







TESSLA: Two-Stage Ensemble Scar Segmentation for the Left Atrium

Shaheim Ogbomo-Harmitt^(✉) , Jakub Grzelak, Ahmed Qureshi ,
Andrew P. King , and Oleg Aslanidi 

School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK
shaheim.ogbomo-harmitt@kcl.ac.uk

Abstract. Atrial fibrillation (AF) is the most common cardiac arrhythmia worldwide; however, the current success rates for catheter ablation (CA) therapy, the first-line treatment for AF, are suboptimal. Therefore, extensive research has focused on the relationship between scar tissue in the left atrium (LA) and AF, and its application for patient stratification and more effective CA therapy strategies. However, quantifying and segmenting LA scar tissue requires significant data pre-processing from well-trained clinicians. Hence, deep learning (DL) has been proposed to automatically segment the LA fibrotic scar from late gadolinium-enhanced cardiac magnetic resonance (LGE-CMR) images. Segmenting LA scar with DL is challenging as fibrosis from LGE-CMR images has a relatively small volume and regions surrounding the scar are also enhanced. Therefore, we propose a two-stage ensemble DL model (TESSLA: two-stage ensemble scar segmentation for the LA) that segments the blood pool of the LA, estimates the LA wall, applies an image intensity ratio with Z-score normalisation and combines a scar segmentation from two independent networks. TESSLA outperformed its constituent models and achieved state-of-art accuracy on the LAScar 2022 challenge evaluation platform for LA scar segmentation with a Dice score of 0.63 ± 0.14 and a Dice score of 0.58 ± 0.11 for the final test phase. Our workflow provides a fully automatic estimation of LA fibrosis from clinical LGE CMR scans.

Keywords: Left Atrial Scar · Deep Learning · Segmentation · Atrial Fibrillation

1 Introduction

Atrial fibrillation (AF) is the most common type of cardiac arrhythmia, affecting more than 46 million people worldwide. It is characterised by rapid and irregular electrical activations of the atrial chambers, resulting in reduced cardiac output [1, 2]. AF is not directly lethal, but it is associated with an increased risk of stroke and heart failure if sustained for long periods (known as persistent AF). In a study investigating the risk factors of cardiovascular diseases with approximately 5000 participants, AF was found to increase the risk of stroke by 1.5% for ages 50–59 years old and 23.5% for ages 80–89 years old [3]. The first-line treatment for AF is catheter ablation (CA) therapy. CA involves using a catheter to ablate (isolate or destroy) arrhythmogenic atrial tissue that

harbours AF triggers, thus restoring sinus rhythm and potentially the heart’s biomechanical function [4]. However, when treating persistent AF, CA therapy has a suboptimal AF reoccurrence rate of up to $\sim 70\%$ post-intervention [5, 6].

Extensive research has focused on the relationship between scar tissue in the left atrium (LA) and AF, and its application for patient stratification and more effective CA therapy strategies, including the DECAAF I and II clinical trial that rely on preprocessed late gadolinium enhanced (LGE) cardiac magnetic resonance (CMR) images [7, 8]. The preprocessing of LGE-CMR images involves manual segmentation of the left atrium (LA) and scar tissue by a well-trained clinician. This creates a bottleneck for routine clinical usage of LA scar quantification and segmentation as it is time-consuming and requires specialists to perform it.

