Deep learning networks in the segmentation of the left atrial appendage in 2D ultrasound: A comparative analysis

Rafael Fernandes, Helena R. Torres, Bruno Oliveira, João Azevedo, Karen Fan, Alex P. Lee, João L. Vilaça and Pedro Morais

***Abstract*— Left atrial appendage (LAA) is the major source of thromboembolism in patients with non-valvular atrial fibrillation. Currently, LAA occlusion can be offered as a treatment for these patients, obstructing the LAA through a percutaneously delivered device. Nevertheless, correct device sizing is a complex task, requiring manual analysis of medical images. This approach is sub-optimal, time-demanding, and highly variable between experts. Different solutions were proposed to improve intervention planning, but, no efficient solution is available to 2D ultrasound, which is the most used imaging modality for intervention planning and guidance. In this work, we studied the performance of recently proposed deep learning methods when applied for the LAA segmentation in 2D ultrasound. For that, it was created a 2D ultrasound database. Then, the performance of different deep learning methods, namely Unet, UnetR, AttUnet, TransAttUnet was assessed. All networks were compared using seven metrics: i) Dice coefficient; ii) Accuracy iii) Recall; iv) Specificity; v) Precision; vi) Hausdorff distance and vii) Average distance error. Overall, the results demonstrate the efficiency of AttUnet and TransAttUnet with dice scores of 88.62% and 89.28%, and accuracy of 88.25% and 86.30%, respectively. The current results demonstrate the feasibility of deep learning methods for LAA segmentation in 2D ultrasound.**

***Clinical relevance*— Our results proved the clinical potential of deep neural networks for the LAA anatomical analysis.**

I. INTRODUCTION

The left atrial appendage (LAA) is a finger-like projection from the main body of the left atrial (LA). The junction is well-defined by a narrowing in the appendage orifice [1]. There are considerable variations in its size, shape, and relationship with adjacent cardiac and extra-cardiac structures, which can be extremely relevant when procedures are performed. These variations can be classified into 4 morphological types: with "chicken wing" being the most common (48%), followed by "cactus" (30%), "windsock" (19%), and "cauliflower" (3%) [1].

Recent studies indicate that it is the source of thromboembolism in approximately 90% of patients presenting with non-valvular atrial fibrillation (NVAF) [2]. Thus, percutaneous LAA occlusion is presented as a solution to reduce the risk of thromboembolic events [3]. This small and little invasive cardiac intervention aims to close the LAA orifice to the LA through the femoral vein [4] by placing the device that seals the connecting orifice between LA and the LAA. Closure devices are available in several predefined sizes, for each patient, an appropriately sized device must be selected.

The implementation of the respective device requires a pre-analysis through echocardiography or another imaging method such as, Computed tomography angiography (CT) [5] where the largest landing zone is measured for the sizing of the device. Some studies have shown that CT angiography images (CCTA) provide better results for the acquisition of LAA images than other image acquisition techniques [6][7][8], and for their sizing and analysis. However, from image acquisition to surgical intervention there can be anatomical variations, which can influence the chosen device [9]. Therefore, during the surgical procedure, specialists prefer to obtain surgical images through a fluoroscopy and 2D ultrasound imaging namely transesophageal echocardiography (TEE) and use these to identify the respective device [10]. However, this process may include some disadvantages such as the time it takes to correctly size the device, the correct identification of the 3D shape of the LAA through peri-procedural data is not that simple, and since this process is all performed manually, there is a lack of intra- and interobserver reproducibility [11].

Currently, there are some techniques for LAA segmentation using 3D models that extract the region of interest (ROI) of the anatomical structure derived from CT images, simple image-based techniques [11], deformable models [12][13], or even the use of machine learning [14][15][16][17]. Segmentation techniques have also been explored from 3D-TEE images, using deformable method [9][18][19]. One of the aspects that we have not seen in any study is the segmentation of LAA in 2D-TEE images, which is still one of the most used imaging modalities for screening and training. This type is more used due to the ease of visualization and comprehension of the data compared to 3D.

