Heliyon 9 (2023) e12945

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Deep learning-based automatic left atrial appendage filling defects assessment on cardiac computed tomography for clinical and subclinical atrial fibrillation patients

Ling Chen [a], Sung-Hao Huang [b,*], Tzu-Hsiang Wang [a], Tzuo-Yun Lan [a], Vincent S. Tseng [c], Hsuan-Ming Tsao [b,d], Hsueh-Han Wang [e], Gau-Jun Tang [a]

[a] Institute of Hospital and Health Care Administration, National Yang Ming Chiao Tung University, Taipei, Taiwan
[b] Division of Cardiology, Department of Internal Medicine, National Yang Ming Chiao Tung University Hospital, Yi-Lan, Taiwan
[c] Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu, Taiwan
[d] School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan
[e] Department of Radiology, National Yang Ming Chiao Tung University Hospital, Yi-Lan, Taiwan

ARTICLE INFO

Keywords:
Left atrial appendage
Atrial fibrillation
Computed tomography
Artificial intelligence
Deep learning

ABSTRACT

Rationale and objectives: Selecting region of interest (ROI) for left atrial appendage (LAA) filling defects assessment can be time consuming and prone to subjectivity. This study aimed to develop and validate a novel artificial intelligence (AI), deep learning (DL) based framework for automatic filling defects assessment on CT images for clinical and subclinical atrial fibrillation (AF) patients.
Materials and methods: A total of 443,053 CT images were used for DL model development and testing. Images were analyzed by the AI framework and expert cardiologists/radiologists. The LAA segmentation performance was evaluated using Dice coefficient. The agreement between manual and automatic LAA ROI selections was evaluated using intraclass correlation coefficient (ICC) analysis. Receiver operating characteristic (ROC) curve analysis was used to assess filling defects based on the computed LAA to ascending aorta Hounsfield unit (HU) ratios.
Results: A total of 210 patients (Group 1: subclinical AF, n = 105; Group 2: clinical AF with stroke, n = 35; Group 3: AF for catheter ablation, n = 70) were enrolled. The LAA volume segmentation achieved 0.931–0.945 Dice scores. The LAA ROI selection demonstrated excellent agreement (ICC $\geq$0.895, $p < 0.001$) with manual selection on the test sets. The automatic framework achieved an excellent AUC score of 0.979 in filling defects assessment. The ROC-derived optimal HU ratio threshold for filling defects detection was 0.561.
Conclusion: The novel AI-based framework could accurately segment the LAA region and select ROIs while effectively avoiding trabeculae for filling defects assessment, achieving close-to-expert performance. This technique may help preemptively detect the potential thromboembolic risk for AF patients.

Abbreviations: AA, Ascending aorta; AF, Atrial fibrillation; AI, Artificial intelligence; AUC, Area under the ROC curve; CI, Confidence interval; DL, Deep learning; ECG, Electrocardiogram; HU, Hounsfield unit; ICC, Intraclass correlation coefficient; LAA, Left atrial appendage; ROC, Receiver operating characteristics; ROI, Region of interest; SD, Standard deviation.

* Corresponding author.
E-mail address: h8368@ms8.hinet.net (S.-H. Huang).

https://doi.org/10.1016/j.heliyon.2023.e12945
Received 8 December 2022; Received in revised form 4 January 2023; Accepted 10 January 2023
Available online 13 January 2023

1. Introduction

The left atrial appendage (LAA) has been identified as a major source of intracardiac thrombosis formation in patients with atrial fibrillation (AF) [1]. Transesophageal echocardiography (TEE) is the current reference standard for evaluating LAA in patients with AF associated embolic events [2], while cardiac computed tomography (CT) has become an emerging accurate imaging modality for the assessment of LAA filling defects [3,4].

Filling defects can be identified on contrast-enhanced cardiac CT as low attenuation areas in the LAA, caused by thrombus or incomplete contrast agents mixing with blood, rather than cardiac structures such as trabeculae or motion artifacts [5]. Studies have used the radiodensity ratio of Hounsfield units (HU) in a selected region of interest (ROI) of the LAA to those in the ascending aorta (AA) to identify LAA thrombus and circulatory stasis in cardiac CT [6,7] and differentiate between them [4,8,9] using TEE as the reference. However, even with the help of commercially available semi-automatic software, the process of selecting an ROI in the LAA for potential filling defects assessment is highly subjective and time consuming. Particularly, the presence of endocardial trabeculae in the LAA may sometimes give the endocardium a "fuzzy" appearance and complicate the task, as shown in Fig. 1 (D). Therefore, automatic ROI suggestion in the LAA may meet clinical demands in clinical practice to reduce efforts in LAA assessment and minimize subjectivity.

Recently, deep learning (DL) demonstrated promising performance in the automated analysis of cardiovascular images [10,11]. Particularly, several DL-based models were developed to identify LAA segments in cardiac CT [12131415]. Machine learning- or DL-based models were also developed to predict LAA thrombotic risk [14,16]. However, none of the existing methods targeted at automatically selecting a representative ROI in the LAA and computing LAA/AA HU ratio for filling defects assessment. In this study, we developed and evaluated an artificial intelligence (AI) based framework to automatically select LAA ROI and assess LAA filling effects on cardiac CT images for clinical and subclinical AF patients.

2. Materials and Methods

2.1. Study design, participants, and scanning parameters

The protocol of this retrospective study was approved by the local Institutional Review Board and the need for informed consent was waived. We retrospectively collected a total of 210 patients in this study, containing three groups of patients with cardiac arrhythmia commonly observed in clinical practice:

Group 1 contained 105 patients with atrial high-rate episodes (AHRE), referred to as subclinical AF (SCAF), detected by cardiac implantable electronic devices (CIEDs). CIED interviews were conducted every 3–6 months to retrieve recorded information. The definition of AHRE considers an atrial rate >170 bpm as the default setting. The data was collected between 2016 and 2020 at NYCUH, YiLan, Taiwan, using an ECG-gated, 64-slice CT scanner (Somatom Definition Flash, Siemens). Axial images were reconstructed to a