The standard approach for LA scar segmentation employs thresholding techniques, which involves evaluating a threshold value based on a fixed number of standard deviations above the average intensity value of the LA wall or blood pool. However, the selection of threshold values is subjective, and the values can be affected by several factors such as scanner variability, acquisition timing after gadolinium administration and whether the LGE-CMR image is pre-or post-ablation [9]. Therefore, to automatically segment LA and scar tissue from LGE-CMR images, deep learning (DL) has been proposed as an efficient and accurate solution. DL was first applied to LA scar segmentation by Yang et al., who used super-pixel over-segmentation for feature extraction and stacked sparse auto-encoders [10]. Meanwhile, other studies adopted a DL model for simultaneous segmentation of the LA wall and LA scar [9]. However, the primary/key issues when segmenting LA scars from LGE-CMR images are its relatively small volume and enhanced regions of intensity surrounding it – creating noise during segmentation. Li et al. addressed this issue by utilising the spatial relationship of the LA and its scar to jointly segment both using an attention mask on the predicted scar probability map for shape attention [11]. Following on from utilising the spatial relationship of the LA and its scar, we propose an ensemble two-stage DL network (TESSLA: two-stage ensemble scar segmentation for the LA). In addition, to overcome the limitation of generalisability of the model developed by Li et al., we propose using an intensity ratio (IIR) normalisation, applied by traditional methods to reduce inter-patient and scanner effects [12, 13]. To summarise, the contributions of this work are:

- First method to feature a two-stage DL model to segment blood pool, estimate LA wall and predict scar segmentation from LA wall and original LGE-CMR images.
- First DL model to implement IIR normalisation for LA scar segmentation.

2 Methods

2.1 Dataset

The dataset was provided by the Left Atrial and Scar Quantification & Segmentation Challenge 2022, which includes 60 LGE-CMR images from patients post- and pre-ablation with corresponding LA blood pool and scar segmentation masks. The images were collected across three centres (University of Utah, Beth Israel Deaconess Medical Center and King’s College London). The spatial resolution of the LGE-CMR images

was either $1.25 \times 1.25 \times 2.5 \text{ mm}^3$ (University of Utah), $1.4 \times 1.4 \times 1.4 \text{ mm}^3$ (Beth Israel Deaconess Medical Center) or $1.3 \times 1.3 \times 4.0 \text{ mm}^3$ (King's College London) [9, 11, 14].

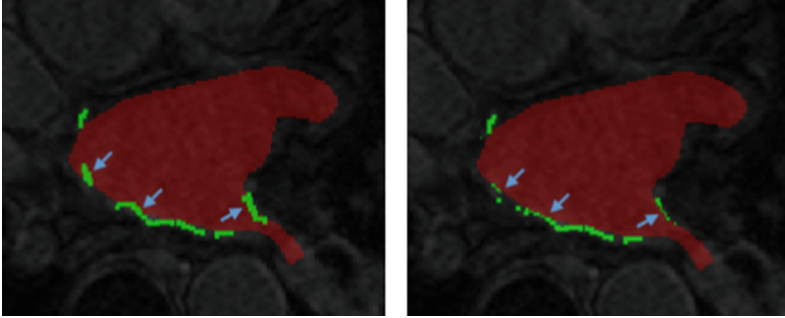


Fig. 1. Ground truth segmentation. Left: scar segmentation (green) and LA blood pool (red) ground truth masks. Right: Same masks with scar projected onto closest point on endocardial wall. Scar in blood pool and subsequent projections are highlighted with arrows (blue).

Due to 20–35% of the voxels of the scar segmentation masks being within the blood pool segmentation masks, prior to being used in model training, scar voxels within the blood pool were projected to the LA surface along the normal direction of the closest LA endocardial wall voxel (see Fig. 1).

2.2 Proposed Model and Implementation

The proposed model consists of three 3D nn-UNets [15], which form two distinct paths to predict scar. In the first path (A in Fig. 2) a nn-UNet predicts the segmentation mask of the LA blood pool from the LGE-CMR image, and a second nn-UNet predicts the LA scar from the LGE-CMR image and IIR normalised LA wall (derived from the blood pool prediction and LGE-CMR image). In the second path (B in Fig. 2) a nn-UNet predicts LA scar directly from the LGE-CMR image (see Fig. 2).

