U-Net transfer learning backbones for lesions segmentation in breast ultrasound images

Mohamed Bal-Ghaoui, My Hachem El Yousfi Alaoui, Abdelilah Jilbab, Abdennacer Bourouhou Electronic Systems, Sensors and Nanobiotechnologies, Mohammed V University in Rabat, Rabat, Morocco

Article Info	ABSTRACT						
Article history:	Breast ultrasound images are highly valuable for the early detection of breast						
Received Jan 3, 2023 Revised Mar 23, 2023 Accepted Apr 7, 2023	cancer. However, the drawback of these images is low-quality resolution and the presence of speckle noise, which affects their interpretability and makes them ra- diologists' expertise-dependent. As medical images, breast ultrasound datasets are scarce and imbalanced, and annotating them is tedious and time-consuming. Transfer learning as a deep learning technique, can be used to overcome the						
Keywords:	dataset deficiency in available images. This paper presents the implementation						
<i>Keyworas:</i> Breast ultrasound images Deep learning Lesions segmentation Transfer learning U-Net backbones	of transfer learning U-Net backbones for the automatic segmentation of breast ultrasound lesions and implements a threshold selection mechanism to deliver optimal generalized segmentation results of breast tumors. The work uses the public breast ultrasound images (BUSI) dataset and implements ten state-of-the- art candidate models as U-Net backbones. We have trained these models with a five-fold cross-validation technique on 630 images with benign and malignant cases. Five out of ten models showed good results, and the best U-Net backbone was found to be DenseNet121. It achieved an average Dice coefficient of 0.7370 and a sensitivity of 0.7255. The model's robustness was also evaluated against normal cases, and the model accurately detected 72 out of 113 images, which is higher than the four best models.						
	This is an open access article under the CC BY-SA license.						

СС

Corresponding Author:

Mohamed Bal-Ghaoui Electronic Systems, Sensors and Nanobiotechnologies, Mohammed V University in Rabat Rabat, Morocco Email: mohamed_balghaoui@um5.ac.ma

INTRODUCTION 1.

Breast ultrasound images play a very important role in the detection and diagnosis of breast cancer. It is a more productive alternative to mammography because it is safer, more accurate, and less expensive [1], [2]. However, these images are known for their difficulty in interpretation due to their low resolution and the presence of speckle noise. For this reason, computer-aided diagnosis (CAD) systems based on deep learning (DL) can be used to ease ultrasound image analysis and provide accurate decision support to radiologists. These intelligent systems have proved their importance in computer vision tasks such as object detection, classification, and segmentation.

Nowadays, DL is regarded as the main approach used in medical image analysis and has proven to outperform conventional techniques [3]. It is known that the effectiveness of a DL model is strongly related to the dataset being used, with larger datasets being better. Transfer learning (TL), on the other hand, is a DL technique that leverages the use of pre-trained models to address the deficiency of the dataset. Behboodi et al. [4] demonstrated that TL using U-Net pre-trained on natural salient object detection images,

such as the XPIE dataset [5], along with simulated ultrasound images for breast lesions yielded good segmentation results on the limited dataset they used. On the other hand, in [6], TL demonstrated good classification results in comparison to a custom-implemented convolutional neural network (CNN) classifier applied to breast ultrasound lesions.

In the medical field, data annotation is an expensive and time-consuming task, as it requires the expertise and efforts of radiologists. However, with TL, pre-trained models can be used to tackle this difficulty and provide a foundation for the feature map. In TL, there are two approaches, depending on how the pre-trained models are used: i) feature extraction and ii) fine-tuning.

Image segmentation is a technique to partition an image into well-defined and recognizable parts. Its aim is to ease the image analysis and provide meaningful representations of the objects in the image. In medical applications, image segmentation can be helpful for detecting and isolating tumors for better diagnosis.

In the literature, breast ultrasound image segmentation techniques are categorized commonly into five classes [7], [8]: i) deformable models, ii) graph-based models, iii) machine learning-based models, and iv) classical approaches. It is also demonstrated that DL-based approaches outperformed conventional techniques. Some of the popular segmentation DL architectures used for medical images are U-Net [9] and its variants [10] such as 3D U-Net [11], attention U-Net [12] and U-Net++ [13], LeNet [14], fully convolutional network (FCN) [15], SegNet [16], and more specific architectures such as small tumor-aware network (STAN) [17] and residual-dilated-attention-gate-UNet (RDAU-Net) [18].