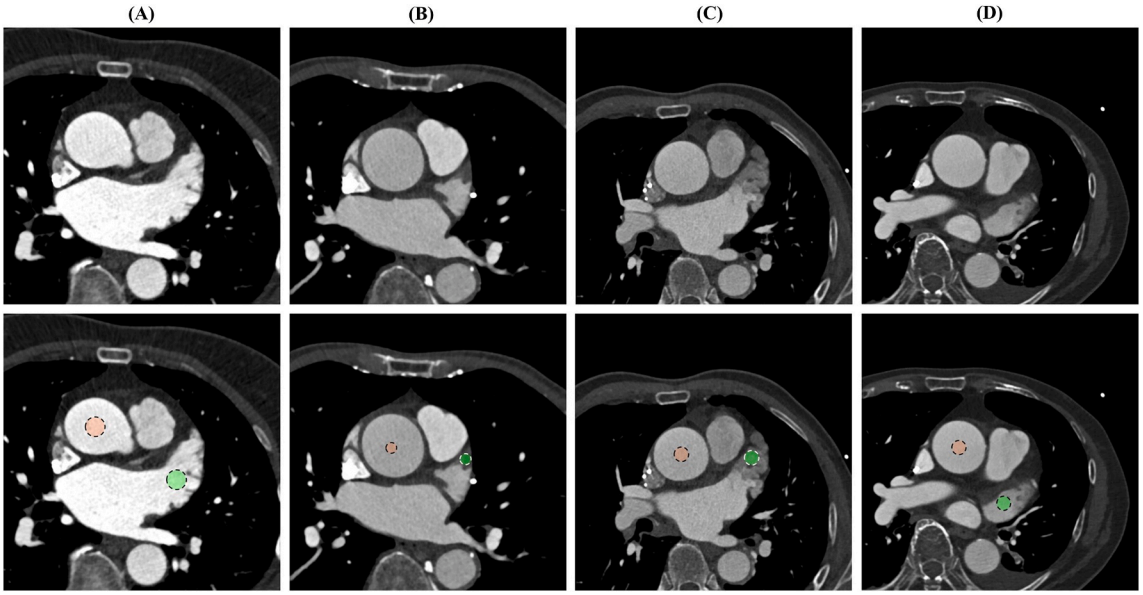

Fig. 1. Examples of the automatically selected LAA ROI (green) and the corresponding AA ROI from the proposed framework. (A) presents a simple clear LAA case, (B) a simple LAA filling defects case, (C) a LAA filling defects case with the presence of trabecula, and (D) a more difficult case with fuzzy-like trabeculae and no filling defect.

† AA, ascending aorta; LAA, left atrial appendage; ROI, region of interest. . (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

512 × 512 matrix using a slice thickness of 3 mm in multiple phases covering a cardiac cycle in the increments of 10% of ECG R-R interval.

Group 2 consisted of 35 patients with clinical AF-related stroke, collected at NYCUH, between 2011 and 2016, using an ECG-gated, 64-slice CT scanner (Brilliance CT, 64-slice, Philips). Axial images were reconstructed to a 512 × 512 matrix using a slice thickness of 0.9 mm in multiple phases covering a cardiac cycle in the increments of 10% of ECG R-R interval.

Group 3 contained 70 clinical AF patients preparing for catheter ablation, collected at Taipei Veterans General Hospital, Taiwan, between 2005 and 2011, using ECG-gated scanners (Sensation 16, Siemens; Brilliance CT, 40-slice, Philips; and Aquilion 64 CFX, Toshiba). Axial images were reconstructed to a 512 × 512 matrix using a slice thickness of 2 mm for the Toshiba scanner, 1 mm for the Siemens scanner, and 0.9 mm for the Philips scanner.

All the cardiac CT scans were performed in the craniocaudal direction during a single breath hold under sinus or pacing rhythm. Delayed CT scanning was performed with the second injection of contrast agent, 180 s after the injection of the first bolus of contrast agent. No β-blockers were used in any of the enrollees for heart rate regulation, because CT was performed to evaluate the intracardiac structure and not the coronary arteries.

The flowchart of the study protocol is displayed in Fig. 2. All data that contained stasis or thrombus in the LAA was identified by experienced cardiologists/radiologists. The framework was developed and tested on the Group 1 subset, and further evaluated on the Group 2 and Group 3 subsets. The Group 1 subset was first randomly divided into a development set (80%) and a test set (20%), with the development set further divided into a training set (80%) and a validation set (20%).

The patients' medical comorbidities, CHA_2DS_2-VASc scores, and medication records were obtained. Patients who had single chamber devices, aged ≥90 years, advanced chronic kidney disease (≥stage 3b) or allergy to contrast medium were excluded.

2.2. Image annotation and preprocessing

We used the Group 1 subset for model development and testing. The LAA volume and myocardial mass were annotated by an expert reader in the cardiovascular department at NYCUH. A semi-automatic region-growing-based software developed by Chung Yuan Christian University was used to annotate reconstructed CT images at phases of 0%, 40% and 80% within the R-R interval for the training of segmentation models. For the proposed framework of filling defect assessment, all the phases from 0% to 90% were used.

The cardiac CT images were saved in the Digital Imaging and Communications in Medicine (DICOM) format with a resolution of 512 × 512 pixels and adjusted according to the slope and intercept extracted from the DICOM attributes. To correct occasional CT metal artifacts, HU values smaller than −1024 were replaced with −1024 and those greater than 3071 were replaced with 3071. The CT images were first resized to 224 × 224 numeric arrays and normalized to values between 0 and 1 before being fitted to the models described below.

2.3. Development of deep learning segmentation models

An overview of our fully automated filling defects assessment framework is shown in Fig. 3. We first built separate DL-based networks for LAA segmentation and AA segmentation, respectively.

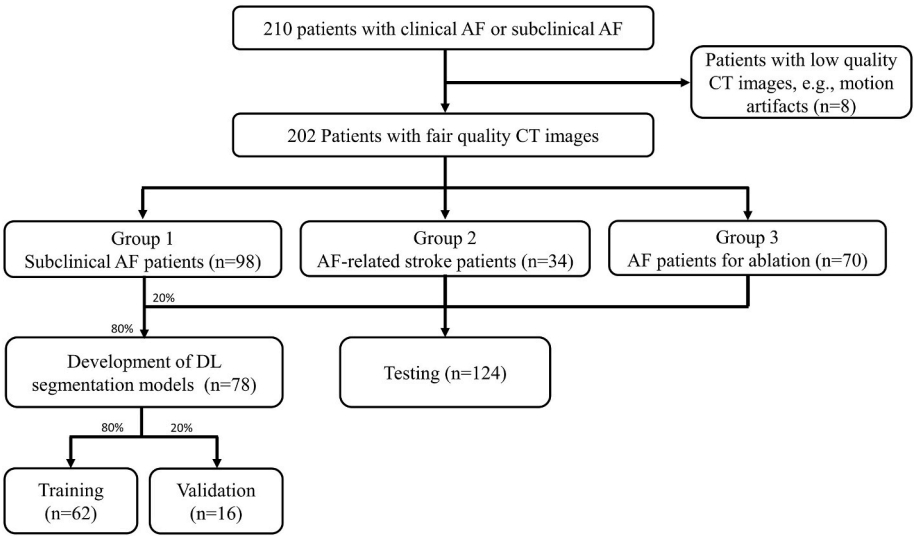

Fig. 2. The flowchart of the study protocol.
† AF, atrial fibrillation; CT, computed tomography; DL, deep learning.

2.4. LAA segmentation model

The LAA segmentation network adopted a U-Net based architecture [17]. To improve accuracy, we added group normalization (GN) at the end of each encoder and decoder block [18]. Furthermore, a negative example learning strategy was designed to randomly sample negative examples (i.e., slices containing no LAA region) into the training batches to avoid false positives, since the LA and LAA did not occur in every cardiac CT slice. Details about the loss functions used in the training, please see AppendixA.

2.5. AA segmentation model

The AA segmentation network had a U-Net encoder-decoder structure with a pre-trained VGG-16 encoder and was fine-tuned with the training data [19]. The AA annotations were first obtained using the Hough circle transform method from the Python OpenCV package to detect circles in images. Detected circles matching no AA region were manually screened out. An ROI at the center of the model segmented area was used later for computing the mean AA HU value as a reference for calculating the LAA/AA ratio [9]. Details about the loss functions used in the training, please see AppendixA.

