Deep Learning-Based Analysis of Aortic Morphology From Three-Dimensional MRI

Jia Guo, Kevin Bouaou, Sophia Houriez–Gombaud-Saintonge, Moussa Gueda, Umit Gencer, Vincent Nguyen, Etienne Charpentier, Gilles Soulat, Alban Redheuil, Elie Mousseaux, et al.

Deep learning-based analysis of aortic morphology from three-dimensional MRI

Jia Guo, PhD,[1,2] Kevin Bouaou, PhD,[1,2] Sophia Houriez--Gombaud-Saintonge, PhD,[1,2,3] Moussa Gueda, MSc, [1, 2], Umit Gencer, MSc,[4,5] Vincent Nguyen, MSc,[1, 2] Etienne Charpentier, MD, PhD,[1,3,6], Gilles Soulat, MD, PhD,[4,5] Alban Redheuil, MD, PhD,[1,2,6] Elie Mousseaux, MD, PhD,[4,5] Nadjia Kachenoura, PhD,[1,2*] Thomas Dietenbeck, PhD[1,2*]

1 Sorbonne Université, INSERM, CNRS, Laboratoire d'Imagerie Biomédicale (LIB), 75006, Paris, France

2 Institute of Cardiometabolism and Nutrition (ICAN), 75013, Paris, France

3 ESME Sudria Research Lab, 75015, Paris, France

4 Université de Paris Cité, PARCC, INSERM, F-75015, Paris, France

5 Assistance Publique Hôpitaux de Paris, Hôpital Européen Georges Pompidou, F-75015, Paris, France

6 Imagerie Cardio-Thoracique (ICT), Sorbonne Université, AP-HP, Groupe Hospitalier Pitié-Salpêtrière, 75013, Paris, France.

* These authors have equally contributed to this work

Corresponding author info

Thomas DIETENBECK (thomas.dietenbeck@sorbonne-universite.fr)

Laboratoire d'Imagerie Biomédicale,

Campus des Cordeliers, Escalier A - 4ème étage

15 rue de l'Ecole de Médecine, 75006 PARIS (France)

Phone: +33 1 44 27 91 19

Grant support:

J. G received a financial support from the China Scholarship Council (CSC) for his PhD fellowship (CSC Grant No.202008070119). We would like to acknowledge the FRM project ING20150532487 for funding K.B, ESME-Sudria for funding S.HGS, FRHTA for partially funding MRI acquisitions used in this study, and EIT health CMRAI (grant number 220636) project for funding V.N.

Running Title: **3D MRI DL analysis of aortic morphology**

ABSTRACT:

Background: Quantification of aortic morphology plays an important role in the evaluation and follow-up assessment of patients with aortic diseases, but often requires labor-intensive and operator-dependent measurements. Automatic solutions would help enhance their quality and reproducibility.

Purpose: To design a deep learning (DL)-based automated approach for aortic landmarks and lumen detection derived from three-dimensional (3D) MRI.

Study Type: Retrospective.

Population: 391 individuals (female:47%, age=51.9±18.4) from three sites, including healthy subjects and patients (hypertension, aortic dilation, Turner syndrome), randomly divided into training/validation/test datasets (n=236/77/78). 25 subjects were randomly selected and analyzed by 3 operators with different levels of expertise.

Field Strength/Sequence: 1.5-T and 3-T, 3D spoiled gradient-recalled or steady-state free precession sequences.

Assessment: Reinforcement learning and a two-stage network trained using reference landmarks and segmentation from an existing semi-automatic software were used for aortic landmark detection and segmentation from sinotubular junction to coeliac trunk. Aortic segments were defined using the detected landmarks while the aortic centerline was extracted from the segmentation and morphological indices (length, aortic diameter, and volume) were computed for both the reference and the proposed segmentations.

Statistical Tests: Segmentation: Dice similarity coefficient (DSC), Hausdorff distance (HD), average symmetrical surface distance (ASSD); landmark detection: Euclidian distance (ED); model robustness: Spearman correlation, Bland-Altman analysis, Kruskal-Wallis test for comparisons between reference and DL-derived aortic indices; inter-observer study: Williams

index (WI). A WI 95% confidence interval (CI) lower bound >1 indicates that the method is within the inter-observer variability. A p-value <0.05 was considered statistically significant.

Results: DSC was 0.90±0.05, HD was 12.11±7.79mm, and ASSD was 1.07±0.63mm. ED was 5.0±6.1mm. A good agreement was found between all DL-derived and reference aortic indices (r>0.95, mean bias <7%). Our segmentation and landmark detection performances were within the inter-observer variability except the sinotubular junction landmark (CI=(0.96;1.04)).

Data Conclusion: A DL-based aortic segmentation and anatomical landmark detection approach was developed and applied to 3D MRI data for achieve aortic morphology evaluation.

Keywords: aorta, 3D cardiac MRI, segmentation, deep learning

INTRODUCTION

Aortic morphology plays an important role in the management of aortic diseases (1, 2). Several imaging studies have focused on the quantification of aortic morphology and have shown strong associations between age and quantitative aortic indices (3–7), including length, diameters, cross-sectional areas, and volumes. Other studies investigated changes in such aortic parameters in the presence of a pathological condition, including aortic dissection (8, 9) and bicuspid aortic valve (10), or further assessed their ability to predict mortality and cardiovascular events (11–13).

In this context, MRI is a non-invasive imaging modality that is able to capture aortic anatomy through volumetric acquisitions. In three-dimensional (3D) MRI of the aorta, a crucial step for the precise extraction of clinically reliable quantitative indices is the segmentation of the aortic lumen and aortic subdivision into standardized anatomical segments (14). Indeed, it is recommended to report the diameter at specific locations or its maximal value within a segment (15). However, current techniques for 3D MRI segmentation require a high level of user interaction (16–18), thus increasing variability and hindering an objective and automated use in large cohorts.

Against this background, this study aimed at developing a deep learning (DL)-based method for quantitative 3D MRI data analysis. Furthermore, we aimed to evaluate performances in terms of segmentation and landmark localization quality, as well as in terms of correlations between aortic indices and reference values.

MATERIALS AND METHODS

Approval of the local ethics committee was obtained (NCT03895541, NCT02701855, NCT03474159, NCT02517944), and written informed consent was collected from all participants allowing research use of their data.

Study population

We studied 391 subjects (183 women, mean age = 51.9 ± 18.4 years), who underwent 3D MRI examinations to assess their aortic morphology at 3 different inclusion sites: site 1 (1.5-T MRI; Aera, Siemens Healthineers, Erlangen, Germany): 232 subjects, site 2 (3-T MRI; Prisma, Siemens Healthineers, Erlangen, Germany): 42 subjects, site 3 (3-T/1.5-T MRI; Discovery MR750w, GE Healthcare, Chicago, IL, USA): 117 subjects. The dataset included 119 (30%) healthy subjects, 95 (24%) hypertensive patients, 116 (30%) patients with an ascending thoracic aorta dilation and either a tricuspid or a bicuspid valve, and 61 (16%) patients with a Turner syndrome. Sites 1 and 2 included hypertensive patients, while sites 1 or 3 included healthy subjects and patients with aortic diseases. Of note, sites 1 and 3 are clinical sites specialized in cardiovascular imaging, while site 2 is a research center specialized in brain imaging where one of our study took place.

Healthy controls were asymptomatic and free of overt cardiovascular disease defined as a clinical history of hypertension, diabetes mellitus, dyslipidemia, renal disease, known inflammatory conditions, and malignancy. Inclusion criteria for hypertensives were: absence of secondary causes of hypertension, absence of personal history of cardiovascular disease, and stable regimen of antihypertensive treatment. Patients with diabetes were not included. ATAA was defined by a maximal diameter of the ascending aorta $\geqslant$ 41 mm or $\geqslant$ 22 mm/m2 when indexed to body surface area (BSA). Exclusion criteria were concomitant presence of aortic valve stenosis, regurgitation graded more than moderate, aortic coarctation, Marfan or Turner syndrome, history of aortic dissection and previous surgery. Aortic valve function as

well as bicuspid aortic valve (BAV) confirmation and fusion pattern were evaluated by using conventionally acquired stacks of two-dimensional cine anatomical and velocity images perpendicular to the aortic root.

