Evaluation of U-net-based Image Segmentation Model to Digital Mammography

Priscilla Cho[a] and Hong-Jun Yoon[b]

[a]Emory University, 201 Dowman Dr, Atlanta, GA, United States
[b]Oak Ridge National Laboratory,* 1 Bethel Valley Road, Oak Ridge, TN, United States

ABSTRACT

Detecting suspicious lesions in medical imaging is the important first step in computer-aided detection (CAD) systems. However, detecting abnormalities in breast tissue is difficult due to the lesion's varying size, shape, margin, and contrast with the background tissue. We focused on mass segmentation, a method that provides notable morphological features by outlining contours of masses. Accurate segmentation is crucial for correct diagnosis. Recent advancements in deep learning have improved object detection and segmentation, and these techniques are also being applied to medical imaging studies. We focused on U-net, which is a recently developed mass segmentation algorithm based on a fully convolutional network. The U-net architecture consists of (1) a contracting path to increase the resolution of the output and (2) a symmetric expanding path to better locate the region of interest. The performance of a U-net model was tested with 63 digital mammograms from INbreast, a publicly available database. We trained the model with images resized to 40x40 pixels and conducted 10-fold cross-validation to prevent overfitting. The model's performance with respect to breast density and the lesion's BI-RADS rating was also investigated. Dice coefficients (DC) were used as a performance measure to compare the predicted segmentation of the model with the ground truth. Logistic regression and an analysis of variance were performed to determine the significance of the DCs with regards to breast density and lesion behavior and to calculate the 95% confidence interval. The average DC was 0.80. The difference between DCs for BI-RADS 2 and 4c and for BI-RADS 2 and 5 were significant, suggesting that the model has more difficulty in segmenting benign lesions.

Keywords: U-net, fully convolutional networks, image segmentation, digital mammography, computer aided diagnosis

1. INTRODUCTION

Breast cancer is the most commonly diagnosed cancer among women worldwide.[1,2] If breast cancer is detected early, the 5-year survival rate of the patient is around 99%, but in the case of metastasis, the rate drops to 27%.[3,4] Therefore, early detection, diagnosis, and treatment are the keys to enhancing the survival rate of breast cancer patients.[5] Mammography is a method of imaging that utilizes a low dose X-ray system to provide two kinds of examinations.[6] Screening mammography is a type of examination performed using mammography as a means of early detection of breast abnormalities.[6,7] After a mammogram is taken, a radiologist analyzes the images to identify any suspicious lesions. However, this task can be difficult due to mammograms having a low contrast.[6] High breast density can also mask the presence of suspicious tumors. As a result, computer-aided detection (CAD) systems have been developed to enhance the accuracy of lesion detection by suggesting a second opinion to radiologists.[8,9]

Mass segmentation is crucial in CAD systems because it outlines the contours of potential masses to provide key morphological features.[1,2] The margin and shape of breast masses can be found with mass segmentation to

*This manuscript has been authored in part by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

Priscilla Cho: E-mail: priscilla.cho@emory.edu, Telephone: (865)722-3782
Hong-Jun Yoon: E-mail: yoonh@ornl.gov, Telephone: (865)241-2626

determine the malignancy of abnormalities. If the breast mass is regular in shape, it is most likely benign. When the breast mass has an irregular margin, it is most likely malignant. Therefore, accurate mass segmentation is necessary for better classification of breast masses. However, mass segmentation is a difficult task for various reasons, such as ill-defined boundaries of masses. Traditional segmentation methods require the fine-tuning of many parameters like thresholds and the manual curation of features, so the results are not very stable. Moreover, traditional methods have challenges in achieving the automatic end-to-end precise segmentation of breast masses. Therefore, segmentation results obtained from traditional methods are not very ideal.[10]

With the recent development of deep learning (DL) techniques, the field of medical imaging has also progressed.[4,11] DL-based methods use convolutional neural networks (CNN) to obtain meaningful features from the training data and have had promising results. CNN can automatically learn local area features and high-level abstract features through multi-layer network structures which is usually better than manual extraction.[4] However, CNN-based mass segmentation uses an image block around a pixel as the network input to classify, which results in low calculation efficiency due to repeated convolution calculation during training and predicting.[12] Ronneberger et al. proposed a more elegant segmentation method called U-net, which is based on a fully convolutional network.[13] The U-net architecture consists of a contracting path to capture context and increase the output's resolution and a symmetric expanding path to improve the lesion's localization.[13] In this study, we evaluate the performance of a general U-net model using digital mammograms from a publicly available database. Reviews of the traditional and latest techniques of breast tumor segmentation are reviewed in Section 2, and the dataset and models utilized in this study are reviewed in the first part of Section 3.