This project was funded by the projects "NORTE-01-0145-FEDER-000045" and "NORTE-01-0145-FEDER-000059", supported by Northern Portugal Regional Operational Programme (Norte2020), under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). It was also financed by national funds, through FCT - Foundation for Science and Technology and FCT/MCTES under the project UIDB/05549/2020, UIDP/05549/2020, CEECINST/00039/2021 and LASI-LA/P/0104/2020. It was also funded by the Innovation Pact Health From Portugal, co-funded from the "Mobilizing Agendas for Business Innovation" of the "Next Generation EU" program of the Recovery and Resilience Plan (RRP), concerning "Capitalization and Business Innovation", under the Regulation of the Incentive System "Agendas for Business Innovation".

R. Fernandes, J. Azevedo, JL. Vilaça and P. Morais are with 2Ai – School of Technology, IPCA, Barcelos, Portugal and LASI-Associate Laboratory of Intelligent Systems, Guimarães, Portugal. (email: jrfernandes@ipca.pt, jvilaca@ipca.pt, pmorais@ipca.pt).

HR. Torres and B. Oliveira are with 2Ai – School of Technology, IPCA, Barcelos, Portugal, with Algoritmi Center, School of Engineering, University of Minho, Guimarães, Portugal, with Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal, and with ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal.

Karen Fan and Alex P.Lee - Division of Cardiology, Department of Medicine and Therapeutics, Prince of Wales Hospital, Hong Kong, China; Laboratory for Cardiac Imaging and 3D Printing, Li Ka Shing Institute of Health Science, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China

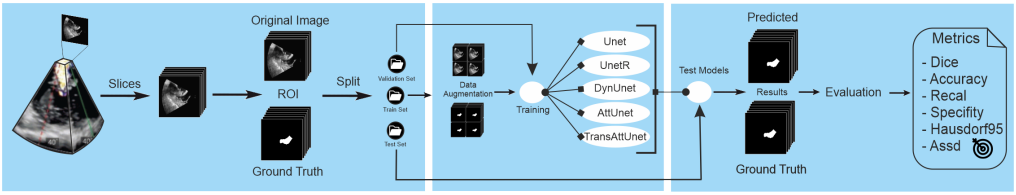


Fig. 1. Graphical summary of the comparative study

In this study, we aim to evaluate and compare convolutional neural networks for LAA segmentation on 2D TEE.

This document is structured as follows. In Section II, a description of the compared networks and its parameters is presented. The implementation details of each network is presented in Section III. In Section IV, the results of the metrics and their description are presented. Section V refers to the discussion of the results presented above and future work, and finally, Section VI refers to the conclusions.

II. METHEDOLOGY

A. Neuronal Networks

Four neural networks were selected for performance study on 2D LAA segmentation. Figure 1 represents a graphical summary of the comparison study. Specifically, the following methods were evaluated, with their respective architectures represented in figure 2: U-Net [17], UNETR [20], AttUnet (Attention U-net) [21], TransAttUnet (Transformer-Attention Unet) [22].

- UNET was developed by Olaf Ronneberger et al [17], for the purpose of medical image segmentation. As a first part it has an encoder used to capture or context of the image. And a second part is the symmetric expansion path (decoder) that is used to enable precise localization using transposed convolutions and dedicated to semantic pixel/voxel prediction. In addition, skip connections are used to combine the outputs of the encoder and decoder at multiple resolutions, allowing the recovery of lost spatial information.

- UNETR Transformer-based architecture utilizes a pure transformer as the encoder to learn sequence representations of the input volume, effectively capturing the global multi-scale information. As in Unet presents via skip connections between encoder and decoder at different resolutions to compute the final semantic segmentation output [20].