For sites 1 and 3, a gadolinium-based contrast agent (Dotarem, Guerbet, or Multihance, Bracco Imaging, Milan, Italy) was typically injected (0.1 to 0.2 mmol/kg) during the MRI exam. For all sites, 3D data were acquired in a sagittal oblique volume encompassing the thoracic aorta during free breathing with electrocardiographic and respiratory gating and with the following scan parameters: site 1: steady-state free precession (SSFP) sequence, voxel size = 0.66×0.66×1.13 mm^3, echo time = 1.5 ms, repetition time = 283 ms, and flip angle = 90°; site 2: SSFP sequence, voxel size = 0.98×0.98×1 mm^3, echo time = 1.3 ms, repetition time = 311 ms, and flip angle = 19°; and site 3: spoiled gradient-recalled (SPGR) sequence, voxel size = 0.67×0.67×3.19 mm^3, echo time = 1.3 ms, repetition time = 3.1 ms, and flip angle = 24°. Controls included on sites 1 and 3 had contrast injection as part of their MRI examination, since they were primarily included for myocardial tissue characterization, and aortic MRI was performed after injection.

Reference segmentation, landmarks, and quantitative aortic indices

The ground truth segmentation, positions of the anatomical landmarks, and aortic morphological indices (aortic length L, maximal diameters D, and volumes V) were obtained for the entire dataset (n = 391) by one operator (Operator 1, T.D. 8 years of experience in cardiovascular image processing), using a previously described semi-automated tool (Mimosa, Sorbonne Université, (7, 16, 19)). Briefly, seven anatomical landmarks were manually placed from the sinotubular junction to the coeliac trunk (sinotubular junction, mid-ascending aorta, brachiocephalic artery, left subclavian artery, mid-descending aorta, diaphragm level, and coeliac trunk) and used to determine the aortic centerline and to delimit aortic segments. The

aortic lumen was then automatically segmented using a 3D active contour and a final centerline was calculated at the center of the segmentation. Such segmentations were controlled visually by the operator and if judged of high quality, global and segmental aortic indices were derived. Otherwise initialization was adjusted to improve final segmentation quality. The dataset was randomly divided into training (236), validation (77), and testing datasets (78). More specifically, in order to keep the proportion of patients with aortic diseases and of magnetic field strengths in each set, the 60/20/20% random split was performed separately for each site and pathology.

To assess DL versus human performance, 25 3D MRI datasets were randomly selected in the testing set. In order to preserve the proportion of pathologies for the inter-observer reproducibility study, 5 subjects were randomly selected in each of the following subsets: vendor 1 + normal aorta, vendor 1 + dilated aorta, vendor 2 + normal aorta, vendor 2 + dilated aorta, Turner syndrome. Three observers (T.D.: 8 years of experience, J.G.: few months of experience, and M.G.: received instructions from an expert during few sessions), segmented the aortic data using the custom in-house software.

Analysis workflow

To achieve fully automated analysis of the 3D MRI datasets of the aorta, we designed an analysis pipeline that combined a reinforcement learning (RL) algorithm to detect anatomical landmarks, with a two-stage network for aortic lumen segmentation, as illustrated in Figure 1. By integrating the outputs of these two techniques, the aortic centerline can be estimated and then used to initialize the segmentation process in the custom in-house software, to subsequently extract the quantitative indices of aortic morphology.

Aortic landmark detection

A double deep Q network (DDQN)-based algorithm (20) was implemented for the detection of the six manually positioned aortic landmarks. Since landmarks were positioned within the aortic lumen, the input of each agent was defined as three 3D patches centered on the agent position of the same size but with different scales (1, 2, and 4), to ensure feeding the network with sufficiently rich and discriminating information. These scales allowed modeling of the current neighborhood at different distances as achieved with feature points (21) or in multi-feature pyramid techniques (22). The model's architecture (Figure 2) consisted of four 3D convolutional layers followed by Leaky Relu activation layers (23) and a 3D max pooling layer. Two fully connected layers (FCL) sized 128 and 6 were used for the action prediction. Furthermore, as the aorta tapers from its proximal to its distal part, the block size used to describe the environment depended on the anatomical landmark position, which might result in a steep drop between the convolutional layers and the FCL. To tackle this issue, an extra FCL of size 512 was added after the last convolutional layer for the sinotubular junction, brachiocephalic trunk and ascending (AAo) and descending (DAo) aorta at the level of the pulmonary artery bifurcation.

The training and validation datasets were used to optimize the hyperparameters of the DDQN architecture. First, cross-sectional diameters at the level of each individual landmark were extracted from the reference segmentations and were used to adjust the patch size in order to account for the aortic tapering from its proximal to its distal part (Table 1). Secondly, each landmarks' relative position was modelled using a normal distribution estimated from landmark coordinates in the training set. During the agent's training, the point was randomly initialized following the fitted distribution while only the mean position of the estimated distribution was used during the test phase.

The process convergence was estimated using the point oscillation and the target Q values of these positions were compared to identify the landmark. To ensure a correct training, the agent was moved toward the landmark if the distance to the target after convergence was higher than 1 voxel.

Segmentation of the aortic lumen

The pipeline of the two-stage network proposed to segment the aortic lumen is illustrated in Figure 1. The first network used an under-sampled volume to obtain a coarse segmentation of the aorta, which was then used to crop the original volume around the aorta and resize it into an isotropic resolution. The second network was used to segment the aorta and the obtained segmentation was resized back to the original volume dimensions.

The proposed segmentation networks (Figure 3) were based on a 3D U-Net network (24, 25) with residual blocks (26), replacing the original convolutional layers in the encoding part of the U-Net architecture to smooth the forward and backward propagation of information. Attention gates (27) were added to the skip connections to limit the activated part to the segmented area and account for class imbalance (background versus aortic lumen). Finally, a multi-feature pyramid (22) was used to fuse the feature maps with strong low-resolution semantic information and high-resolution feature maps with weak semantic information but rich spatial information.

The input volume size was 256 × 256 × 64 (respectively 320 × 192 × 128) for the first (respectively second) network. The model was trained for a maximum of 200 epochs with the Adam optimizer, a batch size of 2 (respectively 1) for the first (respectively second) network, an initial learning rate of 1e-3, and a decay factor of 0.1 with a patience of 10 epochs. The chosen loss function was the Dice loss.

The model was developed in Python (v3.8.11; Python Software Foundation, Wilmington, DE, USA) using Keras (v2.4.3; https://keras.io) and TensorFlow (v2.4.1; Google, Moutain View,

CA, USA) and trained on a Nvidia Quadro GV100 GPU (Nvidia, Santa Clara, CA, USA). The training of the DDQN-based agent took approximately 37 hours on average per landmark, the training of the first segmentation network took 25 epochs of 3 hours, and the training of the second segmentation network took 24 epochs of 5 hours.

Quantitative aortic indices

The active contour used in our custom software required a binary volume that roughly corresponded to the aorta to act as a balloon force guiding the lumen segmentation towards the aortic wall. This binary volume was replaced by our network segmentation output while the detected landmarks allowed for a computation of the aortic centerline leading to a fully automated segmentation of the aorta and its further subdivision into anatomical segments: AAo, arch and DAo. Morphological quantitative indices including length, AAo and DAo maximal diameters, and volumes as well as the aortic arch length computed between mid-AAo and mid-DAo were then extracted, as shown in Figure 4.

Statistical analysis

All statistical evaluations were performed using MATLAB (R2021b; Mathworks, Natick, MA, USA). To evaluate the performances of the RL agents, 3D Euclidean distances (ED) were computed between the detected and the reference landmarks. The Dice similarity coefficient (DSC), Hausdorff distance (HD), and average symmetrical surface distance (ASSD) were used to evaluate the segmentation performances on a global or local scale. The robustness of the segmentation was also evaluated by comparing the algorithm's performances between subgroups (for comparisons per vendor, per field strength, and per aortic dilation status).