For this study of the automatic segmentation of breast abnormalities, we concentrated on how the segmentation performance differs if the subtlety of the tumor region, breast density, and characteristic of the mass changes. In this paper, we made comparisons of the performance of tumor segmentation with respect to breast density and the behavior of tumors as categorized by Breast Imaging Reporting and Data System (BI-RADS) ratings. The study design and evaluation methods are reviewed in Section 3, and the performance scores and confidence intervals are reported in Section 4. Lastly, the discussion and future directions of the study are in Section 5.

2. REVIEW

Image segmentation is the process of partitioning an image into multiple segments in digital image processing and computer vision.[14] The purpose of image segmentation is to change the representation of images into meaningful objects to simplify the analysis. Segmentation to breast imaging and digital mammography is applied widely for several purposes, such as detecting breast contour[15] and identifying pectoral muscle from the mediolateral oblique (MLO) view.[16] In this paper, we are focusing on distinguishing cancerous regions in the images. Segmenting suspicious regions from digital mammograms is essential for the automatic detection and diagnosis of breast cancers for better feature extraction and fast computation.[17]

Suspicious lesion segmentation from the region-of-interest (ROI) of the mammogram has a long history since the emergence of computerized, digital imagery to medical imaging. Thresholding is the primary and most common approach.[18] The method is based upon the characteristic that the pixel intensity of suspicious objects is higher than that of background pixels. Edge detection based on convolution with the mathematical morphological filters such as Sobel operators[19] is another traditional approach.[20] Such operators work better if the contour and shape of the tumors are apparent and simple. However, those traditional techniques do not perform well if the tumor region is ambiguous. Most breast abnormalities do not exhibit such clarity; thus, the related researches focus on establishing robust methods to handle such complexity. Active contour model[21] is another leading technique of image segmentation applied to breast imaging. The active contour models utilize energy functions and constraints to morph and deform a set of connected components and move the components slowly, deforming itself towards the object boundaries.

The emergence of deep learning (DL)[22] and convolutional neural networks (CNN)[23] suggests solutions to complex image processing tasks, including image segmentation. A fully convolutional network (FCN)[13] is another kind of CNN. The last fully connected layer is substituted by another convolution layer with a large receptive field, capturing the global context of the image. Mask R-CNN consists of two networks; the first one utilizes the

FCN layers to generate image masks, then the second network identifies the objects in the image region. Mask R-CNN was also applied to breast tumor detection and segmentation and reported superior performance. In this paper, we adopted a U-net,[13] built upon the concept of the FCN, but applied multiple upsampling layers to achieve finer details of segmentation.

3. METHODS

3.1 Dataset

The INbreast dataset contains 115 cases, where 90 cases are of women with lesions on both breasts and 25 cases are of mastectomy patients.[24] Images are from the breast center at the São João Universitary Hospital Center in Porto, Portugal. The images are captured with MammoNovation Siemens full-field digital mammography containing a solid-state detector of amorphous selenium. There are 116 masses in total, with the sizes ranging from 15 mm^2 to 3689 mm^2.[9] The pixel size of the mammograms is 70 μm, and the bit depth is 14-bit. The masses are annotated by a specialist with the lesion type and detailed contours for each mass. Calcification lesions were excluded because this study focused on mass segmentation. We used images where we could obtain the lesion in the center of the ROI. Images with lesion width or height greater than 512 pixels were also excluded because the ROI will be filled with the lesion, thus having no meaning for segmentation. As a result, we extracted 63 ROIs of 512x512 pixels for the study.

