and perform mass segmentation. In this case, we can observe that the mass has a density very similar to the rest of the breast tissue, making it difficult to diagnose for the model.

Conclusion and Future Work

This work presented a CAD system proposal to aid in the detection and segmentation of digitized mammograms.

Fig. 9 Case study of patient mammography 3048, left CC projection. **a** Pre-processing. **b** Ground truth. **c** Model result



(a) Pre-processing



(b) Ground truth



(c) Model result

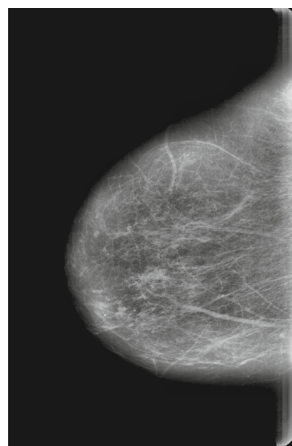
Initially, we analyzed the available public mammogram datasets and chose the DDSM, with the largest set of cases being 2620. Due to the characteristics of DDSM, we used some pre-processing techniques, eliminating patient identification platelets, noises present in the opposite region of the breast, and applying a contrast enhancement. Finally, we perform a data augmentation and process it in models derived from U-Net. Different from the related work, we used the three sets of DDSM images (normal, benign, and malignant). The full use of the DDSM database allows a more direct comparison of the results in future studies. Even using the complete base, our results are similar to the obtained in the considered related work, outperforming the results regarding the sensitivity metric. With a more significant number of cases analyzed, the proposed CAD system can provide greater reliability and objectivity in the identification of cases in real environments.

As a benefit to professionals and clinics, we envision its use as a dual reading system, serving as a second opinion during

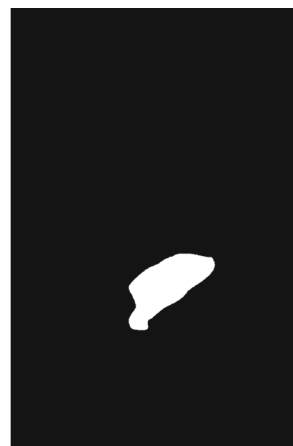
case analysis. As a dual reading device, we aim to assist in decreasing misdiagnosis. Using the method proposed in this article, the professional can still monitor over time whether or not suspicious lesions have grown. In this way, we are incorporating the proposed method into the clinical process, as a screening test, with a partner hospital.

According to the images produced by the models, we perceive that the network tends to classify sets of pixels with a higher intensity as masses. This behavior is desired, but in dense breasts, usually of young women, the network ends up producing false positives. Therefore, a methodology adapted to breast density may contribute to the improvement of classification accuracy. Furthermore, the inclusion of clinical data in the training process, such as age, family history, smoking, weight, and use of hormone replacement, is an aspect to be considered in future models. Another future work would be the use of more data to improve the accuracy of the model.

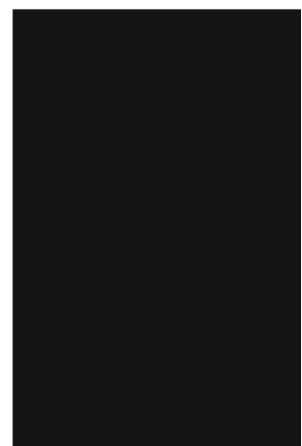
Fig. 10 Case study of patient mammography 1617, left CC projection. **a** Pre-processing. **b** Ground truth. **c** Model result



(a) Pre-processing



(b) Ground truth



(c) Model result

References

- Al-antari MA et al.: A fully integrated computer-aided diagnosis system for digital x-ray mammograms via deep learning detection, segmentation, and classification. *International Journal of Medical Informatics* 117:44–54, 2018
- Al-masni MA et al.: Simultaneous detection and classification of breast masses in digital mammograms via a deep learning yolobased cad system. *Computer Methods and Programs in Biomedicine* 157:85–94, 2018
- Bray F et al.: Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 2018. <https://doi.org/10.3322/caac.21492>
- Caldas FAA et al.: Controle de qualidade e artefatos em mamografia. *Revista Radiologia Brasileira* 38(4):295–300, 2005
- Chakraborty J, Midya A, Rabidas R: Computer-aided detection and diagnosis of mammographic masses using multi-resolution analysis of oriented tissue patterns. *Expert Systems with Applications* 99: 168–179, 2018. <https://doi.org/10.1016/j.eswa.2018.01.010>
- Dhungel N, Carneiro G, Bradley AP: A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Medical Image Analysis* 37:114–128, 2017
- Dhungel N et al.: Deep learning and structured prediction for the segmentation of mass in mammograms. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, 2015, pp. 605–612
- Diniz B et al.: Detection of mass regions in mammograms by bilateral analysis adapted to breast density using similarity indexes and convolutional neural networks. *Computer Methods and Programs in Biomedicine* 156:191–207, 2018. <https://doi.org/10.1016/j.cmpb.2018.01.007>
- Dromain C et al.: Computed-aided diagnosis (CAD) in the detection of breast cancer. *European Journal of Radiology* 82(3):417–423, 2013
- Felzenszwalb PF, Huttenlocher DP: Efficient graph-based image segmentation. *International Journal of Computer Vision* 59(2):167–181, 2004
- Heath, M., et al.: Digital database for screening mammography (2001). URL <http://marathon.csee.usf.edu/Mammography/Database.html>
- Li H, Zhuang S, Li D, Zhao J, Ma Y: Benign and malignant classification of mammogram images based on deep learning. *Biomedical Signal Processing and Control* 51:347–354, 2019. <https://doi.org/10.1016/j.bspc.2019.02.017>
- de Nazaré Silva, J., de Carvalho Filho, A.O., Corrêa Silva, A., Cardoso de Paiva, A., Gattass, M.: Automatic detection of masses in mammograms using quality threshold clustering, correlogram function, and svm. *Journal of Digital Imaging* 28(3), 323–337 (2015). DOI <https://doi.org/10.1007/s10278-014-9739-3>
- Pedro RWD, Machado-Lima A, Nunes FL: Is mass classification in mammograms a solved problem? - a critical review over the last 20 years. *Expert Systems with Applications* 119:90–103, 2019. <https://doi.org/10.1016/j.eswa.2018.10.032>
- Ronneberger O, Fischer P, Brox T: U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*. Cham: Springer International Publishing, 2015, pp. 234–241
- de Sampaio WB, Silva AC, de Paiva AC, Gattass M: Detection of masses in mammograms with adaption to breast density using genetic algorithm, phylogenetic trees, lbp and svm. *Expert Systems with Applications* 42(22):8911–8928, 2015
- Sharma S, Khanna P: Computer-aided diagnosis of malignant mammograms using zernike moments and svm. *Journal of Digital Imaging* 28(1):77–90, 2015
- Zhang X et al.: Whole mammogram image classification with convolutional neural networks. In: *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017, pp. 700–704
- Zuiderveld, K.: Graphics gems iv. In: P.S. Heckbert (ed.) *Graphics gems*, chap. Contrast limited adaptive histogram equalization, pp. 474–485. Academic Press Professional, Inc., San Diego (1994)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.