The segmentation results provided by our method were also compared with those of the segmentation network proposed by Berhane et al. (28). Note that, since this network was originally proposed for the 3D segmentation of 4D flow MRI angiographic images with a

network's input corresponding to a volume cropped around the aorta, we modified the size of the input layer of the network to accept volumes of the same size than our second network and retrained the network on our training dataset. These comparisons were performed using a Student t-test.

Indices of aortic morphology obtained by the proposed method were compared to the reference indices using Spearman correlation coefficient r, Bland-Altman mean bias and limits of agreement (LoA), and a non-parametric Kruskal Wallis test was used to evaluate statistical differences between the 2 measurements. Bland-Altman biases were further expressed in percentage of the reference measures to account for differences in aortic morphology between subgroups (for comparisons per vendor, per field strength, and per aortic dilation status).

The human interobserver variability was quantified using ED for the landmarks detection and using DSC, HD, and ASSD for the segmentation, and it was subsequently compared to the corresponding proposed DL method. The mean differences between DL and human observers (COD) and the mean interobserver difference (IOD) were also computed. The statistical comparison of these differences was performed using the Williams index (WI) along with its confidence interval (CI) (29). If the upper bound of the CI was higher than 1, then there was a statistical evidence that the variability between the DL-based method and the human observers was within the interobserver variability. More details on the WI computation are given in the Supplementary Material. A p-value <0.05 was considered statistically significant.

RESULTS

Basic model characteristics and indices

Basic characteristics and reference morphological aortic indices are summarized in Table 2 for the entire population and per acquisition site. Participants from site 2 were on average older as compared to participants from sites 1 and 3, and had a longer aortic arch and larger DAo on average. Standard deviations of maximal diameters and volumes were high, highlighting the heterogeneity of our database, which included extreme cases in terms of aortic size (small sized aortas: Turner syndrome, normal sized aortas: healthy controls, dilated aortas: patients with hypertension to a small extent and patients with dilated AAo to a larger extent).

Evaluation of landmark detection

Figure 5 shows the performances of the automated detection of aortic landmarks in terms of distances in relation to the expert reference targets and the WI with its 95% CI. On the entire testing dataset, our network detected the aortic landmarks with a distance to the reference below 9 mm on average. The highest mean distances between detected reference landmarks were found for the sinotubular junction (8.7 mm on average) and the coeliac trunk (11.8 mm on average), while the detection errors for the remaining landmarks were all below 5 mm on average. The COD was on average lower than the IOD for all landmarks resulting in WI and its 95% CI being higher than 1, except for the sinotubular junction and coeliac trunk revealing that the DL-based landmark detection was in the same variability range as compared to the human observers with variable levels of experience.

Evaluation of aortic lumen segmentation

Metrics corresponding to the comparison between the DL-based and the reference aortic segmentations on the entire testing dataset are summarized in Table 3 for the whole aorta and for each of the main segments (AAo, Arch, DAo). An overall DSC of 0.90 ± 0.05, HD of 12.11 ± 7.79 mm, and ASSD of 1.07 ± 0.63 mm was obtained for the whole aorta. Of note, 89.7% (70/78) of the testing dataset subjects had a DSC score > 0.85 and two patients with a low DSC value were excluded from the morphological analyses due poor image quality. An example of two segmentations with a DSC < 0.85 is given in Figure 6. Figure 7 summarizes the DSC between the DL-based and the reference segmentations, revealing that the DL-based method had similar numbers of segmentations with DSC less than 0.85 on the 1.5-T data and 3-T data, with slightly lower DSC values for the poorly segmented 3-T data.

Table 4 shows the results of the variability study in terms of DSC, HD, and ASSD on the whole aorta. The COD was slightly lower on average or similar to the IOD for all segmentation similarity metrics, resulting in the upper bound of the WI 95% CI being slightly higher than 1, thus demonstrating that both on a global and local scale the agreement between the DL-based method and human observers, with variable level of expertise, was in the same range as compared to the agreement between the human operators.

Table 3 shows the results of the comparison between the segmentation results of Berhane et al. (28) network against our method. No statistical differences could be found between both methods (DSC: $p = 0.069$, HD: $p = 0.526$, ASSD: $= 0.939$).

Evaluation in terms of aortic morphological measures

Table 5 summarizes the comparison between the aortic quantitative measures (aortic arch length L, AAo and DAo maximal diameters, and volumes) extracted by the proposed method against those extracted from the reference segmentation in the AAo, aortic arch, and DAo.

The two measurements were highly correlated and were not statistically significantly different (aortic arch L: r = 0.95, p = 0.72; AAo D: r = 0.95, p = 0.42; AAo V: r = 0.95, p = 0.91; DAo D: r = 0.96, p = 0.07; DAo V: r = 0.97, p = 0.29). Furthermore, good agreement was found between the two methods, as revealed by low Bland-Altman biases and LoA. Indeed, when such biases were expressed as the percentage of the reference measure, they did not exceed 9% for all the aortic indices: aortic arch length: 1.7%; AAo maximal diameter: 3.9%; DAo maximal diameter: 5.8%; AAo volume: 1.56%; DAo volume: 8.6%.

Table 6 summarizes Bland-Altman analyses for comparisons between the DL-based and the reference approaches in terms of aortic morphological measures per vendor, field strength, (1.5-T versus 3-T), and aortic dilation status (qualified as dilated versus small to normal sized aorta) while Figure 8 shows this analysis for the AAo and DAo V. Table 6 also shows Bland-Altman biases expressed as the percentage of the reference measures, revealing that the DL-based approach was stable among the different subgroups (aortic arch L: aortic dilation status: p = 0.23; AAo maximal D: vendor: p = 0.07, magnetic field: p = 0.78, aortic dilation status: p =0.75; DAo maximal D: vendor: p = 0.43, magnetic field: p = 0.26, aortic dilation status: p =0.13; AAo V: vendor: p = 0.10, magnetic field: p = 0.63, aortic dilation status: p =0.27; AAo maximal D: vendor: p = 0.10, aortic dilation status: p =0.27). One might note however that quite large biases (12.6%) were obtained for the DAo volume from the 1.5-T data.

DISCUSSION

We proposed an automated method based on DL for aortic anatomical landmark detection and segmentation of the aortic lumen from 3D MRI images, which may result in reliable segmental aortic morphological indices. This study proposed a fully automated and comprehensive method, including landmark detection, aortic segmentation, and estimation of aortic morphological indices from 3D MRI datasets originating from different vendors, centers, MRI field strengths, and from a large cohort of patients with variable aortic sizes and disease conditions. Although studies describing the aortic morphology in CT exist (8, 9), there is currently a lack of software and studies providing such indices in 3D MRI (3). Such an automatic software allowing measuring objectively the aortic morphology in a segmental and global way in 3D MRI could ultimately help for a more comprehensive and systematic management of patients with aortic diseases.

The consistency of the proposed DL-based method was demonstrated through: 1) low ED between the detected and the reference aortic anatomical landmarks as well as high DSC, low HD and low ASSD for the 3D aortic lumen segmentation, 2) variability against human annotations, which was comprised within the human interobserver variability, in terms of landmark detection and aortic segmentation, 3) high correlations and low biases for comparisons against references in terms of aortic morphological indices, and 4) equal performances against state-of-the-art segmentation method as well as in terms of aortic morphological indices between vendors, magnetic field strengths, and aortic dilation status, thus potentially revealing the generalizability of our approach.

An important feature of this study may be the database used for training. Specifically, it included 1) healthy subjects as well as patients with a wide spectrum of age, and 2) images acquired at different sites with variable levels of expertise (sites 1 and 3: experts in cardiovascular imaging, site 2: expert in brain imaging) as well as on MRI scanners from

different vendors and with different magnetic field strengths. This heterogeneity may have increased the variability and the representativeness of our data, reducing the selection bias and improving the model generalization. The value of such heterogeneity has been highlighted in various other DL-based studies focused on the aorta (28, 30).