In both model training processes, scheduled learning rates were applied with a starting rate of 0.001, and data augmentation was applied, including random cropping, rotation, contrast, and brightness. The models were implemented using PyTorch framework (www.pytorch.org) and trained for 200 epochs using a NVIDIA RTX 2080 Ti GPU.

2.6. Automatic ROIs selection for LAA

Selecting an ROI for the LAA filling defects assessment can be challenging. Fig. 1 (A-D) shows the examples of a range of different cases, including a clear LAA case, a simple LAA filling defects case, a LAA filling defects case with the presence of trabecula, and a more difficult case with fuzzy-like trabeculae and no filling defect, respectively. We proposed the methods below to automatically differentiate those cases and select an ROI for the LAA filling defects assessment.

2.7. Candidate Region Cutting for automatic phase and slice selection

We proposed a novel method called *Candidate Region Cutting* (CRC) to find candidate regions that potentially contain filling defects in the LAA, based on analyzing the fitted components of a two-component Gaussian mixture. The CRC method first fitted the model to the HU values in the segmented LAA region output from the segmentation model. Fig. 4 (A) shows the original CT and (B) displays the two extracted components. The component with a higher mean HU value and a sharper peak was called the *contrast distribution*, as contrast-filled areas often had higher and more concentrated HU values. The other component was the *candidate distribution*, which could potentially contain an area with filling defects. Since the two components often overlapped, 3 standard deviations (SDs) below the mean of the contrast distribution was set experimentally as the upper bound. To avoid extracting values from potential muscle tissue close to the LAA boundary, 1 SD below the mean of the candidate distribution was set experimentally as the lower bound. Experimental results for different settings please see Appendix C, Table C1. The output was a candidate region, between the upper and lower bounds, colored in purple in Fig. 4 (B), corresponding to the purple region in Fig. 4 (A) (Fig. 4).

A slice was a valid candidate, if the candidate region was large enough to allow a 0.3 cm^2 circular kernel to be fitted in for three consecutive slices. The final selected slice was the valid candidate slice with the largest candidate region amongst all phases and slices. However, if no candidate region was found, the phase with the highest mean HU in the AA was selected and a slice in the phase with the largest LAA area was chosen, since AA was the brightest when the contrast agent fully travelled through, presenting the blood pumping through the AA. In either case, a selected CT slice was ready for further ROI selection.

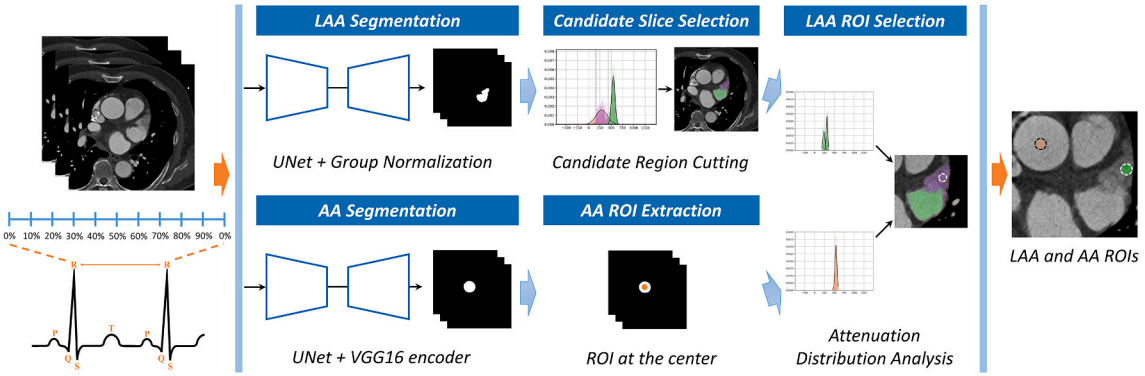

Fig. 3. Overview of the fully automated deep learning-based framework for LAA ROI selection and filling defects assessment. The flow diagrams show the process from LAA and AA segmentation to candidate slice selection, LAA ROI selection, and final obtaining of the LAA/AA HU ratio.
† AA, ascending aorta; HU, Hounsfield units; LAA, left atrial appendage; ROI, region of interest.

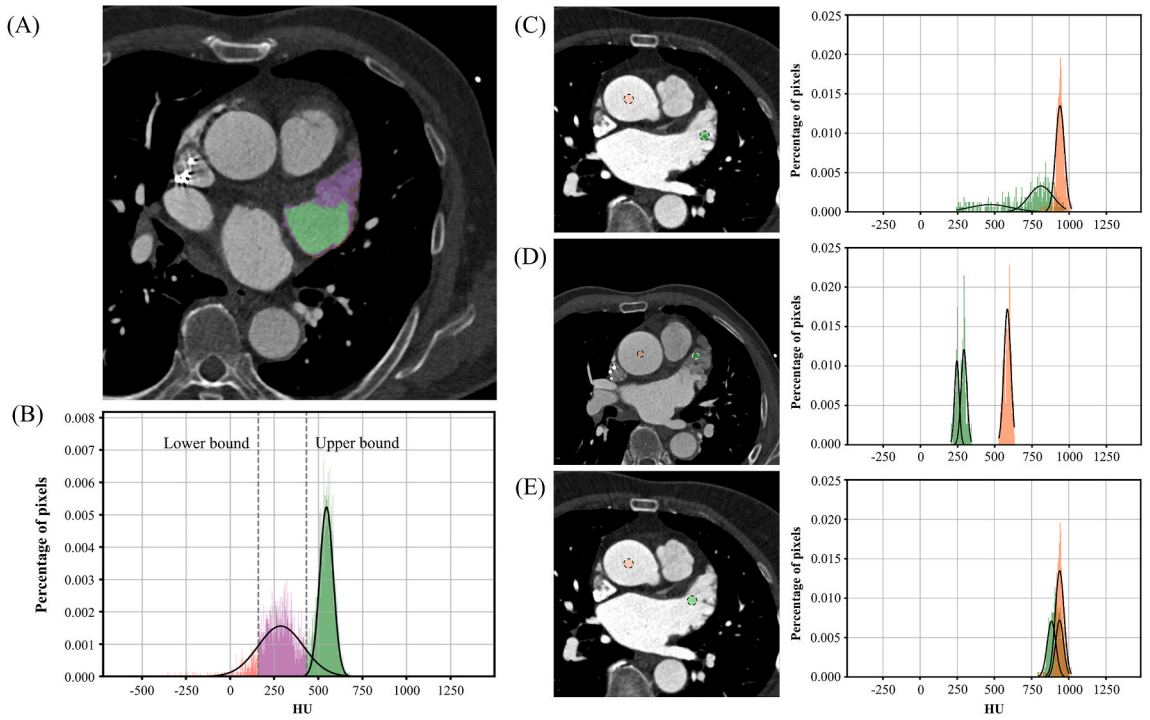

Fig. 4. The illustration of the proposed Candidate Region Cutting (CRC) method and Attenuation Distribution Analysis (ADA) method. (A) shows the original CT slice with the candidate region (purple) identified by the CRC method. (B) is the result of applying the CRC method to the HU values in the LAA region, where the candidate region (purple) was defined as the values between the upper and lower bounds based on the two fitted components. (C)–(E) show the ROIs in the LAA (green) and AA (orange) with their corresponding-colored distributions using the ADA method. (C) presents a LAA ROI with a mixture of trabeculae and contrast agents, where two fitted components demonstrate much wider spread than the AA ROI distribution. (D) shows a stasis case where the means of two components are close together but further left from that of the AA ROI distribution. (E) presents a clear ROI case, where the means of the components and that of the AA ROI distribution are all close together.
† AA, ascending aorta; CT, computer tomography; LAA, left atrial appendage; ROI, region of interest. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