The final LA scar mask is formed by the union of the predictions from the two paths (post-softmax probability thresholding and nn-UNet postprocessing), to provide a final ensemble prediction of LA scar (Fig. 3). nn-UNet was chosen for the segmentation as it automatically configures the optimal U-Net architecture, hyperparameters and image preprocessing and postprocessing steps and has demonstrated state-of-the-art performance in a range of segmentation challenges [15]. Data augmentation was performed during training and included techniques such as rotations, scaling, Gaussian noise, blur, brightness, contrast, low-resolution simulation, gamma correction and mirroring [15].

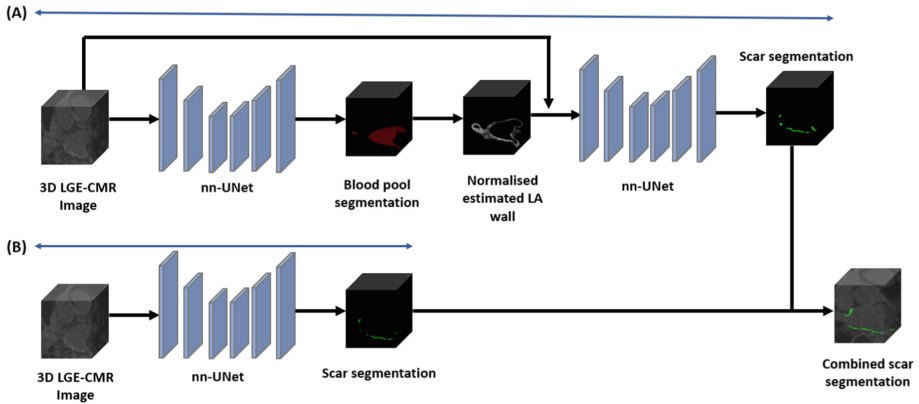


Fig. 2. Diagram of the proposed TESSLA model, outlining each stage of the model: blood pool segmentation, LA wall estimation and normalisation, LA scar segmentation from both nn-UNets and combined scar segmentation. **A)** Represents first path of TESSLA which predicts LA scar from normalised LA wall and LGE-CMR image. **B)** Represents second path of TESSLA which predicts LA scar from LGE-CMR image.

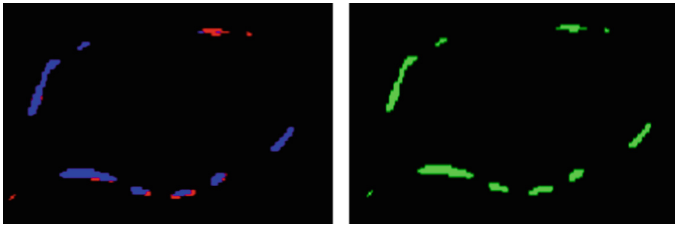


Fig. 3. Example of how the two different scar predictions (colours red and blue representing each scar prediction) (left image) are combined for the final prediction (right image).

2.3 LA Wall Estimation and Normalisation

The LA wall was estimated by obtaining the predicted blood pool segmentation mask to evaluate the boundary of the blood pool/LA endocardial wall and then dilating the wall boundary outwards (in 3D) by 3 voxels (each LGE-CMR image had a resolution of ~ 1 mm). A 3 voxel dilation amount was chosen based on the CT and MRI studies that found LA wall mean thickness was between 2 and 3 mm (Fig. 4) [16, 17].

The outward dilation was achieved by first dilating and eroding the blood pool boundary. Then, as shown in Eq. 1 below, the LA wall segmentation mask (M_{wall}) was found by the matrix subtracting the blood pool segmentation mask (B) and the eroded boundary (B_E) from the dilated boundary (B_D). The subtraction of B_D ensured that no dilation was within the LA blood pool. Lastly, if a voxel's LGE-CMR image intensity was less than 1, it was set to 0. The voxel intensity was then normalised and thresholded, such that a voxel intensity of 1 corresponded to LA wall voxels and 0 corresponded to background voxels. This process took ~ 5 s to run for a single subject on a 12th Gen