Landmark detection performances were better for the AAo than for the DAo. This might be explained by lower contrast around the diaphragm as well as the increased tortuosity of the distal DAo segment with aging or pathology, thus hampering the landmark detection. Furthermore, the AAo is surrounded by enhancing structures (valve, heart, and pulmonary artery) and has a specific shape that may feed the model by specific features rather than the DAo, which has a more tubular and straight shape, surrounded by moderately enhancing or small structures. One might also highlight that our database included patients with subclinically dilated aorta or with pathological dilatation of its proximal segment, thus highlighting the ability of our detection network to capture information at different scales thanks to the pyramidal input.

The DL-based method exhibited lower performance in terms of DSC in the AAo and aortic arch when comparing with the performance for the whole aorta. This disparity could be attributed to the much smaller size of these segments compared to the whole aorta as similar HD and ASSD were obtained in these segments compared to the whole aorta, as well as to small biases when comparing aortic indices extracted from our segmentation with the reference indices. The DL-based method also provided aortic morphological indices that were robust regarding heterogeneity of our data except for the DAo volume where a difference was found depending on the magnetic field strengths. As the DAo ends at the diaphragm, this landmark detection is challenging as revealed by higher detection errors. This incorrect point detection then can lead to an erroneous definition of the segments and higher variability in

the extracted aortic volumes. However, no difference was found for the diameter since maximal diameters did not necessarily occur in end portions of segmentation.

Recent studies also proposed automated segmentation approaches of the aorta in four-dimensional (4D) flow MRI or CT usually based on U-Net architectures(28, 30–36). The comparison between our proposed method and the segmentation network proposed by Berhane et al. (28) revealed that the performances of both methods were similar on a global level as assessed by the DSC and ASSD as well as on a local level assessed by the HD. Hepp et al. (32) used a 3D U-Net to segment the aorta in non-contrast-enhanced MRI data obtaining a mean DSC of 0.85 ± 0.15 on 30 subjects from the general population in a multi-centric study. On 4D flow MRI, 3D U-Net (28, 33) or nnUNet (30, 31) were used for aortic segmentation of the peak systolic angiogram or magnitude data with good performances ((33): N = 36 with bicuspid/tricuspid aortic valve using leave-one-patient-out: DSC = 0.92 ± 0.05, HD = 21.02 ± 24.20 mm, and for AAo+arch: DSC = 0.93 ± 0.02, HD = 9.41 ± 3.45mm, for DAo: DSC = 0.93 ± 0.02, HD = 5.86 ± 6.23mm; (31): N = 81, DSC = 0.95, ASSD = 0.84mm). Similarly, on CT data, 3D U-Net (36) and its variants using residual (34) or dense (35) blocks were used with similar performances ((36): N = 191 healthy subjects and patients with an aortic aneurysm, DSC = 0.95 [0.94; 0.95], HD = 8 [4.47; 10]mm; (34): N = 162, DSC = 0.95 ± 0.02, ASSD = 0.53 ± 0.26mm, HD = 2.13 ± 0.91mm, (35): N = 30, DSC = 0.887). These studies included healthy and pathological subjects (28, 31) or patients with a dilated aorta (33). However, as acknowledged by the authors in their studies (28, 31, 33), a main limitation of their respective work lies in the monocentric design. Regarding aortic landmarks, previous studies (31, 37, 38) usually focused on the detection of only 3 landmarks, namely the sinotubular junction, brachiocephalic artery, and left subclavian artery. The detection error on a single-center cohort of 81 healthy subjects or patients ranged between 6.1 mm and 9.3 mm (38), which is in the same range as the distances obtained in the present study. Our

results were thus in-line with those obtained in the literature ((31, 37, 38) although a direct comparison was not possible due to the differences in MRI data and/or study populations.

Limitations

Two subjects out of the 78 in the testing dataset had to be excluded from the morphological analysis due to low segmentation quality. These 2 subjects were both BAV patients acquired on site 3 with a very low image quality and signal loss in the AAo. A review of the whole dataset revealed that although some other subjects had low image quality, none presented a low contrast or signal loss in the AAo similar to these two subjects. This absence of subjects with similar image quality in the training dataset as well as the fact that one of the two patients came from a very small subset (site 3, 1.5-T, BAV: n = 4 in the whole dataset) might explain the difficulty our network had in generalizing the learned features to this particular type of data. Since the complete processing took only an average of 50 s compared to the 5 min for semi-automated analysis, this spared time could also be used, when required, to correct the segmentation on these complex datasets. Deep Learning network performances are directly linked to the number of training data and their variability and our number of investigated subjects can be seen as small in this regard. However, our database has the advantage of including multiple sources of variability and the different evaluations performed on our dataset showed that our proposed method was robust to those variabilities. Despite the high diversity of our database, we can acknowledge that the absence of patients with very complex aortic morphological changes (e.g., abdominal aneurysms, aortic coarctation) is another limitation of our study. Finally, the human observer variability estimation comprised operators of various levels of expertise rendering the variability evaluation of the DL-based method less challenging. However, one might highlight that such variability might reflect real-life conditions in medical image processing where the different operators may be at various levels of the learning curve when they are asked to perform quantitative assessment.

Conclusion

In this study, an end-to-end automated method for aortic landmark detection and lumen segmentation on 3D MRI data was developed, which may provide aortic morphological indices such as aortic arch length, cross-sectional diameters, and volumes. The method combined RL for aortic landmark detection and a two-stage network for aortic lumen segmentation and was evaluated on data from multiple sites and including various pathologies.

REFERENCES

1. Brady AR, Fowkes FG, Thompson SG, Powell JT: Aortic aneurysm diameter and risk of cardiovascular mortality. *Arterioscler Thromb Vasc Biol* 2001; 21:1203–1207.

2. Lo RC, Lu B, Conrad MF, et al.: Relative Importance of Aneurysm Diameter and Body Size for Predicting AAA Rupture in Men and Women. *Journal of Vascular Surgery* 2013; 57:287–288.

3. Kawel-Boehm N, Hetzel SJ, Ambale-Venkatesh B, et al.: Reference ranges ("normal values") for cardiovascular magnetic resonance (CMR) in adults and children: 2020 update. *J Cardiovasc Magn Reson* 2020; 22:87.

4. Turkbey EB, Jain A, Johnson C, et al.: Determinants and normal values of ascending aortic diameter by age, gender, and race/ethnicity in the Multi-Ethnic Study of Atherosclerosis (MESA). *J Magn Reson Imaging* 2014; 39:360–368.

5. Houriez-Gombaud-Saintonge S, Mousseaux E, Bargiotas I, et al.: Comparison of different methods for the estimation of aortic pulse wave velocity from 4D flow cardiovascular magnetic resonance. *J Cardiovasc Magn Reson* 2019; 21:75.

6. Bouaou K, Bargiotas I, Dietenbeck T, et al.: Analysis of aortic pressure fields from 4D flow MRI in healthy volunteers: Associations with age and left ventricular remodeling. *J Magn Reson Imaging* 2019; 50:982–993.

7. Dietenbeck T, Houriez-Gombaud-Saintonge S, Charpentier E, et al.: Quantitative magnetic resonance imaging measures of three-dimensional aortic morphology in healthy aging and hypertension. *J Magn Reson Imaging* 2021; 53:1471–1483.

8. Heuts S, Adriaans BP, Rylski B, et al.: Evaluating the diagnostic accuracy of maximal aortic diameter, length and volume for prediction of aortic dissection. *Heart* 2020; 106:892–897.

9. Krüger T, Conzelmann LO, Bonser RS, et al.: Acute aortic dissection type A. *Br J Surg* 2012; 99:1331–1344.

10. Bollache E, Guzzardi DG, Sattari S, et al.: Aortic valve-mediated wall shear stress is heterogeneous and predicts regional aortic elastic fiber thinning in bicuspid aortic valve-associated aortopathy. *J Thorac Cardiovasc Surg* 2018; 156:2112-2120.e2.

11. van Ooij P, Markl M, Collins JD, et al.: Aortic Valve Stenosis Alters Expression of Regional Aortic Wall Shear Stress: New Insights From a 4-Dimensional Flow Magnetic Resonance Imaging Study of 571 Subjects. *J Am Heart Assoc* 2017; 6:e005959.

12. Ambale-Venkatesh B, Yang X, Wu CO, et al.: Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. *Circ Res* 2017; 121:1092–1101.