- Attention U-net was created in the context of image segmentation to be able during training to highlight only the relevant activation's and somehow reduce the processing power to be spent on irrelevant activation's in a way that gives the network a better generalization power, so that the network can pay "Attention" to certain parts of the image.Applied to the U-Net architecture this counteracts this problem of inaccuracy in up-sampling by creating jump links. But this brings redundant low-level feature extractions, since the feature representation is poor in the initial layers. The attention implemented in the jump links will actively suppress activation's in irrelevant regions, reducing the number of redundant features brought [21].

- TransAttUnet is the integration of the advantages of multi-level guided attainment and multi-scale skip connections into the standard U-net in order to improve performance in the implementation for medical image segmentation. With this approach, we can model the contextual semantic information and global relations, which can guarantee us the consistency of feature representations and semantic embeddings, reducing the noise perturbation, but also allow the model to mitigate the loss of fine details caused by direct oversampling with large scales [22].

B. Dataset

To be able to perform the network comparison study, it was then necessary to create a database for the respective LAA training one for the respective validation and test. To generate the database it was exported the available in-house databases, namely the 3D US database available to validate some previous 3D segmentation software of the team, as described in Morais et al [23]. Thus, to obtain the 2D images, multiple slices of 3D images aligned with the major axis of the LAA were extracted. All cases were manually reviewed, to confirm the LAA view and segmentation results. Unsatisfactory cases were removed. The final dataset has a total of 188 2D slices, which were then divided into 140 were divided for training, 28 for testing and finally 20 for validation. All 2D images were pre-processed by interpolating them to have the same pixel spacing (0.5 mm) and resolution (256*256).

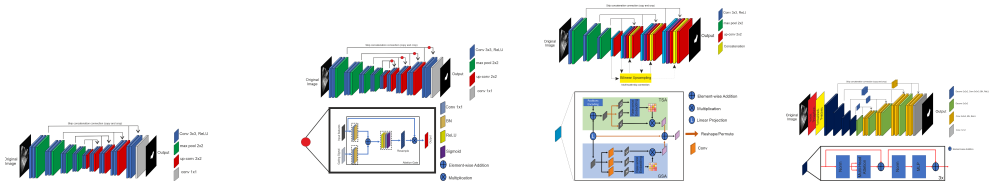
C. Augmentation

Augmentation techniques were used during the training of all networks to increase and overcome overfitting problems. The data augmentation used were: Rotation with a probability of 0.5, Scale Intensity with a probability of 0.5 and a range of -0.5 to 0.5, Zoom with a probability of 0.5 and with a minimum zoom factor of 1 and a maximum of 1.3, Gaussian Noise with a probability of 0.5 using a mean of 0 as the center of distribution and a standard deviation of 0.05, and finally distortion with a probability of 0.5 and a distortion range between (-0.2, 0.2).

D. Loss functions

A weight-balanced loss function was adopted by combining the cross-entropy and Dice loss functions as we see in eq.1 were W_{DC} is (weight for Dice loss function) L_{DC} is (Dice loss function),W_{CE} (weight for Cross Entropy function) and L_{CE} (Cross-Entropy loss function) the weights given to each loss can be changed to a more favorable combination for the best possible result.

$$Loss = (W_{DC} * L_{DC}) + (W_{CE} * L_{CE}) \quad (1)$$



(a) (b) (c) (d)

Fig. 2. Neural network architecture: a) Unet b) UnetR c) AttUnet d) TransAttUnet

In eq.2 we can see the equation referring to the Dice loss function and in eq.3 and eq.4 the corresponding Cross Entropy loss function

$$L_{DC} = 1 - 2 * \frac{|X| \cap |Y|}{|X+Y|} \tag{2}$$

$$L_{CE}(x, y) = \sum_{n=1}^{N} \frac{1}{\sum_{n=1}^{N} W_{y_n} * 1} * l_n \tag{3}$$

$$L_n(x, y) = -W_{y_n} * log \frac{exp(x_{n,y_n})}{\sum_{c=1}^{C} * exp(x_{n,y_n})} \tag{4}$$