2.8. Attenuation Distribution Analysis for automatic ROIs selection

To further select an ROI with potential filling defects, we proposed a novel method called *Attenuation Distribution Analysis* **(**ADA) to differentiate between different types of ROIs in two scenarios. The previously described CRC method was first used to differentiate the scenarios.

In Scenario 1, a candidate region was found by the CRC method. A 1-cm^2 circular kernel would scan through the region to find a candidate ROI. In this scenario, the ROI could be an ROI with trabeculae and contrast, or an ROI with stasis. Our proposed ADA method applied a two-component Gaussian mixture to every single ROI region as the kernel scanned through. The distribution of the corresponding AA ROI was used as a reference for comparison. Fig. 4 (C) shows a LAA ROI with a mixture of trabeculae and contrast, where two fitted components (green) demonstrate much wider spread than the AA ROI distribution (orange). The left-most component with the lowest attenuation range is the distribution with trabeculae. Fig. 4 (D) presents a stasis case where the means of the two components (green) are close together but are further left from that of the AA ROI distribution (orange).

In Scenario 2, no candidate region was found by the CRC method. The proposed ADA method was used to differentiate between an ROI with trabeculae and contrast, and a clear ROI case. As Fig. 4 (E) shows all the means of the distributions are close together in a clear case, as opposed to a trabeculae and contrast case in Fig. 4 (C).

One SD of the AA ROI distribution was experimentally set (Appendix C, Table C2) as the threshold to differentiate between the two cases in both scenarios. If the SDs of the fitted components were not within the AA distribution SD, it belonged to the trabeculae and contrast case. Such a case would be excluded from being a candidate ROI, since a trabeculae region often has lower mean HU value that could be confused with stasis. With this design, out method was able to effectively identify ROIs with potential filling defects and avoid ROIs with trabeculae. If no candidate was found, a kernel of 0.5 cm^2 and 0.3 cm^2 would be used in this order until it finds a candidate or terminates. The candidate with the lowest mean HU value was selected as the final ROI.

2.9. Statistical analysis

Continuous variables were expressed as mean ± SD or median (range) for normally and non-normally distributed data,

respectively. To assess normal distribution, the Kolmogorov-Smirnov test was performed. Continuous variables were compared using Student t-test or paired Wilcoxon signed-rank test as appropriate. Categorical variables were compared using a chi-squared or Fisher exact test. Dice coefficient was used to evaluate the performance of DL-based segmentation. Single rater intraclass correlation coefficient (ICC) analysis was used to assess the agreement between the automatically and manually selected ROIs. Bland-Altman (B-A) analysis, which yielded mean difference (bias) and limits of agreement (LOA) between methods as bias $\pm$ 1.96 SD, was used to assess the agreement of HU values between the automatic and manual selection.

One-way ANOVA and Kruskal–Wallis tests were used for the comparisons of LAA/AA HU ratios amongst subsets. The performance of the filling defects assessment was analyzed by the receiver operating characteristic curve (ROC) and measured by the area under the ROC curve (AUC). Box-and-whisker diagrams were used to compare the LAA/AA HU ratios between the filling defects and no-filling defects groups, as well as subset groups. All statistical analysis was performed using Python programming language (www.python.org) and the SciPy package (www.scipy.org).

3. Results

3.1. Patient characteristics

Eight patients with low quality CT images due to motion artifacts were excluded, with remaining 202 patients included in the study. However, the CT images with artifacts caused by metal implants were not excluded. Patients were elderly (68.1 $\pm$ 13.5 years of age), female (37.6%) and had a CHA_2DS_2-VASc score of 3.13 $\pm$ 1.96. Except Group 3, there were no significant differences in the baseline demographics, including age, gender and underlying comorbidities between Group 1 and Group 2 (Table 1). Group 3 had the youngest age, low CHA_2DS_2-VASc scores as compared to the other two groups.

3.2. Segmentation performance

The total number of annotated CT images was 19,328. The LAA volume segmentation achieved 0.931–0.945 Dice scores. Table 2 compares the performance of the LAA segmentation models on the Group 1 test set. U-Net with group normalization (GN) outperformed the standard U-Net model and U-Net with a pretrained VGG16 encoder. Adding K random negative samples in the training (NegSamlesK) sacrificed the performance slightly; however, it helped avoid false positives especially right before and after positive slices (i.e., slices with LAA regions). AA segmentation was used to extract an ROI $\leq 1\ cm^2$ at the center, as a reference value; therefore, a performance of 0.632 Dice coefficient sufficed for our purpose.

3.3. ROI selection performance

For the ICC analysis on the automatic vs. manual LAA ROI selection, our framework achieved excellent ICC = 0.929 (95% confidence interval, CI: 0.91–0.95, $p < 0.001$) on the entire dataset, good ICC = 0.895 (95% CI: 0.85–0.93, $p < 0.001$) on the Group 1 subset, excellent ICC = 0.948 (95% CI: 0.90–0.97, $p < 0.001$) on the Group 2 subset, and excellent ICC = 0.945 (95% CI: 0.89–0.97, $p < 0.001$) on the Group 3 subset. For the automatic AA ROI selection, the framework demonstrated excellent ICC = 0.967 (95% CI: 0.96–0.97, $p < 0.001$), ICC = 0.946 (95% CI: 0.92–0.96, $p < 0.001$), ICC = 0.963 (95% CI: 0.93–0.98, $p < 0.001$), and ICC = 0.984 (95% CI: 0.97–0.99, $p < 0.001$) on the entire dataset, Group 1, Group 2, and Group 3 subsets, respectively.

The results of B-A analysis on the automatic and manual selection of LAA and AA ROIs are shown in Fig. 5 (A-L). The mean biases of

Table 1
Baseline characteristics of patients.