13. Redheuil A, Wu CO, Kachenoura N, et al.: Proximal aortic distensibility is an independent predictor of all-cause mortality and incident CV events: the MESA study. *J Am Coll Cardiol* 2014; 64:2619–2629.

14. Erbel R, Aboyans V, Boileau C, et al.: 2014 ESC Guidelines on the diagnosis and treatment of aortic diseases: Document covering acute and chronic aortic diseases of the thoracic and abdominal aorta of the adult. The Task Force for the Diagnosis and Treatment of Aortic Diseases of the European Society of Cardiology (ESC). *Eur Heart J* 2014; 35:2873–2926.

15. van Hout MJ, Scholte AJ, Juffermans JF, et al.: How to Measure the Aorta Using MRI: A Practical Guide. *J Magn Reson Imaging* 2020; 52:971–977.

16. Dietenbeck T, Craiem D, Rosenbaum D, et al.: 3D aortic morphology and stiffness in MRI using semi-automated cylindrical active surface provides optimized description of the vascular effects of aging and hypertension. *Comput Biol Med* 2018; 103:101–108.

17. Volonghi P, Tresoldi D, Cadioli M, et al.: Automatic extraction of three-dimensional thoracic aorta geometric model from phase contrast MRI for morphometric and hemodynamic characterization. *Magn Reson Med* 2016; 75:873–882.

18. van Pelt R, Nguyen H, ter Haar Romeny B, Vilanova A: Automated segmentation of blood-flow regions in large thoracic arteries using 3D-cine PC-MRI measurements. *Int J CARS* 2012; 7:217–224.

19. Dietenbeck T, Bouaou K, Houriez-Gombaud-Saintonge S, et al.: Value of aortic volumes assessed by automated segmentation of 3D MRI data in patients with thoracic aortic dilatation: A case-control study. *Diagn Interv Imaging* 2023:S2211-5684(23)00076–1.

20. Hasselt H van, Guez A, Silver D: Deep Reinforcement Learning with Double Q-Learning. *Proceedings of the AAAI Conference on Artificial Intelligence* 2016; 30.

21. Lowe DG: Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 2004; 60:91–110.

22. Moradi S, Oghli MG, Alizadehasl A, et al.: MFP-Unet: A novel deep learning based approach for left ventricle segmentation in echocardiography. *Physica Medica* 2019; 67:58–69.

23. Maas AL, Hannun AY, Ng AY: Rectifier nonlinearities improve neural network acoustic models. In *in ICML Workshop on Deep Learning for Audio, Speech and Language Processing*; 2013.

24. Ronneberger O, Fischer P, Brox T: U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Edited by Navab N, Hornegger J, Wells WM, Frangi AF. Cham: Springer International Publishing; 2015:234–241.

25. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O: 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. 2016.

26. He K, Zhang X, Ren S, Sun J: Identity Mappings in Deep Residual Networks. *arXiv:160305027 [cs]* 2016.

27. Schlemper J, Oktay O, Schaap M, et al.: Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis* 2019; 53:197–207.

28. Berhane H, Scott M, Elbaz M, et al.: Fully automated 3D aortic segmentation of 4D flow MRI for hemodynamic analysis using deep learning. *Magnetic Resonance in Medicine* 2020; 84:2204–2218.

29. Chalana V, Kim Y: A methodology for evaluation of boundary detection algorithms on medical images. *IEEE Transactions on Medical Imaging* 1997; 16:642–652.

30. Aviles J, Talou GDM, Camara O, et al.: Domain Adaptation for Automatic Aorta Segmentation of 4D Flow Magnetic Resonance Imaging Data from Multiple Vendor Scanners. In *Functional Imaging and Modeling of the Heart*. *Volume 12738*. Edited by Ennis DB, Perotti LE, Wang VY. Cham: Springer International Publishing; 2021:112–121. [*Lecture Notes in Computer Science*]

31. Garrido-Oliver J, Aviles J, Córdova MM, et al.: Machine learning for the automatic assessment of aortic rotational flow and wall shear stress from 4D flow cardiac magnetic resonance imaging. *Eur Radiol* 2022; 32:7117–7127.

32. Hepp T, Fischer M, Winkelmann MT, et al.: Fully Automated Segmentation and Shape Analysis of the Thoracic Aorta in Non–contrast-enhanced Magnetic Resonance Images of the German National Cohort Study. *Journal of Thoracic Imaging* 2020; 35:389–398.

33. Marin-Castrillon DM, Geronzi L, Boucher A, et al.: Segmentation of the aorta in systolic phase from 4D flow MRI: multi-atlas vs. deep learning. *MAGMA* 2023.

34. Yang J, Li X, Cheng J-Z, et al.: Segment aorta and localize landmarks simultaneously on noncontrast CT using a multitask learning framework for patients without severe vascular disease. *Computers in Biology and Medicine* 2023; 160:107002.

35. Ma Q, Lucas A, Hammami H, Shu H, Kaladji A, Haigron P: Deep-learning approach to automate the segmentation of aorta in non-contrast CTs. *J Med Imaging (Bellingham)* 2023; 10:024001.

36. Sieren MM, Widmann C, Weiss N, et al.: Automated segmentation and quantification of the healthy and diseased aorta in CT angiographies using a dedicated deep learning approach. *Eur Radiol* 2022; 32:690–701.

37. Guala A, Mejia Cordova M, Morales X, et al.: Machine learning to automatically detect anatomical landmarks on phase-contrast enhanced magnetic resonance angiography. *European Heart Journal - Cardiovascular Imaging* 2021; 22(Supplement_2):jeab090.122.

38. Mejia Cordova M, Guala A, Morales X, et al.: Reinforcement machine learning-based aortic anatomical landmarks detection from phase-contrast enhanced magnetic resonance angiography. *European Heart Journal - Cardiovascular Imaging* 2021; 22(Supplement_1):jeaa356.286.

TABLES

Table 1: Block size used by the Reinforcement Learning agent for each landmark.

	Block size (pixels)
Sinotubular junction	(48, 68, 52)
Ascending aorta at the level of the bifurcation of the pulmonary artery	(40, 70, 56)
Brachiocephalic arterial trunk	(58, 80, 40)
Left subclavian artery	(40, 48, 40)
Descending aorta at the bifurcation of the pulmonary artery	(40, 16, 40)
Hepatic dome / diaphragm	(38, 42, 38)
Coeliac trunk	(38, 40, 38)

Table 2: Subject characteristics for the whole population and for each recruitment site.

	Overall	Site 1	Site 2	Site 3, 1.5T	Site 3, 3T
N (M / F)	391 (205/186)	232 (107/125)	42 (27/15)	10 (6/4)	107 (65/42)
Age (years)	51.9 ± 18.4	46.2 ± 17.3	73.9 ± 5.42	57.9 ± 13.7	53.8 ± 17.6
BMI ($kg.m^{-2}$)	24.5 ± 4.4	24.6 ± 4.1	25.3 ± 3.4	23.6 ± 3.7	24.6 ± 4.0
BSA (m^2)	1.76 ± 0.27	1.75 ± 0.25	1.81 ± 0.19	1.81 ± 0.17	1.81 ± 0.22
cSBP (mmHg)	119 ± 16	120 ± 16	123 ± 12	117 ± 14	115 ± 15
cDBP (mmHg)	81 ± 10	82 ± 11	76 ± 8	84 ± 11	80 ± 10
cPP (mmHg)	38 ± 11	38 ± 11	47 ± 10	34 ± 7	34 ± 11
Number of subjects					
Healthy controls	119	58	0	0	61
Hypertensives	96	53	43	0	0
BAV	49	23	0	4	22
TAV-ATAA	67	37	0	6	24
Reference aortic morphological measures					
Arch length (mm)	129.80 ± 28.00	123.92 ± 27.24	144.31 ± 19.42	154.94 ± 29.23	134.49 ± 28.44
AAo maximal diameter (mm)	32.58 ± 7.86	31.27 ± 8.34	34.62 ± 4.00	41.10 ± 3.85	33.91 ± 7.37
DAo maximal diameter (mm)	24.40 ± 4.87	22.79 ± 4.60	29.17 ± 2.68	28.09 ± 5.34	25.75 ± 4.28
AAo volume (mL)	49.89 ± 33.91	46.50 ± 36.16	49.97 ± 15.51	87.43 ± 30.45	53.84 ± 34.44
DAo volume (mL)	64.57 ± 32.85	53.75 ± 28.55	86.97 ± 18.43	97.45 ± 36.81	76.49 ± 35.38

Abbreviations: BMI: body mass index, BSA: body surface area, cSBP: central systolic blood pressure, cDBP: central diastolic blood pressure, cPP: central pulse pressure, BAV: bicuspid aortic valve, TAV-ATAA: ascending thoracic aorta aneurysm and tricuspid aortic valve, AAo: ascending aorta, DAo: descending aorta.