In deep learning, image segmentation is narrowed down to a pixel classification problem where DLbased models predict and attribute each pixel to its corresponding class [19]. This work focuses mainly on binary segmentation types of classes of ultrasound breast lesions using DL-based methods. Lesions regions of interest (RoI) are considered the positive class, and the background as negative. We have also implemented a threshold selection mechanism to determine the optimal prediction threshold and generalize good segmentation results on unseen data. U-Net backbone models such as VGG16, VGG19, ResNet50, ResNet50, Se-ResNet50, Se-ResNet50, MobileNetV2, and DenseNet121 were implemented and evaluated using a crossvalidation technique.

2. METHOD

2.1. Dataset preparation

We used a publicly available dataset known as the breast ultrasound images (BUSI) [20] dataset, which includes three image categories: normal, benign, and malignant. For this study, we chose the Benign and Malignant classes. Normal cases were excluded from the dataset because of the assumption that the model would be less sensitive to breast lesions, given the quality of ultrasound images, speckle noise, and the difficulty of extracting features and detecting lesions boundaries. In order to ensure the model's sensitivity to breast lesions, all training samples contained lesions. Introducing ground-truth masks with 0-pixel values of normal cases in the training dataset could bias the model's sensitivity to tumors. However, the normal class was considered in evaluating the models' robustness. We trained U-Net backbones on 630 images with their corresponding ground-truth masks, and we used the partitioning shown in Table 1 for training, validation, and testing sets.

Table 1. BUSI dataset preparation	on
-----------------------------------	----

Train + Val set (90%) ^a			Test set (Test(100%) ^c		
Total	80% Train	20% Validation	80% Malignant	20% Benign	Normal	
Total	453	114	50	13	113	
^a Training and Validation set the split was done in cross-validation						

with 20% of data for validation and 80% for training

^bTest set, a combination of Malignant and Benign images, acts as unseen data

^cTest set, a total of 113 Normal breast ultrasound images

The images in the BUSI dataset have different sizes and were averaged to a size of 500×500 . As a pre-processing step, the images were resized to 256×256 dimensions and then normalized to values between 0 and 1. To ensure that the study was more objective, reproducible, and meaningful for TL, no data augmentation was performed.

2.2. U-Net: the base architecture

U-Net [9] was used as the main architecture. It is a well-known architecture for segmentation tasks and was originally implemented for biomedical images. There are many variants of U-Net. In the context of TL, Amiri *et al.* in [21] demonstrated that fine-tuning the contracting path of U-Net yields better results than fine-tuning the expanding path.

From this context, pre-trained models were leveraged as feature extractor backbones based on the U-Net architecture to segment breast ultrasound lesions. The backbones were loaded with ImageNet weights, and all encoder layers were frozen. The models were trained for 100 epochs with a reduced batch size of 8, using the Adam optimizer initiated with a learning rate of 0.0001. An adaptive learning rate process was also implemented during the training with a reduction factor of 0.1, and patience of 10 epochs while monitoring validation loss during all 5 cross-validation folds, and only saving the best models. Figure 1 provides an overview of the study workflow.



Figure 1. Study workflow overview

2.3. Cross-validation

Due to dataset limitations, the following partitioning was chosen: 90% of the data was used as seed for cross-validation (CV) with 5-folds, and 10% was reserved as the test set to evaluate the final prediction stage as shown in Table 1. CV is an excellent approach to assessing model performance on a limited dataset. By allowing models to learn and validate features from all samples, CV helps mitigate the risk of overfitting. A five-fold CV was used in this study, and it is commonly preferred in many papers, such as in [4] and [7].