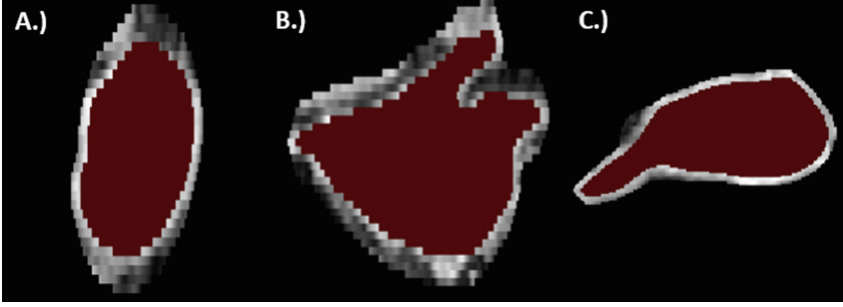


Fig. 4. Example of LA wall estimation with IIR and Z-score normalised LGE-CMR values with highlighted blood pool (red) from different anatomical views. **A)** LA wall from sagittal view. **B)** LA wall from coronal view. **C)** LA wall from axial view

Intel(R) Core (TM) i7-12700KF 3.61 GHz.

$$M_{wall} = B_D - B_E - B \quad (1)$$

An element-wise product, \odot , of the LA wall segmentation mask (M_{wall}) and LGE-CMR image (I) was applied to determine I_{wall} . The estimated LA wall region (LA_{wall}) was then defined to be all voxels where the intensities of image I_{wall} were above zero.

$$I_{wall} = I \odot M_{wall} \quad (2)$$

$$LA_{wall} = \{X | X \in I_{wall}, X > 0\} \quad (3)$$

$$I_{IIR} = \frac{LA_{wall}}{\tilde{X}_B} \quad (4)$$

IIR normalisation was then performed by dividing the LA_{wall} by, \tilde{X}_B , the mean blood pool voxel intensity (Eq. 4) [13]. Note that only the voxel intensities of the estimated LA wall from I are divided by \tilde{X}_B as they are the only non-zero intensity voxels. The motivation for LA wall IIR normalisation was to provide the second nn-UNet with a set of features which are partially homogenised across multiple inter-individual factors (interscan variability in intensities, contrast dose, the delay time of image acquisition after contrast injection, body mass index, hematocrit, and renal function) to assist in model generalisability [18].

$$I^* = \frac{I_{IIR} - \mu}{\sigma} \quad (5)$$

Using the rule-based preprocessing of the nn-UNet pipeline, all IIR LA wall (I_{IIR}) voxels were Z-score normalised to get I^* (Eq. 5, where the mean voxel intensity of I_{IIR} is μ and the standard deviation is σ) as better convergence can be achieved during backpropagation if the average of each input variable over the training set is close to zero. The latter quality has led Z-score normalisation to becoming a de-facto practice in computer vision [19].

3 Results

3.1 Model Implementation and Training

Each nn-UNet model was trained independently for their respective task (blood pool or LA scar segmentation). The proposed framework was trained using 48 LGE-CMR images and a further 12 such images were used as a validation set. Each nn-UNet was trained for 1000 epochs using a combined cross-entropy and Dice loss function, Stochastic gradient descent was used with Nesterov momentum ($\mu = 0.99$) with an initial learning rate of 0.01 on an NVIDIA RTX 48GB A6000 GPU; each nn-UNet took ~ 24 h to train [20–22].

3.2 Validation and Test Set Results

As reflected in Tables 1 and 2, TESSLA outperformed both of its constituent models, justifying the use of an ensemble prediction. Figure 5 further supports this justification, as it illustrates how the two different nn-UNets predict LA scars that are overlapping but also different which is reflected in the increase of Dice score and sensitivity. Therefore, combining the two scar segmentations can predict better LA scar coverage.

On the LAScar 2022 test phase (hold out test set of 24 LGE-CMR images), TESSLA had a LA scar Dice score of 0.581 ± 0.112 and sensitivity of 0.529 ± 0.145 .