Table 3: Average performances of deep learning-based aortic segmentation and comparison with a state-of-the-art method.

	AAo	Arch	DAo	Overall	Berhane et al. (28)
DSC	0.85 ± 0.10	0.84 ± 0.06	0.91 ±0.04	0.90 ± 0.05	0.91 ± 0.04 p = 0.069
HD (mm)	8.13 ± 3.35	5.10 ± 3.50	5.33 ± 4.50	12.11 ± 7.79	11.3 ± 9.1 p = 0.526
ASSD (mm)	1.75 ± 2.67	4.99 ± 1.88	0.93 ± 0.86	1.07 ± 0.63	1.12 ± 1.01 p = 0.939

Abbreviations: AAo: ascending aorta, DAo: descending aorta, DSC: Dice similarity coefficient, HD: Hausdorff distance, ASSD: average symmetrical surface distance. P-value for a Student t-test computed between our method and Berhane et al. (28).

Table 4: Comparison between the proposed method segmentation results and the inter-observer's variability.

	Inter-observer study (n = 25)		
	COD	IOD	WI (95% CI)
DSC	0.87 ± 0.03	0.87 ± 0.04	1.02 (1.01, 1.04)
HD	19.36 ± 12.18 mm	21.95±14.15 mm	1.04 (1.01, 1.08)
ASSD	0.15 ± 0.057 mm	0.16 ± 0.061 mm	1.03 (1.01, 1.05)

Abbreviations: DSC: Dice similarity coefficient, HD: Hausdorff distance, ASSD: average symmetrical surface distance, COD: mean Deep Learning to human observers difference, IOD: mean difference between human observers, WI: William index, CI: confidence interval.

Table 5: Comparison of deep learning-based aortic morphological measures against reference measurements within the testing dataset (n = 76).

	Ground truth	Results	R	p-values	μ [LoA]
Aortic arch length (mm)	129.46±29.36	129.69±27.99	0.95	0.72	1.71 [-16.32; 19.74]
AAo maximal Diameter (mm)	32.18±7.79	33.20±8.10	0.95	0.42	1.15 [-3.54; 5.83]
DAo maximal Diameter (mm)	24.24±5.06	25.32±4.62	0.96	0.07	1.28 [-1.27; 3.83]
AAo Volume (mL)	47.52±31.62	46.43±32.10	0.95	0.91	-0.27 [-22.43; 23.00]
DAo Volume (mL)	63.59±33.04	65.77±29.54	0.97	0.29	4.39 [-8.47; 17.26]

Abbreviations: AAo: ascending aorta, DAo: descending aorta, LoA: limits of agreement.

P-values are computed for comparison between methods using a Kruskall-Wallis test, r is the Spearman correlation coefficient and mean bias [limits of agreement] are given for the Bland-Altman analysis.

Table 6: Bland-Altman analyses for comparisons between DL-based and reference aortic morphological measures, per vendors, field strength, aortic dilation status within the testing dataset (n=76).

		Vendor		Field strength		Aortic dilation status	
		Vendor 1 (53)	Vendor 2 (23)	1.5 T (46)	3T (30)	Dilated aorta (28)	Small to normal aorta (48)
Arch length (mm)	p-values	< 0.0001		<0.0001		0.23	
	μ(μ %)	4.62(4.0%)	-4.99(-3.5%)	5.63(4.7%)	-4.30(-2.9%)	0.03(-0.2%)	2.69(2.8%)
	LoA	[-11.54:20.77]	[-20.27:10.27]	[-7.04:18.29]	[-23.13:14.53]	[-20.96:21.01]	[-13.30:18.67]
AAo maximal Diameter (mm)	p-values	0.07		0.78		0.75	
	μ(μ%)	0.71(2.8%)	2.15(6.3%)	0.97(3.5%)	1.42(4.3%)	1.17(3.0%)	1.34(4.4%)
	LoA	[-3.50:4.92]	[-3.07:7.36]	[-2.87:4.81]	[-4.35:7.20]	[-5.28:7.62]	[-2.20:4.67]
DAo maximal Diameter (mm)	p-values	0.43		0.26		0.13	
	μ(μ%)	1.18(5.6%)	1.51(6.3%)	1.42(6.5%)	1.08(4.7%)	0.93(3.8%)	1.49(7.0%)
	LoA	[-1.42:3.79]	[-0.90:3.92]	[-0.70:3.53]	[-2.01:4.17]	[-2.18:4.04]	[-0.60:3.57]
AAo Volume (mL)	p-values	0.10		0.63		0.27	
	μ(μ%)	-0.75(-0.2%)	2.65(5.7%)	0.67(3.0%)	-0.32(-0.7%)	-1.26(-3.2%)	1.18(4.3%)
	LoA	[-24.07:22.58]	[-18.33:23.62]	[-22.24:23.59]	[-23.04:22.41]	[-37.99:35.48]	[-5.07:7.42]
DAo Volume (mL)	p-values	0.10		0.0018		0.27	
	μ(μ%)	5.15(10.9%)	2.65(3.5%)	6.12(12.6%)	1.75(2.6%)	3.51(4.5%)	4.91(11.1%)
	LoA	[-4.69:14.99]	[-15.19:20.50]	[--3.66:15.90]	[-13.55:17.05]	[-11.10:18.11]	[-6.87:16.69]

Abbreviations: AAo: ascending aorta, DAo: descending aorta, LoA: limits of agreement

P-values are computed for comparison between two subgroups. μ and LoA mean Bland and Altman Bias and limits of agreement, μ% mean Bland and Altman bias expressed in percentage of the reference measures.

FIGURE LEGENDS

Figure 1: Overview of the DL-based method for 4D aorta quantitative analysis.

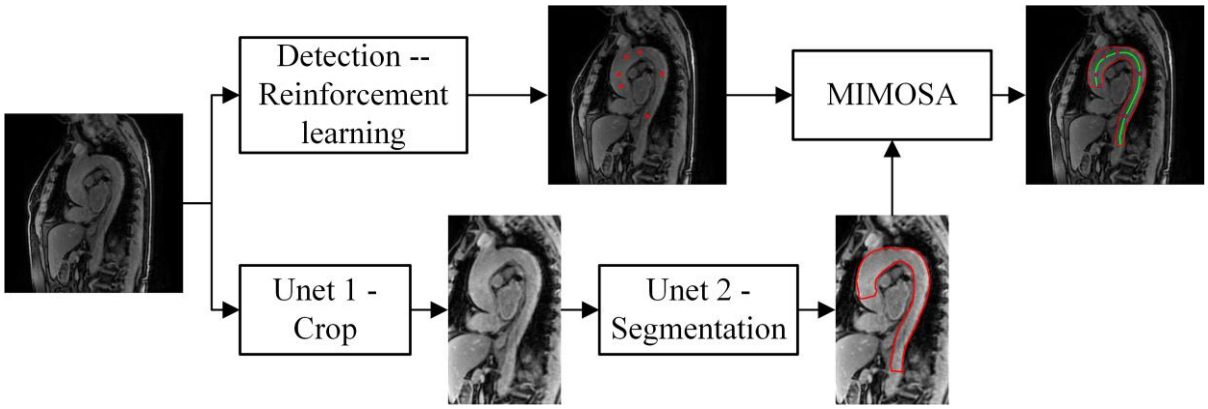


Figure 2: Architecture of the network used for anatomical landmark detection.