3.2 U-net

U-net is a family of fully convolutional networks developed for object segmentation of digital images. The structure of U-net consists of a contractive downsampling path followed by an expansive upsampling path.[2] It uses the high-resolution information of lesions obtained by the shallow layer and supplies the missing detailed spatial information during the upsampling process to have better segmentation results.[4] We have noticed that several breast cancer segmentation studies applied image resizing from 512x512 pixels to 40x40 pixels, presumably due to the limited number of training samples.[5,11] Therefore, We also resized the images to 40x40 pixels and assigned two layers of downsampling and upsampling. The U-net model was implemented using Keras and TensorFlow backend on a DGX station with a V100 GPU.

3.3 Performance Measure

We used the Dice coefficient (DC) as a measure of the model's performance.[25] The DC measures how well the model segmented the input image compared to the ground truth image manually segmented by experienced radiologists. The formula is shown below in Equation (1). For this study, a in the equation represents the number of elements in the predicted output of the model while b represents the number of elements in the ground truth. The U-net model architecture and performance measure applied to the study is illustrated in Figure 1

$$DC = \frac{2|a \cap b|}{|a| + |b|} \tag{1}$$

3.4 Study Design

Due to the limited number of training samples, we applied image augmentation with flip, rotate, shift, shear, and zoom. We also applied 10-fold cross-validation (CV) tests for the evaluation of the U-net model for mass segmentation. We divided the data into 10 segments and conducted 10 iterations of training and validation. CV tests are conducted so that in each iteration, a different segment of the data is used for validation while the remaining 9 segments are used for training.[26] The 10-fold CV and data augmentation are both standard practices of model training and evaluation to prevent overfitting of the data during training when the dataset is limited.

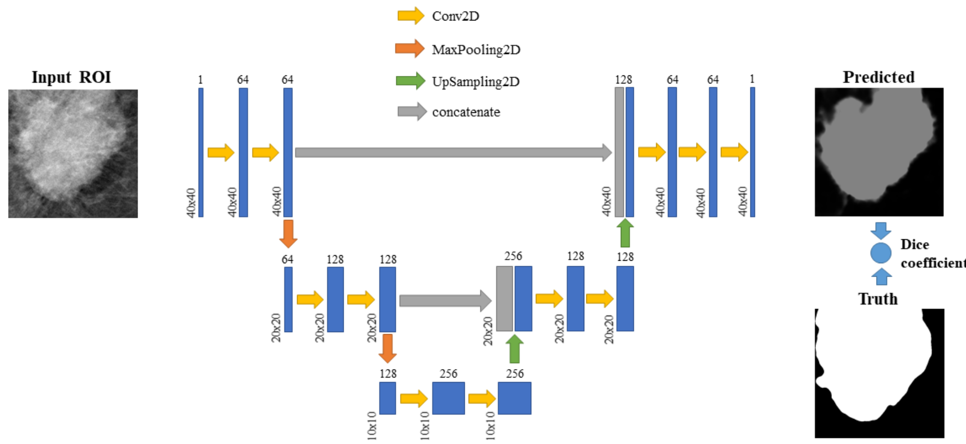


Figure 1: The U-net architecture employed in this study. We reduced the size of the ROI from 512x512 to 40x40, applied the U-net model to determine the contour, and compared the predicted output with the reviewer-annotated ground truth. Performance measure is with the Dice coefficient.

3.5 Evaluation

In addition to obtaining the average DC for all ROIs, we performed further investigation of mass segmentation with respect to breast density and lesion behavior. Breast density is determined by the amount of stromal tissue, epithelial tissue, and fat found in the breast.[27] A high breast density level means that more stromal and epithelial tissue is present. We used the breast density ratings of 1 through 4 reported by the INbreast dataset, with 1 being the least dense and 4 being the most dense. For lesion behavior, Breast Imaging Reporting and Data System (BI-RADS) ratings of 2, 3, 4a, 4b, 4c, 5, and 6 were used.[28] A BI-RADS rating of 2 is typically benign, and ratings 4a-6 indicate differing chances of malignancy with rating 6 being a histologically confirmed malignancy.[29] However, the 63 ROIs used in the study only contained images with density levels of 1-3 and BI-RADS ratings of 2, 3, 4a, 4c, 5, and 6. The average DC was found for each breast density level and BI-RADS score present in the ROIs used for this study. Logistic regression was performed to determine if the average DCs were dependent on breast density and lesion behavior. Analysis of variance was performed to find the significance of the average DCs with regards to breast density and lesion behavior and to calculate the 95% confidence interval (CI). The logistic regression and analysis of variance are conducted with the glm function in R.