2.4. Evaluation metrics

Although segmentation is regarded as a classification problem, evaluation metrics are different [19]. In fact, some of the metrics such as accuracy (TP+TN)/(TP+TN+FN+FP) are considered misleading and don't reflect the actual model performance. Accuracy includes the True Negative, representing in this case, the back-ground, which has the majority of the pixels. Thus, Accuracy would be always much higher compared to other metrics [22]. Müller *et al.* [22] highlighted important guidelines for metrics used in medical image segmentation. Their research includes the usage of the dice similarity coefficient (DSC) (1) and Intersection-over-Union (IoU) (2) as popular metrics in the category of F-measure-based metrics. Sensitivity (3) and specificity (4) are also considered established standards for performance evaluation in medicine [23], [24]. In addition to computed metrics, visual interpretation is always recommended in evaluating segmentation tasks. However, lower computational metric values might not always imply bad predictions. Cheng *et al.* [25] obtained accurate visual performance despite having a lower DSC when segmenting ultrasound lung.

$$Dice = \frac{2*TP}{2*TP + FP + FN} \tag{1}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{2}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

2.5. Segmentation loss function

This is an important parameter in image segmentation. During the training process, the model tries to minimize the objective function to cover edge cases. Choosing a well-suited loss function can enhance model performance and aid in fast convergence, and it must be selected according to the nature of the task. The loss semantic segmentation survey paper in [26] summarizes well-known loss functions and categorizes them into four types: i) distributed-based, ii) region-based, iii) boundary-based, and iv) compounded losses. Furthermore, it highlights different scenarios for choosing the appropriate loss function for segmentation. In this article, the Focal Tversky loss was implemented, as it demonstrated good results in recognizing difficult examples [27], particularly in ultrasound images.

2.6. The optimal prediction threshold

Performing binary segmentation involves two classes: the foreground region of interest (RoI) and the background, which represents the majority of the pixels. In this work, the ground truth highlights lesions with white color and the background as black. Based on a particular threshold, each pixel's prediction would be either 1 or 0. For instance, pixels with a probability of 0.5 or higher would be set to 1, defining a lesion, while pixels with lower probabilities would be given the value 0 and considered part of the background.

It is noteworthy that no optimal prediction segmentation threshold is known, and many papers choose to use 0.5 as a default threshold. In order to determine the optimal threshold for a given model during the prediction phase, we have implemented a selection mechanism. The process steps are described below and can also be visualized in Figure 2.

- Define a custom threshold range, starting from 0.4 and ending at 0.9 with a step of 0.05.
- For each threshold, compute both the Dice coefficient and the mean Intersection-over-Union.
- At the end of each prediction, map the maximum Dice value and the maximum mean IoU value to their corresponding threshold index.
- Choose the corresponding index for future predictions.



Figure 2. Prediction threshold selection process

3. RESULT AND DISCUSSION

All the training and testing were accomplished using Google Colab Computing Units with access to GPU Tesla V100-SXM2 of type NVIDIA-SMI 460.32.03. Each backbone was trained separately using five-fold cross-validation and evaluated using metrics (described in section 2.4.) such as dice coefficient, mean IoU, sensitivity, and specificity. The overall training time and the model output size were both considered for the evaluation. Table 2 shows the obtained results averaged across the five folds.

Models selection and evaluation results in Table 2, show that five models demonstrated promising results, namely VGG16, VGG19, ResNet50, ResNeXt50 and DenseNet121. This latter had the highest Dice and the lowest Loss with 0.7370 and 0.3713 values respectively. In addition, it had the smallest size of 87.8MB among the five models. Figure 3 shows training and validation Dice across 100 epochs for the two best-performing backbones.

During cross-validation folds, and for each backbone, best-performing models were saved. The majority of the backbones have been seen to perform better on folds 2 and 4. This clearly demonstrated the benefit of cross-validation and showed that most relevant training data were found in the two folds. Table 3 shows CV scores per fold. Note that only the best folds were reported.