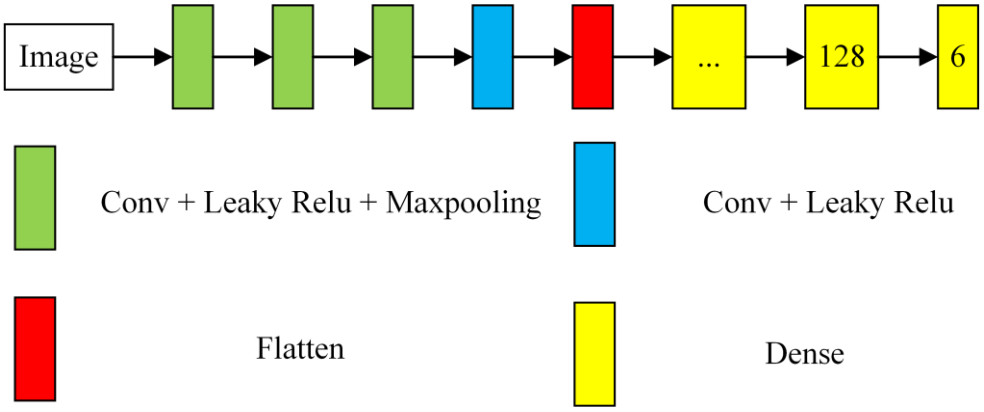


Figure 3: Architecture of the proposed 2-stage network. Top row: U-net producing the ROI around the aorta, bottom row: U-net for aortic lumen segmentation.

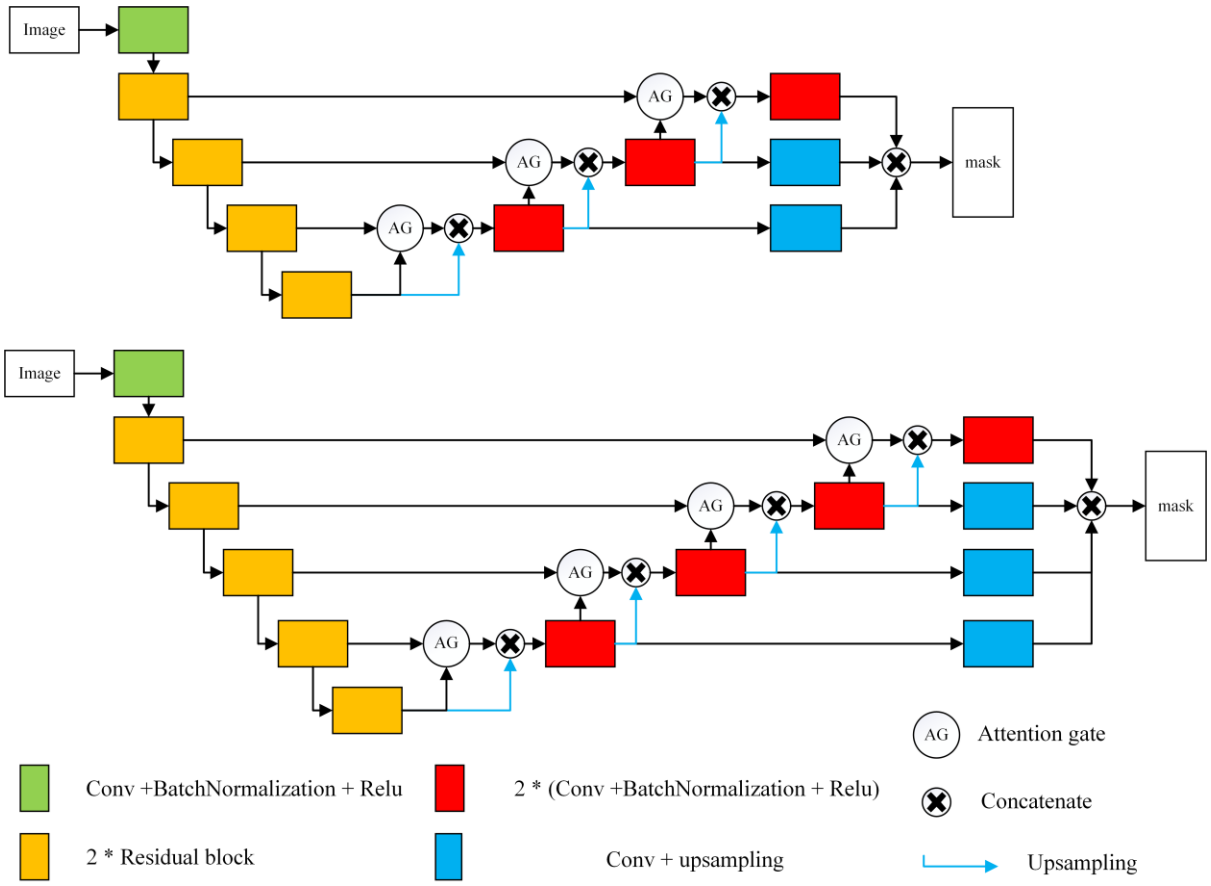


Figure 4: Illustration of the aortic morphological indices.

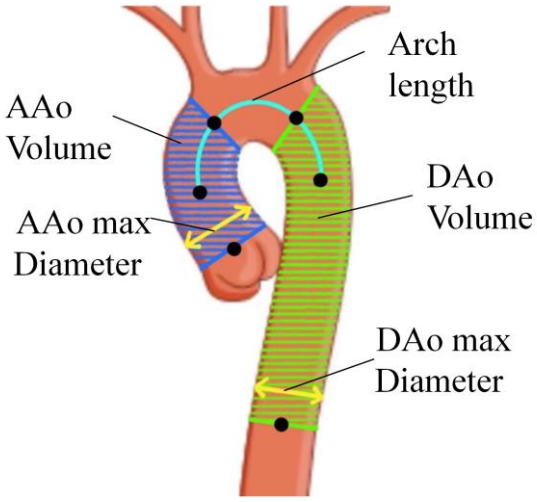


Figure 5: Landmarks detection performances on the test set (n = 78) (left); Comparison between the DL-based method and the human observers landmarks in terms of variability (right).

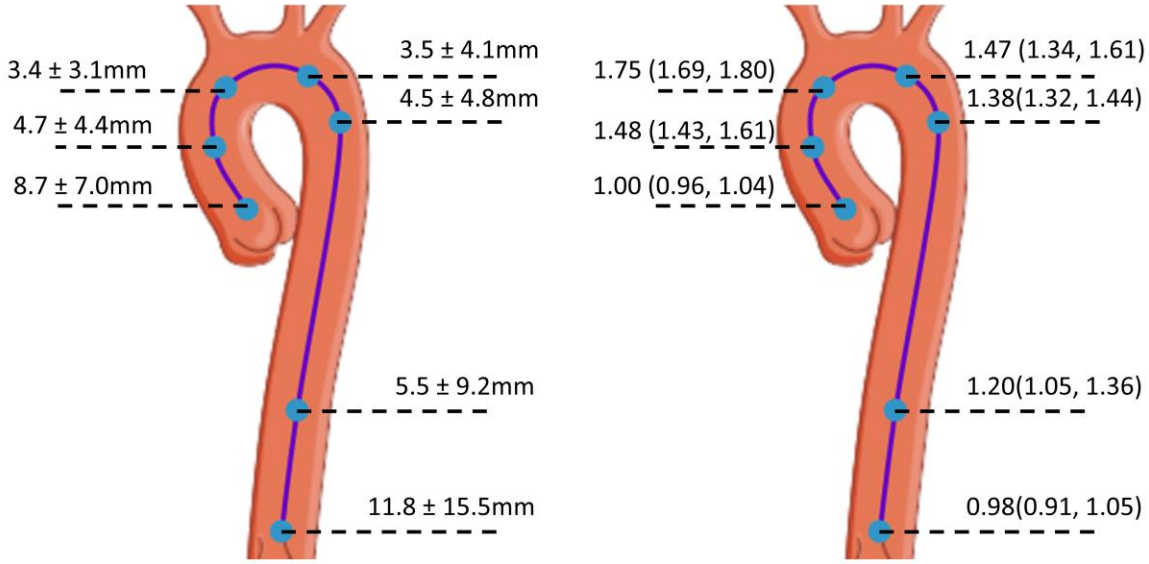


Figure 6: Examples of the data with low contrast (left) or signal heterogeneity (right). The red (yellow) contour correspond to the reference (proposed method) segmentation.