E. Evaluations Metrics

Seven widely used segmentation metrics are used to assess the performance of the models, namely the Dice coefficient (DC), accuracy (Acc), recall (Rec), specificity (Spe), precision (Pre), Hausdorff distance (HD) and average distance error (Assd). DC measures the similarity between the segmentation result and the ground truth. Acc, Rec, Spe, and Pre compute pixel-wise classification accuracy as described in detail in [24]. Hausdorff distance is the maximum distance of a set to the nearest point in the other set and it is based on the calculation of the 95th percentile of the distances to eliminate the impact of a very small subset of the outliers. Averssd computes the average Euclidean distance between both surfaces, using each one as a reference and then averaging (to remove bias).

III. IMPLEMENTATION DETAILS

A. Train Procedure

During the training of all networks, Adam was used as an optimizer with a learning rate of 1e-4 and a batch size of 10. The number of epochs varied from net to net because they were trained until the values stabilized. A weight-balanced loss function was adopted by combining the cross-entropy and Dice loss functions. To ensure that the previous results were the best in each neural network, we made sure that the number of epochs were enough for the loss curve to stabilize and the validation of the curve to be the best possible score. The architectures were implemented in Python 3.8.10 and Pytorch 1.7.0 and run using a GeForce GTX 1070 and Ubuntu 18.04, running the system, CUDA11.0 and cuDNN8.0 and 8GB of GPU memory.

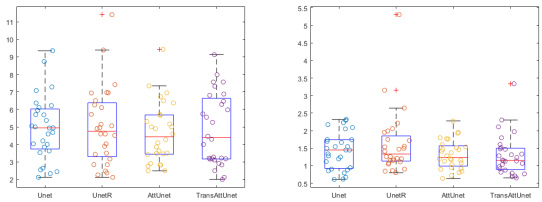
IV. RESULTS

The test images were input into the models to obtain the predicted labels. With the obtained results, it was compared the performance of each tested method. Table I summarizes the average values of each metric of the respective network and the respective standard deviation. Analyzing table I we can see that TransAttUnet achieved the best results in DC metrics with 89.28%, Acc with 99.09%, Recall with 93.09% and Assd with 1.30. Unet and AttUnet, on the other hand, outperformed this, in precision with 91.09% and HD95 with 4.66mm respectively. In this study, and in comparison with related works for this anatomy, we emphasized the results on the HD95 and Assd metrics, so in figure 3 we presented the corresponding boxplots, respectively. As a more visual and clear way to compare the performance of networks, we have the figure 4 that shows us the result of the superposition of the ground truth with the prediction of the respective network and the respective dice score. This eases the perception of each networks performance, allowing a qualitative comparison among them. It is possible to see in fig 4 that TransAttUnet had the best results in almost all cases.

V. DISCUSSION

In a generic way, the obtained results demonstrated that the current state-of-the-art networks achieved satisfactory results for this segmentation task, which may allow, in real-time, the enhancement of the LAA segmentation into the US image throughout the intervention, ultimately easing the identification of the relevant structure, as well as, simplifying the extraction of relevant indexes.

In this study, Unet was the most generic method under study, showed acceptable results and e also able to conclude that by making some changes, as in the cases of UNETR, AttUnet, and TransAttUnet, higher performances can be achieved. However, TransAttUnet stood out the most in the results due to multi-level guided attention that highlight only relevant activation's, reducing the influence of redundant low-level features obtained during the extraction stage. In fact, this attention feature is critical in this stage due to typical noise level observed in US, which hard its automatic analysis. Moreover, another advantage of the proposed network is the use of multi-scale skip connections that flexibly aggregates contextual feature maps from decoders of varying



(a) (b)

Fig. 3. diagram and the respective scatter of points of each network. a) Hausdorff95 values b) Averssd values.