Table 1. Validation set results for proposed model and constituent models.

Model	Scar		Blood
	Dice Score	Sensitivity	Dice Score
TESSLA	0.529 ± 0.070	0.531 ± 0.133	0.927 ± 0.020
nn-UNet (LGE-CMR)	0.506 ± 0.090	0.454 ± 0.134	N/A
nn-UNet (LGE-CMR + LA wall)	0.510 ± 0.080	0.474 ± 0.125	0.927 ± 0.020

Table 2. LAScarQ 2022 evaluation platform results from the hold-out test set (10 3D LGE-CMR) images from the challenge evaluation platform for the proposed model and constituent models.

Model	Scar		Blood Pool
	Dice Score	Sensitivity	Dice Score
TESSLA	0.634 ± 0.142	0.578 ± 0.162	0.890 ± 0.075
nn-UNet (LGE-CMR)	0.608 ± 0.149	0.524 ± 0.165	N/A
nn-UNet (LGE-CMR + LA wall)	0.593 ± 0.152	0.497 ± 0.152	0.890 ± 0.075

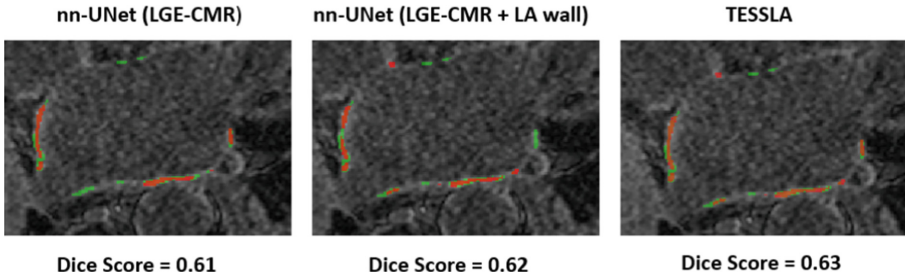


Fig. 5. Comparison of model prediction from a validation set subject with corresponding scar segmentation Dice score, where predicted scar is in red and ground truth is in green.

LGE-CMR image quality plays a vital role in accurate TESSLA LA scar prediction, as shown in Fig. 6 and 7, as images with non-prevalent fibrotic gadolinium binding or motion artefacts had worse Dice scores. Furthermore, this explains the increase in LA scar Dice score in the LAScar 2022 evaluation platform results compared to the validation set results in training, as the evaluation platform set of LGE-CMR images only had one image with poor quality while the validation set had three. Meanwhile, this also explains the decrease of scar Dice score on the test phase dataset compared to the evaluation platform results.

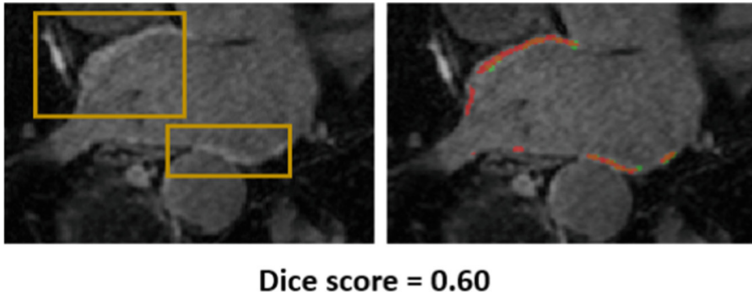


Fig. 6. Example of good quality LGE-CMR image with highlighted regions (gold boxes) of good gadolinium binding to fibrotic tissue (image on left) and corresponding Dice score. Predicted (red) and ground truth (green) scar highlighted on the image on the right.