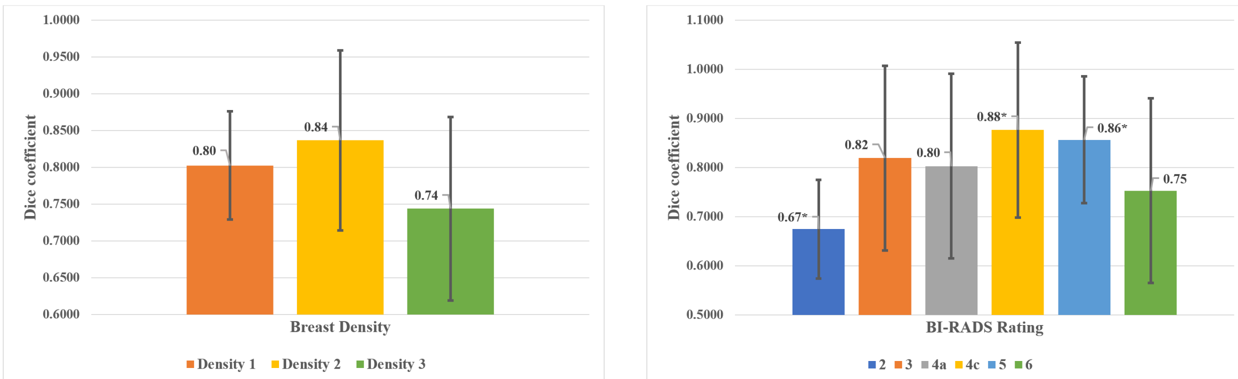
4. RESULTS

Based on the 10-fold CV, the average DC was 0.80, which is comparable to plain U-net implementations reported by other studies.[30,31] We performed a further analysis with respect to breast density, listed in Table 1. The table shows that the segmentation performance of a dense breast is lower than that of a fatty breast, but there was not a significant performance difference. Table 1 also lists the average DCs with respect to BI-RADS ratings. We observed a significant difference between DCs for BI-RADS ratings 2 and 4C and for BI-RADS ratings 2 and 5. These results suggest that the model's segmentation performance is higher for malignant tumors than benign ones. Average DCs and their confidence intervals are illustrated in Figure 2.

Density	# cases	Average DC
1	30	0.80 (0.73-0.88)
2	17	0.84 (0.71-0.96)
3	16	0.74 (0.62-0.87)

BIRADS	# cases	Average DC
2	15	0.67* (0.57-0.78)
3	6	0.82 (0.63-1.01)
4a	6	0.80 (0.61-0.99)
4c	7	0.88* (0.70-1.05)
5	23	0.86* (0.73-0.99)
6	6	0.75 (0.56-0.94)

Table 1: Average DCs with respect to breast density levels and BI-RADS ratings of the mammographic images. Significant performance differences are shown by marking the corresponding DCs with an asterisk.



(a) (b)

Figure 2: Plot of the average DCs and their 95% confidence intervals with respect to (a) breast density and (b) BI-RADS rating of the mammographic images.

5. DISCUSSION

In this study, we reviewed the latest DL-based image segmentation model called U-net, implemented the model on the popular DL software packages, and performed evaluation studies with masses on digital mammography. DCs were applied for the segmentation performance measure. We used data augmentation and designed a 10-fold CV to avoid overfitting with the limited sample size.

One previous study by de Moor et al. was similar to this study except with free receiver operating characteristic (FROC) analysis as the performance measure and showed promising results.[32] Another study by Zeiser et al. tested the segmentation performance of six models derived from U-net to digital mammography and resulted in a DC of 0.79 for the best performing model.[33]

Based on the DC values with respect to breast density, we observed that the segmentation performance became lower for dense breasts, meaning that the segmentation of breast abnormalities from the images with more dense tissue becomes difficult. The difference between DCs with regards to breast density levels however was not significant. On the contrary, the model performed better with malignant tumors than the benign ones, suggesting that the subtle boundaries and low intensity of benign tumors made them more difficult to segment. However, we were unable to see significance in many cases, and we are not certain if this is due to the limited number of training samples.

