European Radiology
EXPERIMENTAL

EUROPEAN SOCIETY
OF RADIOLOGY

ORIGINAL ARTICLE

Open Access

Deep learning-aided extraction of outer aortic surface from CT angiography scans of patients with Stanford type B aortic dissection

Risto Kesävuori[1*] , Tuomas Kaseva[1], Eero Salli[1], Peter Raivio[2], Sauli Savolainen[1,3] and Marko Kangasniemi[1]

Abstract

Background Guidelines recommend that aortic dimension measurements in aortic dissection should include the aortic wall. This study aimed to evaluate two-dimensional (2D)- and three-dimensional (3D)-based deep learning approaches for extraction of outer aortic surface in computed tomography angiography (CTA) scans of Stanford type B aortic dissection (TBAD) patients and assess the speed of different whole aorta (WA) segmentation approaches.

Methods A total of 240 patients diagnosed with TBAD between January 2007 and December 2019 were retrospectively reviewed for this study; 206 CTA scans from 206 patients with acute, subacute, or chronic TBAD acquired with various scanners in multiple different hospital units were included. Ground truth (GT) WAs for 80 scans were segmented by a radiologist using an open-source software. The remaining 126 GT WAs were generated via semi-automatic segmentation process in which an ensemble of 3D convolutional neural networks (CNNs) aided the radiologist. Using 136 scans for training, 30 for validation, and 40 for testing, 2D and 3D CNNs were trained to automatically segment WA. Main evaluation metrics for outer surface extraction and segmentation accuracy were normalized surface Dice (NSD) and Dice coefficient score (DCS), respectively.

Results 2D CNN outperformed 3D CNN in NSD score (0.92 versus 0.90, $p = 0.009$), and both CNNs had equal DCS (0.96 versus 0.96, $p = 0.110$). Manual and semi-automatic segmentation times of one CTA scan were approximately 1 and 0.5 h, respectively.

Conclusions Both CNNs segmented WA with high DCS, but based on NSD, better accuracy may be required before clinical application. CNN-based semi-automatic segmentation methods can expedite the generation of GTs.

Relevance statement Deep learning can speeds up the creation of ground truth segmentations. CNNs can extract the outer aortic surface in patients with type B aortic dissection.

Key points

• 2D and 3D convolutional neural networks (CNNs) can extract the outer aortic surface accurately.

• Equal Dice coefficient score (0.96) was reached with 2D and 3D CNNs.

• Deep learning can expedite the creation of ground truth segmentations.

Keywords Computed tomography angiography, Convolutional neural network, Machine learning, Neural networks (computer), Aortic dissection

---

*Correspondence:
Risto Kesävuori
risto.kesavuori@hus.fi
Full list of author information is available at the end of the article

Springer Open

Kesävuori *et al. European Radiology Experimental*   (2023) 7:35

Page 2 of 13

Graphical Abstract

Deep learning-aided extraction of outer aortic surface from CT angiography scans of patients with type B aortic dissection

- **2D and 3D convolutional neural networks (CNNs) can extract the outer aortic surface accurately.**

- **Equal Dice coefficient score (0.96) was reached with 2D and 3D CNNs.**

- **Deep learning can expedite the creation of ground truth segmentations.**

  2D CNNs (blue) 3D CNNs (yellow) can extract the outer aortic surface accurately with high Dice coefficient scores. However, CNNs may occasionally fail the discrimination between the aortic wall and periaortic structures, especially when the periaortic structures have similar attenuation values with the aortic wall. Ground truth segmentations in red.
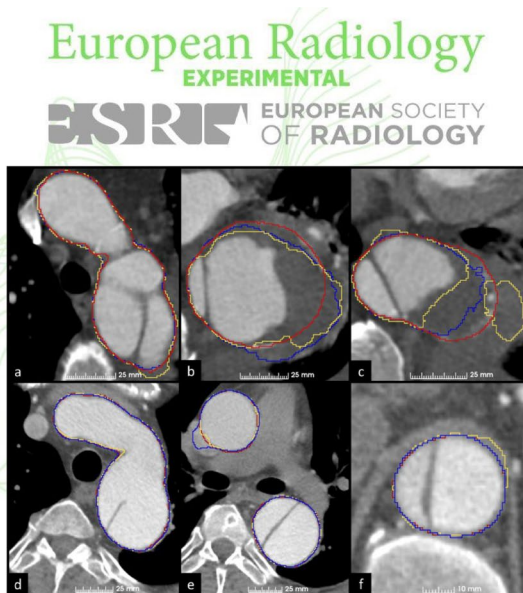
**Deep learning can speeds up the creation of ground truth segmentations.
CNNs can extract the outer aortic surface in patients with type B aortic dissection.**

Eur Radiol Exp (2023) Kesävuori R et al. DOI: 10.1186/s41747-023-00342-z

Background

Aortic dissections require prompt diagnosis and treatment to prevent aortic rupture and other major complications. In aortic dissection, blood enters the aortic wall through a tear in the inner layer of the aorta forming a false lumen inside the wall. Aortic dissections that do not involve the ascending aorta are classified as Stanford type B dissections (TBAD) [1, 2]. Follow-up imaging aims to find patients, who are at risk of developing complications. Reliable tools for aortic diameter measurements are required, because large diameter and fast growth rate of the aorta or false lumen are major risk factors for complications, and these factors guide surgical decision-making [1, 3]. Computed tomography angiography (CTA) is the primary imaging modality in aortic dissection, and aortic diameters should be measured perpendicular to its long axis using multiplanar reconstruction [1–5]. Manual aortic dimension measurements have suboptimal inter- and intra-rater reproducibility [3, 4, 6–13]. Therefore, observed aortic diameter changes of ≤ 3–5 mm during follow-up should be interpreted with caution [3].