	SCAF (n = 98)	AF-related stroke (n = 34)	AF for ablation (n = 70)	p-value
Age, years	75.0 $\pm$ 10.4	73.7 $\pm$ 10.3	55.8 $\pm$ 9.6	<0.001
Female, n (%)	42 (43)	11 (32)	23 (33)	0.322
Hypertension, n (%)	72 (72)	27 (79)	34 (49)	<0.001
Diabetes mellitus, n (%)	31 (32)	12 (35)	8 (11)	<0.005
HF, n (%)	15 (15)	15 (44)	2 (3)	<0.001
TIA/Prior stroke, n (%)	0	34 (100)	4 (6)	<0.001
Vascular disease, n (%)	21 (21)	12 (35)	11 (16)	0.076
CHA_2DS_2-VASc (range)	3.5 $\pm$ 1.3	5.7 $\pm$ 1.4	1.4 $\pm$ 1.2	<0.001
Medication, n (%)				
Oral anticoagulant, n (%)	41 (42)	18 (53)	15 (21)	<0.005
Anti-platelet agent, n (%)	24 (24)	13 (38)	28 (40)	0.075
Antiarrhythmic drugs, n (%)	35 (36)	15 (44)	30 (43)	0.545
Beta-blocker, n (%)	46 (47)	18 (53)	22 (31)	0.054
Calcium channel blocker	2 (2)	8 (24)	9 (13)	<0.001
ACEI/ARB, n (%)	49 (50)	20 (59)	22 (31)	0.012

[†] Data is presented as n (%) or mean (SD) unless otherwise specified. ACEI, angiotensin-converting enzyme inhibitors; AF: atrial fibrillation; ARB, angiotensin receptor blocker; HF, heart failure; SCAF, subclinical AF; TIA, transient ischemic attack.

Table 2
Segmentation performance.

LAA segmentation model	LAAV Dice	LAAM Dice
U-Net	0.6275	0.5778
U-Net + VGG16 encoder	0.6259	0.5774
U-Net + GN	0.9448	0.8675
U-Net + GN + NegSamplesK(K = 5)	0.9310	0.8607
U-Net + GN + NegSamplesK(K = 10)	0.9388	0.8571

† GN: Group Normalization; LAAV: left atrial appendage volume; LAAM: left atrial appendage myocardial mass; NegSamplesK: using random K negative slices before and after the positive slices.

the LAA/AA HU ratios were remarkably close to zero, which were −0.021, −0.029, −0.009 and −0.015 for the entire dataset, Group 1, Group 2, and Group 3 subsets, respectively. The 95% LOA ranges were −0.233 to 0.192, −0.253 to 0.195, −0.264 to 0.246 and −0.182 to 0.152, respectively. Further analysis revealed that there were no significant differences amongst the mean biases of subsets ($p = 0.561$).

In summary, the proposed DL-based framework showed stable performance across the patients with SCAF, clinical AF-related stroke and AF for ablation, as well as three scanner models from different manufacturers.

3.4. Performance of filling defects assessment

Our framework achieved an AUC score of 0.997 (95% CI, 0.997–0.998) on the Group 1 subset, 0.924 (95% CI, 0.916–0.925) on the Group 2 subset, and 0.979 (95% CI, 0.9792–0.981) on the entire dataset, as shown in Fig. 6 (A). Group 3 was not included in the ROC analysis because it did not contain filling defects cases. The optimal HU ratio thresholds for filling defects assessment identified from the ROC curve on the Group 1, Group 2, and entire dataset were 0.524, 0.561, and 0.561, respectively. These results demonstrated that the framework performed well on all the subsets analyzed.

The box-and-whisker diagram in Fig. 6 (B) shows significant differences in the LAA/AA HU ratios suggested by our automatic approach between the LAA filling defects group and no filling defects group ($p < 0.001$).

3.5. Comparing AI predicted LAA/AA HU ratios amongst groups

The mean ± SD of AI predicted HU ratios for Group 1, Group 2, and Group 3 were 0.905 ± 0.149, 0.813 ± 0.197, and 0.957 ± 0.092, respectively. The box-and-whisker diagram in Fig. 6 (C) compares the AI predicted HU ratios amongst three groups. The differences amongst three groups ($p < 0.001$) and between all pairs of groups (Group 1 vs. Group 2, $p = 0.005$; Group 2 vs. Group 3, $p < 0.001$; Group 3 vs. Group 1, $p = 0.012$) were statistically significant.

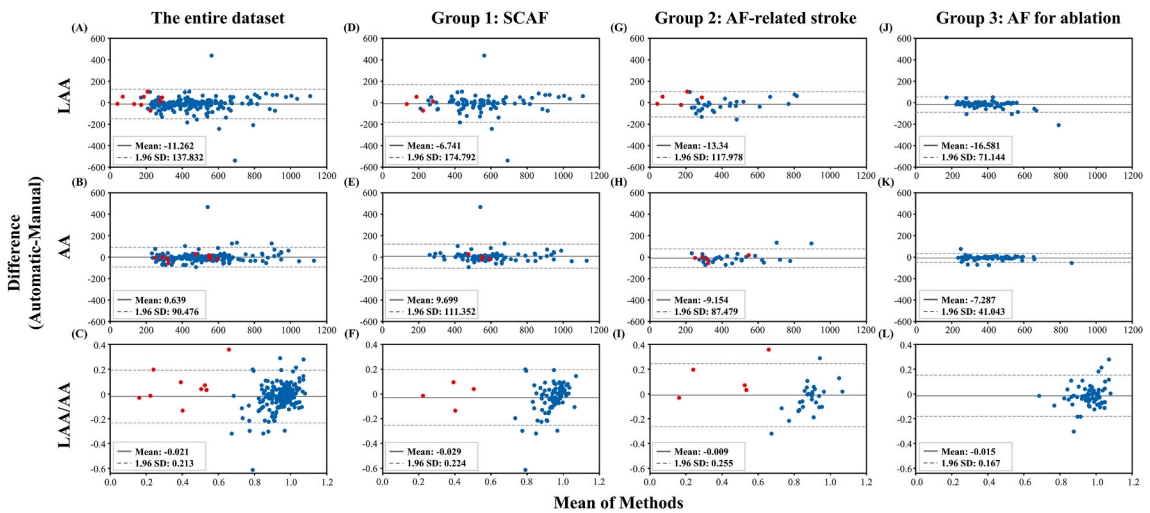

Fig. 5. The results of a Bland-Altman analysis. The columns are different subsets, and the rows are LAA, AA, LAA/AA HU values. The y-axis represents the difference (automatic - manual) in HU values, and the x-axis, the mean HU values of the methods compared. Filling-defects cases are marked red and non-filling-defects cases are marked blue. The solid line indicates mean bias, dotted line 95% LOA. The proposed framework showed stable performance across the patients with SCAF (D–F), clinical AF-related stroke (G–I) and AF for ablation (J–L), and the entire dataset (A–C).
† AA, ascending aorta; HU, Hounsfield units; LAA, left atrial appendage; LOA, limits of agreement. . (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