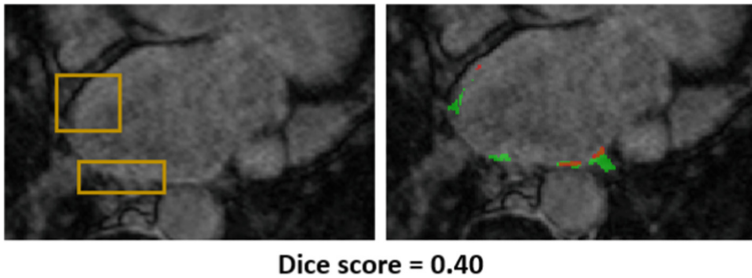


Fig. 7. Example of poor quality LGE-CMR image with highlighted regions of poor (gold boxes) gadolinium binding to fibrotic tissue (image on left) and corresponding Dice score. Predicted (red) and ground truth (green) scar highlighted on the image on the right.

4 Discussion and Conclusion

This study introduces a novel DL model, TESSLA, a two-stage ensemble model for LA scar segmentation. TESSLA achieved accuracy comparable to the state-of-the-art on the LAscar 2022 evaluation platform test set by utilising LA wall IIR-Z-score normalisation and nn-UNets. To effectively compare these results to the gold standard, a study of inter-observer variability in manual segmentation would be required. Moreover, the model results show that poor image quality significantly impacts its performance. Hence, future work should focus on implementing a method to reduce its effects for better robustness. The study also demonstrated that integrating an IIR-Z-score normalisation to the LA wall into an ensemble model can predict better LA scar coverage. Therefore, our proposed DL model provides better model generalisability for LA scar segmentation than a single model with LGE-CMR input.

The novel automatic tool for LA scar tissue quantification developed in this study can be applied in the clinic. Previous studies have suggested that DL LA scar segmentation tend to have higher Dice score accuracies than conventional thresholding-based methods [9]. Hence, DL models like TESSLA can be superior to the current gold standard of IIR thresholding, both in accuracy and speed of the LA scar assessment. This could lead to better patient stratification and AF treatment planning. Furthermore, this will also aid the clinical implementation of emerging technologies for AF management, such as digital twins, patient-specific models and AI therapy predictors [23–26].

Acknowledgements. This work was supported by funding from the Medical Research Council [MR/N013700/1], the British Heart Foundation [PG/15/8/31130], and the Wellcome/EPSRC Centre for Medical Engineering [WT 203148/Z/16/Z].

References

1. Hart, R.G., Halperin, J.L.: Atrial fibrillation and stroke: concepts and controversies. *Stroke* **32**, 803–808 (2001)
2. Chugh, S.S., et al.: Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. *Circulation* **129**, 837–847 (2014)
3. Wolf, P.A., Abbott, R.D., Kannel, W.B.: Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke* **22**, 983–988 (1991)
4. Amin, A., Houmsse, A., Ishola, A., Tyler, J., Houmsse, M.: The current approach of atrial fibrillation management. *Avicenna J. Med.* **6**, 8–16 (2016)
5. Brundel, B.J.J.M., Ai, X., Hills, M.T., Kuipers, M.F., Lip, G.Y.H., de Groot, N.M.S.: Atrial fibrillation. *Nat Rev Dis Primers.* **8**, 21 (2022). <https://doi.org/10.1038/s41572-022-00347-9>
6. Karamichalakis, N., et al.: Managing atrial fibrillation in the very elderly patient: challenges and solutions. *Vasc. Health Risk Manag.* **11**, 555–562 (2015)
7. Marrouche, N.F., et al.: Association of atrial tissue fibrosis identified by delayed enhancement MRI and atrial fibrillation catheter ablation: the DECAAF study. *JAMA* **311**, 498–506 (2014)
8. Marrouche, N.F., et al.: DECAAF II Investigators: Efficacy of LGE-MRI-guided fibrosis ablation versus conventional catheter ablation of atrial fibrillation: The DECAAF II trial: Study design. *J. Cardiovasc. Electrophysiol.* **32**, 916–924 (2021)
9. Li, L., Zimmer, V.A., Schnabel, J.A., Zhuang, X.: Medical image analysis on left atrial LGE MRI for atrial fibrillation studies: A review. *Med. Image Anal.* **77**, 102360 (2022)