		-ugea ue			in out in ones	
Loss	Dice	Sens. ^a	Spec. ^b	Mean IoU	Training Time ^c	Size (MB) ^d
0.4619	0.6587	0.6199	0.9744	0.5994	20	272.1
0.3997	0.6989	0.7118	0.9680	0.7211	12	159.8
0.3982	0.6966	0.7207	0.9657	0.6998	11	180.1
0.3988	0.7161	0.6926	0.9764	0.6917	20	194.2
0.3858	0.7132	0.7316	0.9685	0.6531	54	195.0
0.4252	0.6821	0.6771	0.9698	0.6667	58	204.9
0.3713	0.7370	0.7255	0.9758	0.6246	30	87.8
0.4867	0.6146	0.6201	0.9603	0.6565	22	177.6
0.4266	0.6675	0.7016	0.9608	0.6742	25	204.2
0.4515	0.6599	0.6414	0.9704	0.6966	15	76.0
	Loss 0.4619 0.3997 0.3982 0.3988 0.4252 0.3713 0.4867 0.4266 0.4515	Loss Dice 0.4619 0.6587 0.3997 0.6989 0.3982 0.6966 0.3988 0.7161 0.3858 0.7132 0.4252 0.6821 0.3713 0.7370 0.4867 0.6146 0.4266 0.6675 0.4515 0.6599	Loss Dice Sens. ^a 0.4619 0.6587 0.6199 0.3997 0.6989 0.7118 0.3982 0.6966 0.7207 0.3988 0.7161 0.6926 0.3858 0.7132 0.7316 0.4252 0.6821 0.6771 0.3713 0.7370 0.7255 0.4867 0.6146 0.6201 0.4266 0.6675 0.7016 0.4515 0.6599 0.6414	Loss Dice Sens. ^a Spec. ^b 0.4619 0.6587 0.6199 0.9744 0.3997 0.6989 0.7118 0.9680 0.3982 0.6966 0.7207 0.9657 0.3988 0.7161 0.6926 0.9764 0.3858 0.7132 0.7316 0.9685 0.4252 0.6821 0.6771 0.9698 0.3713 0.7370 0.7255 0.9758 0.4867 0.6146 0.6201 0.9603 0.4266 0.6675 0.7016 0.9608 0.4515 0.6599 0.6414 0.9704	Loss Dice Sens. ^a Spec. ^b Mean IoU 0.4619 0.6587 0.6199 0.9744 0.5994 0.3997 0.6989 0.7118 0.9680 0.7211 0.3982 0.6966 0.7207 0.9657 0.6998 0.3988 0.7161 0.6926 0.9764 0.6917 0.3858 0.7132 0.7316 0.9685 0.6531 0.4252 0.6821 0.6771 0.9698 0.6667 0.3713 0.7370 0.7255 0.9758 0.6246 0.4867 0.6146 0.6201 0.9603 0.6565 0.4266 0.6675 0.7016 0.9608 0.6742 0.4515 0.6599 0.6414 0.9704 0.6966	Loss Dice Sens. ^a Spec. ^b Mean IoU Training Time ^c 0.4619 0.6587 0.6199 0.9744 0.5994 20 0.3997 0.6989 0.7118 0.9680 0.7211 12 0.3982 0.6966 0.7207 0.9657 0.6998 11 0.3988 0.7161 0.6926 0.9764 0.6917 20 0.3858 0.7132 0.7316 0.9685 0.6531 54 0.4252 0.6821 0.6771 0.9698 0.6667 58 0.3713 0.7370 0.7255 0.9758 0.6246 30 0.4867 0.6146 0.6201 0.9603 0.6565 22 0.4266 0.6675 0.7016 0.9608 0.6742 25 0.4515 0.6599 0.6414 0.9704 0.6966 15

Table 2. Metrics averaged across folds for different backbones

^aSensitivity

^bSpecificity

^cModels training time expressed in minutes.

^dModels saved weights size in megabyte (MB)

Models saved on folds 1, 2, and 4 of the five backbones were used to predict unseen test data with 63 images, see Table 1. Fold-specific threshold (described in section 2.6. and reported in Table 2) was used for the prediction. To interpret the results accurately, visual representations are essential. Dataset images are grouped into two categories: i) lesions with well-defined boundaries, and ii) samples with undefined boundaries. For both categories, four samples were plotted. Figure 4(a) and 4(b) represent the segmentation results for the best-performing backbone DenseNet121. The four samples in Figure 4 are structured as follows: an original grayscale image, a ground-truth mask, and a predicted mask displayed from left to right for each category.



Figure 3. Training and validation dice losses for the two best-performing backbones

5752

The model's robustness was also evaluated by predicting untrained classes, such as normal breast cases [7]. Table 3 shows that DenseNet121 had the highest true rate of 63.7%, correctly predicting 72 images out of 113. ResNet50 followed with 46 images, VGG19 with 28, ResNeXt50 with 24 images, and finally, VGG16 with 22 images.