TABLE I

QUANTITATIVE COMPARISON OF THE DIFFERENT METHODS

Network	Evaluation Metrics						
	DC	Acc	Rec	Spe	Pre	HD95	Assd
Unet	88.3±4.21	98.96±0.44	86.46±7.65	99.28±0.27	**91.0±5.69**	4.94±5.69	1.42±0.52
UnetR	87.6±4.39	98.94±0.46	89.78±6.99	99.22±0.44	86.22±6.94	5.01±2.21	1.61±0.91
AttUnet	88.26±3.49	99.00±0.45	89.78±6.21	**99.45±0.37**	88.25±7.30	**4.66±1.71**	1.31±0.41
TransAttUnet	**89.2±4.37**	**99.09±0.46**	**93.09±5.93**	99.37±0.42	86.30±7.79	4.85±2.05	**1.30±0.58**

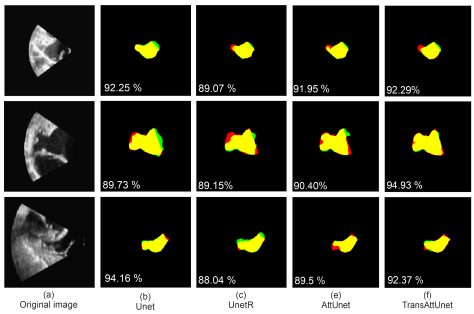


Fig. 4. Results (good, median and bad results according to Dice value) of the different methods trained, in the first column the original image and the remaining columns are the resulting masks for each network, the yellow color represents the intersection of the predict with the ground truth, the green color represents the predict of the network and red the ground truth

semantic scales to generate discriminative feature representations. TransAttUnet, had the second best performance on HD95, and the best in AVSSD.

As future work, it is intended to go on to explore a bit more the importance of different training modules, namely loss functions, optimizers, data augmentation, and adaptive learning rate strategies. In addition, comparison of the performance of deep learning strategies to create indexes for sizing LAA occlusion devices is also envisioned.

VI. Conclusion

We proposed a comparative analysis of four deep learning methods for the segmentation of the LAA in 2D TEE. In summary, all networks demonstrated their potential to be used for the LAA segmentation task. However, the network that showed the best results was TransAttUnet, slightly outperforming the remaining networks. However, additional experiments should be conducted to improve the comparison of the methods, namely by studying different loss functions and Hyper-parameters.

References

[1] Roy Beigel et al. "The left atrial appendage: anatomy, function, and noninvasive evaluation". In: *JACC: Cardiovascular imaging* 7.12 (2014), pp. 1251–1265.

[2] Nina C Wunderlich et al. "Percutaneous interventions for left atrial appendage exclusion: options, assessment, and imaging using 2D and 3D echocardiography". In: *JACC: Cardiovascular Imaging* 8.4 (2015), pp. 472–488.

[3] Indranill Basu Ray et al. "Meta-analysis comparing WatchmanTM and Amplatzer devices for stroke prevention in atrial fibrillation". In: *Frontiers in cardiovascular medicine* 7 (2020), p. 89.

[4] Pedro Morais et al. "A novel interventional guidance framework for transseptal puncture in left atrial interventions". In: *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation*. Springer, 2018, pp. 93–101.

[5] Jacqueline Saw et al. "Comparing measurements of CT angiography, TEE, and fluoroscopy of the left atrial appendage for percutaneous closure". In: *Journal of cardiovascular electrophysiology* 27.4 (2016), pp. 414–422.

[6] José Angel Cabrera, Farhood Saremi, and Damián Sánchez-Quintana. "Left atrial appendage: anatomy and imaging landmarks pertinent to percutaneous transcatheter occlusion". In: *Heart* 100.20 (2014), pp. 1636–1650.

[7] Orly Goitein et al. "Cardiac CT Angiography (CCTA) predicts left atrial appendage occluder device size and procedure outcome". In: *The international journal of cardiovascular imaging* 33.5 (2017), pp. 739–747.