10. Yang, G., et al.: A fully automatic deep learning method for atrial scarring segmentation from late gadolinium-enhanced MRI images. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), pp. 844–848 (2017)
11. Li, L., Zimmer, V.A., Schnabel, J.A., Zhuang, X.: AtrialJSQnet: A New framework for joint segmentation and quantification of left atrium and scars incorporating spatial and shape information. *Med. Image Anal.* **76**, 102303 (2022)
12. Roy, A., Varela, M., Aslanidi, O.: Image-Based Computational Evaluation of the Effects of Atrial Wall Thickness and Fibrosis on Re-entrant Drivers for Atrial Fibrillation. *Front. Physiol.* **9**, 1352 (2018)
13. Roy, A., et al.: Identifying locations of re-entrant drivers from patient-specific distribution of fibrosis in the left atrium. *PLoS Comput. Biol.* **16**, e1008086 (2020)
14. Li, L., Zimmer, V.A., Schnabel, J.A., Zhuang, X.: AtrialGeneral: Domain Generalization for Left Atrial Segmentation of Multi-center LGE MRIs. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12906, pp. 557–566. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87231-1_54
15. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods.* **18**, 203–211 (2021)
16. Beinart, R., et al.: Left atrial wall thickness variability measured by CT scans in patients undergoing pulmonary vein isolation. *J. Cardiovasc. Electrophysiol.* **22**, 1232–1236 (2011)
17. Varela, M., et al.: Novel MRI Technique Enables Non-Invasive Measurement of Atrial Wall Thickness. *IEEE Trans. Med. Imaging.* **36**, 1607–1614 (2017)
18. Benito, E.M., et al.: Left atrial fibrosis quantification by late gadolinium-enhanced magnetic resonance: a new method to standardize the thresholds for reproducibility. *Europace* **19**, 1272–1279 (2017)
19. LeCun, Y., Bottou, L., Orr, G.B., Müller, K.-R.: Efficient BackProp. In: Orr, G.B., Müller, K.-R. (eds.) *Neural Networks: Tricks of the Trade*. LNCS, vol. 1524, pp. 9–50. Springer, Heidelberg (1998). https://doi.org/10.1007/3-540-49430-8_2
20. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. In: Cardoso, M.J., et al. (eds.) *DLMIA/ML-CDS -2017*. LNCS, vol. 10553, pp. 240–248. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_28
21. Sabuncu, Z.: Generalized cross entropy loss for training deep neural networks with noisy labels. *Adv. Neural Inf. Process. Syst.*, **31** (2018)
22. Paszke, A., et al: PyTorch: An imperative style, high-performance deep learning library. In: Wallach, H., Laroche, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. (eds.) *Adv. Neural Inf. Process. Syst.* Curran Associates, Inc. **32**, pp. 8026–8037 (2019)
23. Corral-Acero, J., et al.: Others: The Digital Twin to enable the vision of precision cardiology. *Eur. Heart J.* **41**, 4556–4564 (2020)
24. Muffoletto, M., et al.: Toward Patient-Specific Prediction of Ablation Strategies for Atrial Fibrillation Using Deep Learning. *Front. Physiol.* **12**, 674106 (2021). <https://doi.org/10.3389/fphys.2021.674106>
25. Roney, C.H., et al.: Predicting atrial fibrillation recurrence by combining population data and virtual cohorts of patient-specific Left atrial models. *Circ. Arrhythm. Electrophysiol.* **15**, e010253 (2022)
26. Muizniece, L., et al.: Reinforcement Learning to Improve Image-Guidance of Ablation Therapy for Atrial Fibrillation. *Front. Physiol.* **12**, 733139 (2021). <https://doi.org/10.3389/fphys.2021.733139>