Table 3. Best performing metrics per fold and prediction results for normal classes

1	0	1	1		
	Fold 1		Fold 2		Fold 4
Scores/Fold	Vgg19	Vgg16	Resnext50	DenseNet121	Resnet50
Loss	0.4022	0.377	0.3289	0.3375	0.371
Dice	0.6998	0.7097	0.723	0.7659	0.7328
Sensitivity	0.7074	0.7591	0.8619	0.7637	0.7158
M.IoU	0.7129	0.7433	0.7689	0.6134	0.7485
Pred. Threshold	d 0.9	0.4	0.9	0.9	0.7
Pred. Normal case	es ^a 28	22	24	72	46

^aNumber of correctly predicted images in the normal class with a total of 113 images



Figure 4. Visual interpretation of DenseNet121 prediction results (a) visually easy-samples and (b) visually hard-samples

4. CONCLUSION

Deep learning is a great tool for medical image analysis, particularly in the case of ultrasound images which often have poor image quality, making them challenging to analyze and interpret. DL techniques are proven to accurately provide good results. However, in medicine, data annotation is still a barrier for most datasets. To tackle this, multiple strategies could be used. Data augmentation, for example, could help increase the number of samples by applying multiple image transformations. Transfer learning, on the other hand, can empower the feature engineering task by loading pre-trained models on natural images such as ImageNet. Recent research in TL for ultrasound images demonstrated that pre-trained models on natural salient object detection datasets such as XPIE can outperform ImageNet-based pre-trained models. In the literature, different DL segmentation architectures exist, the most known for biomedical images is U-Net. This article implements ten well-known architectures of TL U-Net backbones for automatic breast ultrasound lesions segmentation along with a threshold selection mechanism for optimal segmentation results. Five backbones out of ten demonstrated good results. One particular model showed promising results, with an average dice coefficient

ISSN: 2088-8708

per fold of 0.737, and sensitivity of 0.7255. This model also proved robust against normal cases, as it yielded good segmentation results, with a detection rate of 64% which is 72 correctly detected out of 113 normal breast cases.

REFERENCES

- [1] R. Guo, G. Lu, B. Qin, and B. Fei, "Ultrasound imaging technologies for breast cancer detection and management: a review," *Ultrasound in Medicine and Biology*, vol. 44, no. 1, pp. 37–70, 2018, doi: 10.1016/j.ultrasmedbio.2017.09.012.
- [2] B. O. Anderson *et al.*, "Breast cancer in limited-resource countries: an overview of the breast health global initiative 2005 guidelines," *The Breast Journal*, vol. 12, no. 1, pp. 3–15, Jan. 2006.
- [3] A. Esteva *et al.*, "Deep learning-enabled medical computer vision," *npj Digital Medicine*, vol. 4, no. 1, pp. 1–9, 2021. doi:10.1038/s41746-020-00376-2.
- [4] B. Behboodi, M. Amiri, R. Brooks, and H. Rivaz, "Breast lesion segmentation in ultrasound images with limited annotated data," in 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Apr. 2020, pp. 1834–1837. doi: 10.1109/ISBI45749.2020.9098685.
- [5] C. Xia, J. Li, X. Chen, A. Zheng, and Y. Zhang, "What is and what is not a salient object? learning salient object detector by ensembling linear exemplar regressors," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2017, pp. 4399–4407.
- [6] M. Bal-Ghaoui, M. H. Yousfi Alaoui, A. Jilbab, and A. Bourouhou, "Enhanced ultrasound breast cancer classification based on sparse data and two customized deep learning approaches," *IREMOS*, vol. 15, pp. 82 031–82 057, 2022, doi: 10.15866/iremos.v15i3.21504.
- [7] Y. Zhang *et al.*, "BUSIS: a benchmark for breast ultrasound image segmentation," *Healthcare*, vol. 10, no. 4, Apr. 2022, doi: 10.3390/healthcare10040729.
- [8] A. E. Ilesanmi, U. Chaumrattanakul, and S. S. Makhanov, "Methods for the segmentation and classification of breast ultrasound images: a review," *Journal of Ultrasound*, vol. 24, no. 4, pp. 367–382, Dec 2021, doi: 10.1007/s40477-020-00557.
- [9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9351, Springer International Publishing, 2015, pp. 234–241.
- [10] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, vol. 9, pp. 82 031–82 057, 2021, doi: 10.1109/AC-CESS.2021.3086020.
- [11] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-net: Learning dense volumetric segmentation from sparse annotation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), 2016, vol. 9901, pp. 424–432, doi: 10.1007/978-3-319-46723-8_49.
- [12] O. Oktay et al., "Attention U-Net: learning where to look for the pancreas," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2018)*, Apr. 2018.
- [13] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: a nested u-net architecture for medical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*). Springer, 2018, vol. 11045, pp. 3–11, doi: 10.1007/978-3-030-00889-5.
- [14] Y. Le Cun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998, doi: 10.1109/5.726791.
- [15] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2015, pp. 431–440, doi: 10.1109/CVPR.2015.7298965.
- [16] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.
- [17] B. Shareef, A. Vakanski, P. E. Freer, and M. Xian, "ESTAN: enhanced small tumor-aware network for breast ultrasound image segmentation," in *Healthcare*, vol. 10, no. 11. IEEE, Nov. 2022, doi: 10.1109/ISBI45749.2020.9098691.
- [18] Z. Zhuang, N. Li, A. N. J. Raj, V. G. V. Mahesh, and S. Qiu, "An RDAU-NET model for lesion segmentation in breast ultrasound images," *PLOS ONE*, vol. 14, no. 8, Aug. 2019, doi: 10.1371/journal.pone.0221535.
- [19] M. Chala, B. Nsiri, M. H. E. Y. Alaoui, A. Soulaymani, A. Mokhtari, and B. Benaji, "An automatic retinal vessel segmentation approach based on convolutional neural networks," *Expert Systems with Applications*, vol. 184, 2021.
- [20] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data in Brief*, vol. 28, Feb. 2020.
- [21] M. Amiri, R. Brooks, and H. Rivaz, "Fine tuning U-net for ultrasound image segmentation: which layers?" in Lec-