[8] Jacqueline Saw et al. "Comparing measurements of CT angiography, TEE, and fluoroscopy of the left atrial appendage for percutaneous closure". In: *Journal of cardiovascular electrophysiology* 27.4 (2016), pp. 414–422.

[9] Pedro Morais et al. "Fast segmentation of the left atrial appendage in 3-D transesophageal echocardiographic images". In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 65.12 (2018), pp. 2332–2342.

[10] Wenjuan Bai et al. "Assessment of the left atrial appendage structure and morphology: comparison of real-time three-dimensional transesophageal echocardiography and computed tomography". In: *The international journal of cardiovascular imaging* 33.5 (2017), pp. 623–633.

[11] Hongning Song et al. "Morphologic assessment of the left atrial appendage in patients with atrial fibrillation by gray values–inverted volume-rendered imaging of three-dimensional transesophageal echocardiography: A comparative study with computed tomography". In: *Journal of the American Society of Echocardiography* 29.11 (2016), pp. 1100–1108.

[12] Pol Grasland-Mongrain, Jochen Peters, and Olivier Ecabert. "Combination of shape-constrained and inflation deformable models with application to the segmentation of the left atrial appendage". In: *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE. 2010, pp. 428–431.

[13] Pedro Morais et al. "3D segmentation of the left atrial appendage in computed tomography for planning of transcatheter occlusion". In: *Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling*. Ed. by Cristian A. Linte and Jeffrey H. Siewerdsen. Vol. 12034. International Society for Optics and Photonics. SPIE, 2022, 120340Q. DOI: 10.1117/12.2610705. URL: https://doi.org/10.1117/12.2610705.

[14] Cheng Jin et al. "Left atrial appendage segmentation using fully convolutional neural networks and modified three-dimensional conditional random fields". In: *IEEE journal of biomedical and health informatics* 22.6 (2018), pp. 1906–1916.

[15] Hrvoje Leventić et al. "Left atrial appendage segmentation from 3D CCTA images for occluder placement procedure". In: *Computers in biology and medicine* 104 (2019), pp. 163–174.

[16] Lei Wang et al. "Left atrial appendage segmentation based on ranking 2-D segmentation proposals". In: *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges: 7th International Workshop, STACOM 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 7*. Springer. 2017, pp. 21–29.

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

[18] Pedro Morais et al. "Assessment of LAA Strain and Thrombus Mobility and Its Impact on Thrombus Resolution—Added-Value of a Novel Echocardiographic Thrombus Tracking Method". In: *Cardiovascular Engineering and Technology* 13 (6 Dec. 2022), pp. 950–960. ISSN: 18694098. DOI: 10.1007/S13239-022-00629-Z/TABLES/5. URL: https://link.springer.com/article/10.1007/s13239-022-00629-z.

[19] Pedro Morais et al. "Semiautomatic Estimation of Device Size for Left Atrial Appendage Occlusion in 3-D TEE Images". In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 66.5 (2019), pp. 922–929. DOI: 10.1109/TUFFC.2019.2903886.

[20] Ali Hatamizadeh et al. "Unetr: Transformers for 3d medical image segmentation". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2022, pp. 574–584.

[21] Ozan Oktay et al. "Attention u-net: Learning where to look for the pancreas". In: *arXiv preprint arXiv:1804.03999* (2018).

[22] Bingzhi Chen et al. "Transattunet: Multi-level attention-guided u-net with transformer for medical image segmentation". In: *arXiv preprint arXiv:2107.05274* (2021).

[23] Pedro Morais et al. "Feasibility and Accuracy of Automated Three-Dimensional Echocardiographic Analysis of Left Atrial Appendage for Transcatheter Closure". In: *Journal of the American Society of Echocardiography* 35.1 (2022), pp. 124–133.

[24] Abdel Aziz Taha and Allan Hanbury. "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool". In: *BMC medical imaging* 15.1 (2015), pp. 1–28.