Figure 7: Box plot representation of the Dice score (DSC) of the proposed segmentation results for different vendors (left), magnetic field strengths and for dilated versus small to normal sized aortas (right).

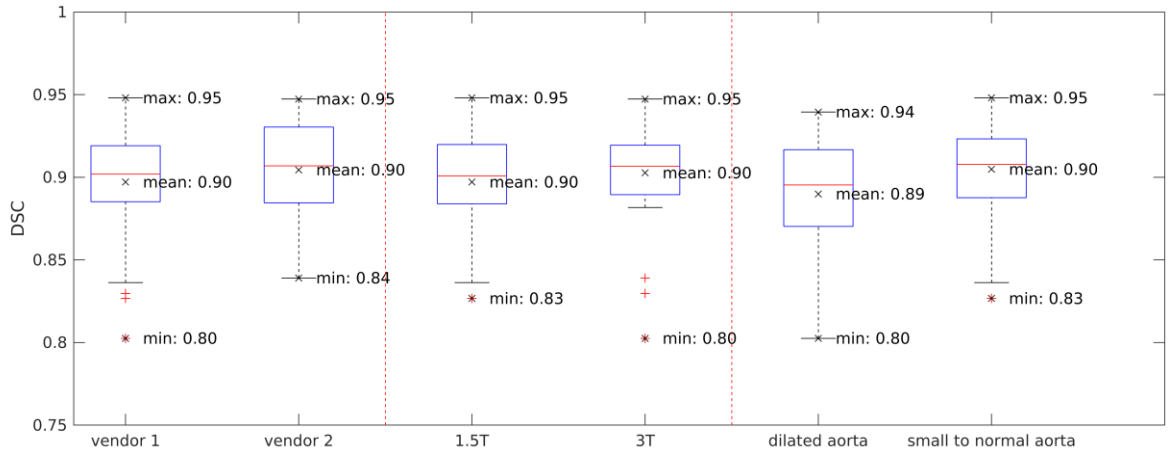
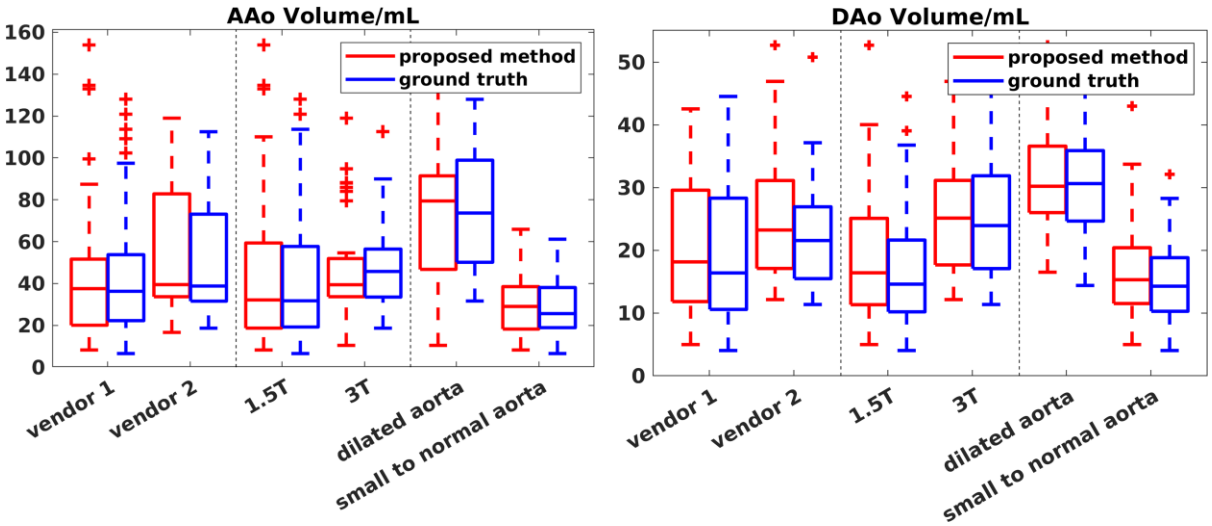


Figure 8: Box plot representation of ascending aorta (AAo) (left) and the descending aorta (DAo) (right) volume of the proposed method (red) and the reference (blue) for different vendors, magnetic field strengths and aortic size status.



Supplementary information 1

Table R1: Subject characteristics for the whole population and for each pathology

	Overall	Healthy subject	Hypertensives	TAV-ATAA	BAV	TS
N (M / F)	391 (203/183)	119 (61/58)	96 (55/41)	67 (50/17)	49 (40/9)	61 (0/61)
Age (years)	51.9 ± 18.4	48 ± 14.3	61.9 ± 14.5	64.3 ± 11.9	52.5 ± 16.8	29.5 ± 11.8
BMI ($kg.m^{-2}$)	24.5 ± 4.4	23.7 ± 2.9	25.7 ± 4.0	25.9 ± 4.3	24.5 ± 4.1	23.9 ± 4.7
BSA (m^2)	1.76 ± 0.27	1.76 ± 0.19	1.83 ± 0.21	1.89 ± 0.20	1.91 ± 0.18	1.48 ± 0.16
cSBP (mmHg)	119 ± 16	113 ± 13	127 ± 12	126 ± 17	119 ± 16	106 ± 12
cDBP (mmHg)	81 ± 11	79 ± 9	84 ± 11	82 ± 11	82 ± 10	78 ± 12
cPP (mmHg)	38 ± 11	34 ± 7	43 ± 10	42 ± 17	37 ± 12	27 ± 4
Reference aortic morphological measures						
Arch length (mm)	129.8 ± 28.0	120.2 ± 19.8	133.8 ± 21.6	156.4 ±24.4	146.7 ± 25.9	99.5 ± 15.0
AAo maximal diameter (mm)	32.6 ± 7.9	28.8 ± 4.0	32.0 ± 4.3	41.2 ± 4.9	41.7 ± 6.9	24.2 ± 5.0
DAo maximal diameter (mm)	24.4 ± 4.9	23.0 ± 3.2	25.8 ± 4.1	29.2 ± 3.8	26.0 ± 4.1	18.5 ± 2.9
AAo volume (mL)	49.9 ± 33.9	32.4 ± 12.4	43.0 ± 15.1	89.3 ± 28.8	90.2 ± 36.8	19.2 ± 10.5
DAo volume (mL)	64.6 ± 32.9	55.0 ± 20.5	68.5 ± 25.0	101.6 ± 33.0	76.6 ± 28.3	27.2 ± 9.8

Abbreviations: TAV-ATAA: ascending thoracic aorta aneurysm and tricuspid aortic valve, BAV: bicuspid aortic valve, TS: Turner syndrome, BMI: body mass index, BSA: body surface area, cSBP: central systolic blood pressure, cDBP: central diastolic blood pressure, cPP: central pulse pressure, AAo: ascending aorta, DAo: descending aorta.

Supplementary information 2

The average disagreement $D_{j,j'}$ between the two observers j and j' is defined as $D_{j,j'} = \frac{1}{N}\sum_{i=1}^{N} e\left(x_{ij}, x_{ij'}\right)$ and $WI = \frac{\frac{1}{n}\sum_{j=1}^{N}\frac{1}{D_{0,j}}}{\frac{2}{n(n-1)}\sum_{j}\sum_{j':j'\neq j}\frac{1}{D_{j,j'}}}$ The CI for this index is estimated using a jackknife non-parametric sampling technique where the standard error of the Willams index is given by

$$se = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left[WI'_{(i)} - \frac{1}{N}\sum_{i=1}^{N} WI'_{(i)}\right]}$$

where $WI'_{(i)}$ is the Williams index computed without the *i*-th measurement.