ture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, 2019, vol. 11795, pp. 235–242.

- [22] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," in *Image Analysis for Medical Physics*, 2022.
- [23] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool," *BMC Medical Imaging*, vol. 15, no. 1, pp. 1–28, 2015.
- [24] A. Popovic, M. de la Fuente, M. Engelhardt, and K. Radermacher, "Statistical validation metric for accuracy assessment in medical image segmentation," *International Journal of Computer Assisted Radiology and Surgery*, vol. 2, no. 3-4, pp. 169–181, Dec. 2007.
- [25] D. Cheng and E. Y. Lam, "Transfer learning U-Net deep learning for lung ultrasound segmentation," in *Proceedings* of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Oct 2021.
- [26] S. Jadon, "A survey of loss functions for semantic segmentation," in 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Oct. 2020, pp. 1–7, doi: 10.1109/CIBCB48159.2020.9277638.
- [27] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention U-Net for lesion segmentation," in 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Apr. 2019, pp. 683–687, doi: 10.1109/ISBI.2019.8759329.

BIOGRAPHIES OF AUTHORS



Mohamed Bal-Ghaoui B B C received his Engineer's degree in Biomedical Engineering at the University of Mohammed V, ENSAM Rabat. He is a Ph.D. student researcher in Artificial Intelligence applied to Biomedical engineering. His research interest includes AI applied to Ultrasound Images for the early detection of cancer. He is part of E2SN Team - Biomedical Engineering Research Laboratory ENSAM-Rabat. He can be contacted at email: mohamed_balghaoui@um5.ac.ma.



My Hachem El Yousfi Alaoui (D) X (a) is Professor of Biomedical Engineering at the University of Mohammed V, ENSAM-Rabat, he is a member of the research laboratory E2SN - Biomedical Engineering Research Laboratory at ENSAM-Rabat, Mohamed V University in Rabat, Morocco. Members of the E2SN research group, Prof. EL yousfi's current research work is focused on biomedical data processing, AI, IoT and the hardware implementation of associated circuits. He can be contacted at email: h.elyousfi@um5r.ac.ma.





Abdennacer Bourouhou **b** is Professor of Electrical Engineering at ENSAM, Mohammed V University Rabat. He acquired his Ph.D. in Electronics and Signal Processing from Ibn Tofail University in Kenitra, Morocco in April 2008. A. Bourouhou is a member of the Electronic Systems Sensors and Nanobiotechnology (E2SN) team of the STIS research center of UM5 Rabat. his research interests include biosignal processing, wireless sensor networks (WSNs) for environmental protection, and medical monitoring. He can be contacted at email: a.bourouhou@um5r.ac.ma.