For future studies, different U-net models, such as conditional residual U-net[2] and attention dense-U-net,[4] can be tested on whether their segmentation performance differs with regards to breast density and lesion behavior. More samples can be obtained for a larger dataset. Well performing U-net models could also be found for each step of density level detection, mass segmentation, and lesion behavior diagnosis. The different U-net models can be tested on the accuracy of assessing breast density levels or BI-RADS ratings in addition to mass segmentation performance by comparing predicted assessments to the manually labeled ratings reported in the INbreast dataset.

REFERENCES

[1] Zhu, W., Xiang, X., Tran, T. D., Hager, G. D., and Xie, X., "Adversarial deep structured nets for mass segmentation from mammograms," in [*2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*], 847–850, IEEE (2018).

[2] Li, H., Chen, D., Nailon, W. H., Davies, M. E., and Laurenson, D., "Improved breast mass segmentation in mammograms with conditional residual u-net," in [*Image Analysis for Moving Organ, Breast, and Thoracic Images*], 81–89, Springer (2018).

[3] DeSantis, C. E., Ma, J., Goding Sauer, A., Newman, L. A., and Jemal, A., "Breast cancer statistics, 2017, racial disparity in mortality by state," *CA: a cancer journal for clinicians* **67**(6), 439–448 (2017).

[4] Li, S., Dong, M., Du, G., and Mu, X., "Attention dense-u-net for automatic breast mass segmentation in digital mammogram," *IEEE Access* **7**, 59037–59047 (2019).

[5] Xu, S., Adeli, E., Cheng, J.-Z., Xiang, L., Li, Y., Lee, S.-W., and Shen, D., "Mammographic mass segmentation using multichannel and multiscale fully convolutional networks," *International Journal of Imaging Systems and Technology* (2020).

[6] Tang, J., Rangayyan, R. M., Xu, J., Naqa, I. E., and Yang, Y., "Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances," *IEEE Transactions on Information Technology in Biomedicine* **13**(2), 236–251 (2009).

[7] Samala, R. K., Chan, H.-P., Hadjiiski, L., Helvie, M. A., Wei, J., and Cha, K., "Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography," *Medical physics* **43**(12), 6654–6666 (2016).

[8] Giger, M. L., Karssemeijer, N., and Schnabel, J. A., "Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer," *Annual review of biomedical engineering* **15**, 327–357 (2013).

[9] Min, H., Wilson, D., Huang, Y., Liu, S., Crozier, S., Bradley, A. P., and Chandra, S. S., "Fully automatic computer-aided mass detection and segmentation via pseudo-color mammograms and mask r-cnn," in [*2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*], 1111–1115, IEEE (2020).

[10] Oliver, A., Freixenet, J., Marti, J., Perez, E., Pont, J., Denton, E. R., and Zwiggelaar, R., "A review of automatic mass detection and segmentation in mammographic images," *Medical image analysis* **14**(2), 87–110 (2010).

[11] Dhungel, N., Carneiro, G., and Bradley, A. P., "Deep learning and structured prediction for the segmentation of mass in mammograms," in [*International Conference on Medical image computing and computer-assisted intervention*], 605–612, Springer (2015).

[12] Ciresan, D., Giusti, A., Gambardella, L. M., and Schmidhuber, J., "Deep neural networks segment neuronal membranes in electron microscopy images," in [*Advances in neural information processing systems*], 2843–2851 (2012).

[13] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*International Conference on Medical image computing and computer-assisted intervention*], 234–241, Springer (2015).

[14] Jayaraman, S., Esakkirajan, S., and Veerakumar, T., "Digital image processing," *Recuperado de https://books. google. com. pe/books* (2009).

[15] Martí, R., Oliver, A., Raba, D., and Freixenet, J., "Breast skin-line segmentation using contour growing," in [*Iberian conference on pattern recognition and image analysis*], 564–571, Springer (2007).