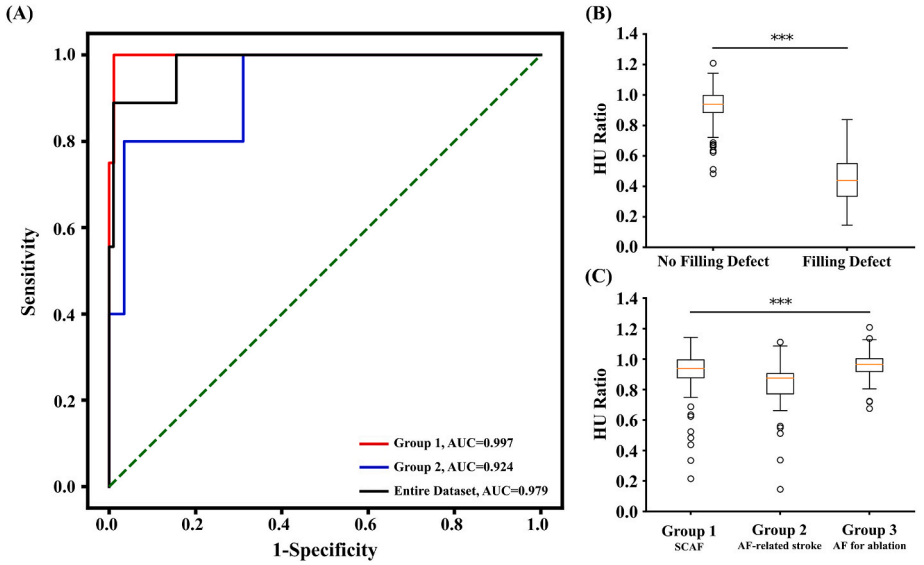

Fig. 6. (A) The performance of filling defects assessment was analyzed based on the thresholds of LAA/AA HU ratios on the ROC curves and measured by the area under the ROC curve (AUC). (B) and (C) are the box-and-whisker diagrams used to compare the LAA/AA HU ratios between the filling defects and no-filling defects groups, and amongst three study subsets respectively.
† AA, ascending aorta; HU, Hounsfield units; LAA, left atrial appendage; ROC, receiver operating characteristic. *** indicated $p < 0.001$.

4. Discussion

4.1. Major findings

We presented a fully automatic AI-based framework for filling defects assessment on cardiac CT for clinical and subclinical AF patients, based on DL segmentation networks and the proposed novel CRC and ADA ROI selection methods. Particularly, the proposed ADA method was capable of differentiating between an ROI with contrast only, an ROI with stasis, and an ROI with a mixture of trabeculae and contrast, which can be difficult by human eyes to differentiate. Therefore, our framework was able to accurately identify filling defects while avoiding trabeculae with excellent AUC scores ($\geq$0.924). The proposed approach also demonstrated an outstanding performance for LAA segmentation (Dice = 0.939), and good-to-excellent reliability for the automatic LAA ROI selection against manual selection (ICCs $\geq$0.895). Our approach achieved close-to-zero biases and excellent 95% LOAs in B-A analyses for automatic LAA/AA HU ratio suggestions. It may help preemptively detect the potential thromboembolic risk for AF patients in the various clinical scenarios.

4.2. The accuracy of LAA segmentation in different DL-based approaches

For automatic LAA segmentation, Morais et al. achieved a Dice score of 0.88 for their semi-automatic 3D LAA segmentation framework [20]; however, it relied on manual setting of the LAA centerlines. Takayashiki et al. utilized a U-Net based model for LAA segmentation on heart CT images [21], but the reported accuracies were poor. Jin et al. proposed a DL-based segmentation model for AF patients using a fully convolutional neural network with conditional random fields to handle spatial order in the slice sequences [12]. Their approach achieved a Dice score of 0.947 on LAA segmentation. Our LAA segmentation networks based on U-Net with GN achieved comparable results of 0.931–0.945 Dice scores.

4.3. The feasibility of LAA thrombotic risk prediction amongst different AI models

Ebrahimian et al. introduced a radiomics approach to differentiate between LAA thrombi and mixing artifacts on single-phase CT, using extracted first-order, shape-based and texture encoding features, without AA HU assessment [16]. Their approach achieved a 0.94 AUC performance in thrombus vs. no thrombus predication when radiomics combined with radiologist assessment, and 0.87 AUC in radiomics only. When distinguishing between thrombus and artifact, an AUC of 0.85 was reported in the solely radiomics model. Our framework achieved an AUC score of 0.979 on the entire dataset based on fully automated LAA ROI selection, followed by LAA/AA HU ratios calculation for filling defects assessment. Ferez et al. introduced a Point-Net inspired model avoiding the need to run computational fluid dynamics (CFD) simulations to predict endothelial cell activation, potentially as an in-silico index of LAA thrombotic risk [14]. This AI model focused on hemodynamic changes associated CFD for thrombogenesis prediction, rather than an anatomic approach for filling defects detection. Our approach was closer to the real-world practice and provided better interpretability in daily CT assessment.

4.4. The clinical implications of LAA/AA HU ratios and its thrombogenic threshold in the AI models

Romero et al. reported the LAA/AA HU ratio as the most frequently used measure to quantify LAA contrast attenuation [22]. The previous studies showed that a reduced LAA/AA HU ratio reflected circulation stasis within the LAA and correlated with decreased TEE velocity [232425]. A markedly reduced LAA/AA HU ratio may also correlate with the clinical situation of spontaneous echo contrast (SEC) or thrombus formation in patients with AF [8]. Multi-modality imaging techniques for the evaluation of patients with AF were recommended by the AF guidelines from major cardiology societies [26,27]. The findings of left atrial size, LAA configuration, LAA filling defects and LAA/AA HU ratios on cardiac CT will define management and drive treatment in patients with AF.

AI predicted LAA/AA HU ratios reflected the nature of our study groups. Clinical AF-related stroke patients (Group 2) would typically have a worse clinical condition than SCAF patients (Group 1) and thus had the lowest mean HU ratio (p-values $\leq$0.005 compared to the other two groups). This finding echoed our previous study on LA function in stroke patients, where patients with AF-related stroke had significantly reduced booster-pump function of LA and LAA, as compared to those with clinical AF without stroke. Reduced booster-pump function was shown to be independently associated with the occurrence of stroke [28]. On the other hand, patients preparing for catheter ablation of AF (Group 3) had the youngest mean age, lowest mean CHA_2DS_2-VASc scores, and highest mean HU ratio as compared to the other two groups, since these AF patients for ablation were supposed to have no filling defects.

The optimal threshold for LAA/AA HU ratios from our automatic approach was 0.561, determined by the ROC analysis. This threshold aligned with existing findings of LAA/AA HU ratios in relation to filling defects using TEE [6,7]. Patel et al. reported a threshold of LAA/AA HU ratios >0.75 for excluding the occurrence of LAA thrombus or dense non-clearing severe SEC [7]. Kim et al. reported an optimal cutoff of 0.25 for distinguishing between severe SEC or thrombus and lesser grades of SEC [6]. The threshold determined by our automatic approach fell in between the upper bound [7] and lower bound [6].

4.5. Study limitations

This study had some limitations. First, motion artifacts that cast shadows on CT images could contribute to inaccurate selection of ROIs; therefore, they were excluded in this study. Second, our approach partly relied on the correct phasing within R-R interval, particularly when no candidate region was found by the CRC method. Therefore, incorrect ECG-gating at the time of CT scanning or early start of P-phase and prolonged P-R phase could lead to suboptimal selection of LAA ROIs. Third, since our dataset did not have TEE as reference standard, a few potential filling defects cases were not used as positive cases due to the difficulty in judging by the experts. Finally, the number of cases with filling defects were limited in our dataset. Although the framework generalized well to the Group 2 test set, more data with filling defects cases in the future may help further verify our approach.