[16] Raba, D., Oliver, A., Martí, J., Peracaula, M., and Espunya, J., "Breast segmentation with pectoral muscle suppression on digital mammograms," in [*Iberian conference on pattern recognition and image analysis*], 471–478, Springer (2005).

[17] Sahiner, B., Petrick, N., Chan, H.-P., Hadjiiski, L. M., Paramagul, C., Helvie, M. A., and Gurcan, M. N., "Computer-aided characterization of mammographic masses: accuracy of mass segmentation and its effects on characterization," *IEEE transactions on medical imaging* **20**(12), 1275–1284 (2001).

[18] Otsu, N., “A threshold selection method from gray-level histograms,” *IEEE transactions on systems, man, and cybernetics* **9**(1), 62–66 (1979).

[19] Kanopoulos, N., Vasanthavada, N., and Baker, R. L., “Design of an image edge detection filter using the sobel operator,” *IEEE Journal of solid-state circuits* **23**(2), 358–367 (1988).

[20] Anjna, E. A. and Er, R. K., “Review of image segmentation technique,” *International Journal of Advanced Research in Computer Science* **8**(4) (2017).

[21] Kass, M., Witkin, A., and Terzopoulos, D., “Snakes: Active contour models,” *INTERNATIONAL JOURNAL OF COMPUTER VISION* **1**(4), 321–331 (1988).

[22] Hinton, G. E. and Osindero, S., “A fast learning algorithm for deep belief nets,” *Neural Computation* **18**, 2006 (2006).

[23] Krizhevsky, A., Sutskever, I., and Hinton, G. E., “Imagenet classification with deep convolutional neural networks.,” in [*In Advances in the Neural Information Processing System,*], (2012).

[24] Moreira, I. C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M. J., and Cardoso, J. S., “Inbreast: toward a full-field digital mammographic database,” *Academic radiology* **19**(2), 236–248 (2012).

[25] Dice, L. R., “Measures of the amount of ecologic association between species,” *Ecology* **26**(3), 297–302 (1945).

[26] Singh, G. and Panda, R. K., “Daily sediment yield modeling with artificial neural network using 10-fold cross validation method: a small agricultural watershed, kapgari, india,” *International Journal of Earth Sciences and Engineering* **4**(6), 443–450 (2011).

[27] Wang, A. T., Vachon, C. M., Brandt, K. R., and Ghosh, K., “”breast density and breast cancer risk: A practical review”,” *Mayo Clinic Proceedings* **89**(4), 548–557 (2014).

[28] BIRADS, B. I.-R., “Data system (bi-rads),” *American College of Radiology* (1993).

[29] Balleyguier, C., Ayadi, S., Van Nguyen, K., Vanel, D., Dromain, C., and Sigal, R., “”birads™ classification in mammography”,” *European Journal of Radiology* **61**(2), 192–194 (2007).

[30] Jiang, J., Hu, Y.-C., Liu, C.-J., Halpenny, D., Hellmann, M. D., Deasy, J. O., Mageras, G., and Veeraraghavan, H., “Multiple resolution residually connected feature streams for automatic lung tumor segmentation from ct images,” *IEEE transactions on medical imaging* **38**(1), 134–144 (2018).

[31] Dong, H., Yang, G., Liu, F., Mo, Y., and Guo, Y., “Automatic brain tumor detection and segmentation using u-net based fully convolutional networks,” in [*annual conference on medical image understanding and analysis*], 506–517, Springer (2017).

[32] de Moor, T., Rodriguez-Ruiz, A., Gubern Mérida, A., Mann, R., and Teuwen, J., “”automated lesion detection and segmentation in digital mammography using a u-net deep learning network”,” in [*14th International Workshop on Breast Imaging (IWBI 2018)*], Krupinski, E. A., ed., **10718**, 23–29, International Society for Optics and Photonics, SPIE (2018).

[33] Zeiser, F. A., da Costa, C. A., Zonta, T., Marques, N. M. C., Roehe, A. V., Moreno, M., and da Rosa Righi, R., “”segmentation of masses on mammograms using data augmentation and deep learning”,” *Journal of Digital Imaging* **33**(4), 858–868.