5. Conclusion

We presented a novel AI-based, fully automated framework for filling defects assessment on cardiac CT images. The framework, based on DL segmentation networks and the proposed CRC and ADA ROI selection methods, proved to be effective in identifying the LAA regions and selecting ROIs for filling defects assessment, while avoiding ROIs with trabeculae, contrast, or their mixture that can be difficult to differentiate by human eyes. This technique may help preemptively detect the potential thromboembolic risk for clinical AF and SCAF patients.

Author contribution statement

LC and ShH conceived and designed the experiments. LC and ThW performed the experiments and analyzed the data. ShH, HhW, HmT, and GjT analyzed and interpreted the data. TyL and VsT contributed analysis tools. ShH and HmT contributed the data. LC and ShH wrote the paper. All authors revised the article.

Funding statement

Dr. Ling Chen and Gau-Jun Tang were supported by National Science and Technology Council, Taiwan, R.O.C. [MOST 110-2222-E-A49A-501-MY3 & MOST 111-2321-B-A49-010].

Data availability statement

Data will be made available on request.

Declaration of interest's statement

The authors declare no competing interests.

Appendices

Appendix A. Loss functions for training the segmentation models

Loss functions for the LAA segmentation model.

The training of the LAA segmentation model used three loss functions. Cross Entropy (CE) loss is defined in Eq. (A.1), where for a CT slice i, Y_i is the ground truth annotation, $\widehat{Y}_i$ is the model output segmentation, j indicates the j^{th} class in Y_i, c is the total number of classes.

$$L_{CE}(Y_i, \widehat{Y}_i) = -\sum_{j}^{c} Y_i^j \, log \, \widehat{Y}_i^j \tag{A.1}$$

Dice loss is defined based on the Dice similarity coefficient (DSC), which measures the similarity between two sets of data. DSC is defined in Eq. (A.2), where $A(\bullet)$ is a function that returns the area, and Dice loss is defined in Eq. (A.3) accordingly.

$$DSC(Y_i, \widehat{Y}_i) = \frac{2 \times A(Y_i \cap \widehat{Y}_i)}{A(Y_i) + A(\widehat{Y}_i)} \tag{A.2}$$

$$L_{DSC}(Y_i, \widehat{Y}_i) = 1 - DSC(Y_i, \widehat{Y}_i) \tag{A.3}$$

Additionally, we designed a loss function called False Positive (FP) loss to further penalize false positives for the negative cases. The loss is defined in Eq. (A.4) below, where N is the set of negative slices, i.e., containing no LAA region.

$$L_{FP}(Y_i, \widehat{Y}_i) = \begin{cases} \sum_{j} \left(Y_{ij} - \widehat{Y}_{ij}\right)^2 \; if \; i \in N \\ 0 \quad otherwise \end{cases} \tag{A.4}$$

The final loss is the sum of the three losses: $L_{LAA} = L_{CE} + L_{DSC} + L_{FP}$.

Loss functions for the AA segmentation model.

The training of the AA segmentation used Binary cross entropy (BCE) loss and Dice loss (Eq. (A.3)). BCE loss is defined based on the CE loss (Eq. (A.1)) when $c = 2$ and $Y_i^0 = 1 - Y_i^1$. The final loss is the sum of two losses: $L_{AA} = L_{BCE} + L_{DSC}$.

Appendix B. Examples of LAA segmentation results

Examples of LAA segmentation results are listed in Figure B1, with the original CT slice in the first row, ground truth annotation in the second row, and AI segmentation in the third row.

Appendix C. Parameter tuning for automatic LAA ROI selection

To experimentally find the optimal upper bound and lower bound for the proposed Candidate Region Cutting (CRC) method for automatic slice selection, we ran experiments with the combinations of upper bound and lower bound thresholds, in the range of 1, 2, and 3 standard deviations (SDs). Table C1 compares the intraclass correlation coefficient (ICC) between the mean HU values of automatically and manually selected LAA region of interests (ROI) for different upper bound and lower bound settings. It can be seen that the upper bound threshold has more impact on the performance and that 3SDs yields the best results. The lower bound threshold, on the other hand, has ignorable effect on the results. Therefore, the optimal thresholds were set as 3 SDs below the mean of the contrast distribution for the upper bound and 1 SD below the mean of the candidate distribution for the lower bound, yielding ICC = 0.89455.

To further find the optimal threshold for the proposed Attenuation Distribution Analysis (ADA) for automatic LAA ROI selection, we experimented with thresholds in the rage of 1, 1.5, and 2 SDs. Table C2 shows that the optimal threshold is 1 SD of the ascending aorta (AA) distribution on the same slice, yielding ICC = 0.895.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.heliyon.2023.e12945.

References

[1] R. Beigel, N.C. Wunderlich, S.Y. Ho, R. Arsanjani, R.J. Siegel, The left atrial appendage: anatomy, function, and noninvasive evaluation, JACC Cardiovasc Imaging 7 (12) (2014) 1251–1265, https://doi.org/10.1016/j.jcmg.2014.08.009.

[2] A. Bernard, T. Leclercq, P.O. Comby, G. Duloquin, F. Ricolfi, Y. Béjot, et al., High rate of cardiac thrombus diagnosed by adding cardiac imaging in acute stroke computed tomography protocol, Int. J. Stroke 16 (6) (2021) 692–700, https://doi.org/10.1177/1747493020967623.

[3] P. Spagnolo, M. Giglio, D. Di Marco, P.M. Cannaò, E. Agricola, P.E. Della Bella, et al., Diagnosis of left atrial appendage thrombus in patients with atrial fibrillation: delayed contrast-enhanced cardiac CT, Eur. Radiol. 31 (3) (2021) 1236–1244, https://doi.org/10.1007/s00330-020-07172-2.

[4] S.H. Chun, Y.J. Suh, K. Han, S.J. Park, C.Y. Shim, G.R. Hong, et al., Differentiation of left atrial appendage thrombus from circulatory stasis using cardiac CT radiomics in patients with valvular heart disease, Eur. Radiol. 31 (2) (2021) 1130–1139, https://doi.org/10.1007/s00330-020-07173-1.

[5] K. Ouchi, T. Sakuma, H. Ojiri, A predictor for defects in filling of the left atrial appendage on cardiac computed tomography in patients with atrial fibrillation, J. Comput. Assist. Tomogr. 44 (2) (2020) 284–288, https://doi.org/10.1097/RCT.0000000000000984.

[6] Y.Y. Kim, A.L. Klein, S.S. Halliburton, Z.B. Popovic, S.A. Kuzmiak, S. Sola, et al., Left atrial appendage filling defects identified by multidetector computed tomography in patients undergoing radiofrequency pulmonary vein antral isolation: a comparison with transesophageal echocardiography, Am. Heart J. 154 (6) (2007) 1199–1205, https://doi.org/10.1016/j.ahj.2007.08.004.

[7] A. Patel, E. Au, K. Donegan, R.J. Kim, F.Y. Lin, K.M. Stein, et al., Multidetector row computed tomography for identification of left atrial appendage filling defects in patients undergoing pulmonary vein isolation for treatment of atrial fibrillation: comparison with transesophageal echocardiography, Heart Rhythm 5 (2) (2008) 253–260, https://doi.org/10.1016/j.hrthm.2007.10.025.

[8] J. Hur, H.N. Pak, Y.J. Kim, H.J. Lee, H.J. Chang, Y.J. Hong, et al., Dual-enhancement cardiac computed tomography for assessing left atrial thrombus and pulmonary veins before radiofrequency catheter ablation for atrial fibrillation, Am. J. Cardiol. 112 (2) (2013) 238–244, https://doi.org/10.1016/j.amjcard.2013.03.018.

[9] J. Hur, Y.J. Kim, H.J. Lee, J.W. Ha, J.H. Heo, E.Y. Choi, et al., Left atrial appendage thrombi in stroke patients: detection with two-phase cardiac CT angiography versus transesophageal echocardiography, Radiology 251 (3) (2009) 683–690, https://doi.org/10.1148/radiol.2513090794.

[10] J. Wang, H. Zhu, S.-H. Wang, Y.-D. Zhang, A review of deep learning on medical image analysis, Mobile Network. Appl. 26 (1) (2021) 351–380, https://doi.org/10.1007/s11036-020-01672-7.

[11] G. Litjens, F. Ciompi, J.M. Wolterink, B.D. de Vos, T. Leiner, J. Teuwen, et al., State-of-the-Art deep learning in cardiovascular image analysis, JACC Cardiovasc Imaging 12 (8 Pt 1) (2019) 1549–1565, https://doi.org/10.1016/j.jcmg.2019.06.009.

[12] C. Jin, J. Feng, L. Wang, H. Yu, J. Liu, J. Lu, et al., Left atrial appendage segmentation using fully convolutional neural networks and modified three-dimensional conditional random fields, IEEE J. Biomed. Health Inform. 22 (6) (2018) 1906–1916, https://doi.org/10.1109/JBHI.2018.2794552.

[13] Actor-critic reinforcement learning for automatic left atrial appendage segmentation, in: W.A. Al, I.D. Yun (Eds.), IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2018, https://doi.org/10.1109/BIBM.2018.8621575, 2018 3-6 Dec. 2018.

[14] X. Morales Ferez, J. Mill, K.A. Juhl, C. Acebes, X. Iriart, B. Legghe, et al., Deep learning framework for real-time estimation of in-silico thrombotic risk indices in the left atrial appendage, Front. Physiol. 12 (2021), https://doi.org/10.3389/fphys.2021.694945.

[15] B. Jeon, S. Jung, H. Shim, H.J. Chang, Bayesian estimation of geometric morphometric landmarks for simultaneous localization of multiple anatomies in cardiac CT images, Entropy 23 (1) (2021), https://doi.org/10.3390/e23010064.

[16] S. Ebrahimian, S.R. Digumarthy, F. Homayounieh, A. Primak, F. Lades, S. Hedgire, et al., Use of radiomics to differentiate left atrial appendage thrombi and mixing artifacts on single-phase CT angiography, Int. J. Cardiovasc. Imag. 37 (6) (2021) 2071–2078, https://doi.org/10.1007/s10554-021-02178-3.

[17] U-net: convolutional networks for biomedical image segmentation, in: O. Ronneberger, P. Fischer, T. Brox (Eds.), International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, https://doi.org/10.1007/978-3-319-24574-4_28.

[18] X. Zhou, G. Yang, Normalization in training U-net for 2-D biomedical semantic segmentation, IEEE Rob. Autom. Lett. 4 (2) (2019) 1792–1799, https://doi.org/10.1109/LRA.2019.2896518.

[19] Automatic instrument segmentation in robot-assisted surgery using deep learning, in: A.A. Shvets, A. Rakhlin, A.A. Kalinin, V.I. Iglovikov (Eds.), IEEE International Conference on Machine Learning and Applications, 17th, ICMLA, 2018, https://doi.org/10.1109/ICMLA.2018.00100, 2018 17-20 Dec. 2018.

[20] 3D segmentation of the left atrial appendage in computed tomography for planning of transcatheter occlusion, in: P. Morais, D. Nelles, J.W. Schrickel, G. Nickenig, J. D'hooge, A. Sedaghat, et al. (Eds.), Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling, SPIE, 2022, https://doi.org/10.1117/12.2610705.

[21] I. Takayashiki, A. Doi, T. Kato, H. Takahashi, S. Sekimura, M. Hozawa, et al. (Eds.), Method for Left Atrial Appendage Segmentation Using Heart CT Images, IEEE 10th International Conference on Awareness Science and Technology (iCAST), 2019, https://doi.org/10.1109/ICAwST.2019.8923258, 2019 23-25 Oct. 2019.

[22] J. Romero, J.J. Cao, M.J. Garcia, C.C. Taub, Cardiac imaging for assessment of left atrial appendage stasis and thrombosis, Nat. Rev. Cardiol. 11 (8) (2014) 470, https://doi.org/10.1038/nrcardio.2014.77.

[23] J. Hur, Y.J. Kim, H.-J. Lee, J.E. Nam, Y.J. Hong, H.Y. Kim, et al., Cardioembolic stroke: dual-energy cardiac CT for differentiation of left atrial appendage thrombus and circulatory stasis, Radiology 263 (3) (2012) 688–695, https://doi.org/10.1148/radiol.12111691.

[24] N. Funabashi, H. Takaoka, M. Uehara, T. Murayama, K. Ozawa, Y. Kobayashi, LAA CT contrast defects correlate with TEE LAA velocity and $CHADS_2$-score and are a prognostic indicator for embolism in subjects with atrial fibrillation or flutter, Int. J. Cardiol. 185 (2015) 297–300, https://doi.org/10.1016/j.ijcard.2015.02.051.

[25] A.D. Chang, G.C. Ignacio, R. Akiki, B. Mac Grory, S.S. Cutting, T. Burton, et al., Increased left atrial appendage density on computerized tomography is associated with cardioembolic stroke, J. Stroke Cerebrovasc. Dis. 29 (4) (2020), 104604, https://doi.org/10.1016/j.jstrokecerebrovasdis.2019.104604.

[26] E. Donal, G.Y. Lip, M. Galderisi, A. Goette, D. Shah, M. Marwan, et al., EACVI/EHRA Expert Consensus Document on the role of multi-modality imaging for the evaluation of patients with atrial fibrillation, Eur. Heart J. Cardiovasc. Imag. 17 (4) (2016) 355–383, https://doi.org/10.1093/ehjci/jev354.

[27] G. Hindricks, T. Potpara, N. Dagres, E. Arbelo, J.J. Bax, C. Blomström-Lundqvist, et al., ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): the Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC, Eur. Heart J. 42 (5) (2020) 373–498, https://doi.org/10.1093/eurheartj/ehaa612, 2021.

[28] H.M. Tsao, W.C. Hu, P.H. Tsai, C.L. Lee, H.H. Wang, S.L. Chang, et al., Functional remodeling of both atria is associated with occurrence of stroke in patients with paroxysmal and persistent atrial fibrillation, Acta Cardiol. Sin. 33 (1) (2017) 50–57, https://doi.org/10.6515/acs20160411a.