Aortic dimensions can be measured between the external (outer-to-outer wall) or internal (luminal) surfaces of the aortic wall, thus including or excluding the wall, respectively. On high-quality CTA scans, the aortic wall is often visible as a thin line encircling the contrast-enhanced lumen, as depicted in Fig. 1. The thickness of the aortic wall is typically from 1.5 to 2.5 mm, and it tends to thicken with aging [14, 15]. In aortic dissection, outer-to-outer wall measurements are preferred, because luminal measurements omit false lumen thrombosis and intramural hematoma (IMH) and may therefore underestimate the maximal diameter of the aorta, as depicted in Fig. 1 [2, 3, 5]. Recently published Society for Vascular Surgery and Society of Thoracic Surgeons reporting standards for type B aortic dissections and the 2010 American College of Cardiology Foundation/American Heart Association guidelines both recommend outer-to-outer wall dimension measurements in aortic dissection imaging [2, 5]. Similarly, a recent statement by American Heart Association recommends using the outer aortic contour as the landmark for aortic dimension measurements in patients with mural thrombus [3].

To enable automatic outer-to-outer wall aortic dimension measurements and volumetric analysis of the aorta, the whole aorta (WA) including the aortic wall needs to be segmented. Automatic segmentation
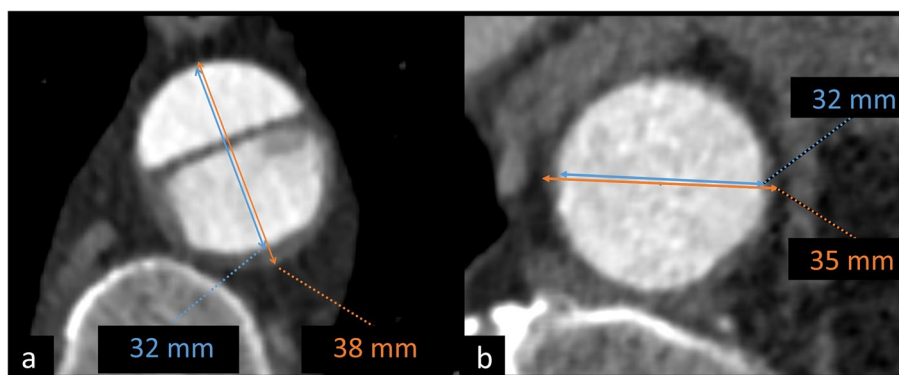
Fig. 1 The differences between luminal (blue arrow) and outer-to-outer wall (orange arrow) dimension measurements. **a** TBAD patient with partly thrombosed false lumen. **b** Intact aorta of a TBAD patient with slight wall thickening at the level of diaphragm. *TBAD* Stanford type B aortic dissection

of abdominal and thoracic aorta including the aortic wall has been studied previously [16–21]. While most of the works concentrating on the automatic segmentation of aortic dissection focus on the segmentation of true and false lumina [22–30], a few studies segmenting WA using convolutional neural networks (CNN) do exist [31–35]. However, the datasets in these studies are small, they comprise mainly of patients with other aortic diseases than aortic dissection, the imaging data is significantly preprocessed before CNN training, or the accuracy of outer aortic wall delineation in the manual segmentations is not described in detail [31–35].

In this study, we present a dataset consisting of 206 TBAD CTA scans with fully segmented ground truth (GT) WAs and high variation in TBAD imaging findings. The techniques and challenges faced in the manual and semi-automatic segmentation of WAs are discussed. Two-dimensional (2D) and three-dimensional (3D) CNNs are trained to segment the WAs automatically and evaluated both on the tasks of segmentation of WA and extraction of the outer aortic surface.

Methods

The Helsinki University Hospital's ethical committee approved this retrospective study, and patients' informed consent was waived.

Dataset and segmentation process

A total of 240 consecutive patients diagnosed with TBAD at our institution from January 2007 to December 2019 were retrospectively reviewed for this study. Patients, who did not have any aortic CTA scans or who had only > 3-mm axial reformats available in our institution's picture archiving and communication systems,

were excluded from this study. From each patient, one CTA scan with TBAD imaging findings was included in the dataset. In an aim to include more complex imaging findings of TBAD to the dataset, the latest available CTA scan was often selected, because false lumen thrombosis and aneurysmal degeneration usually develop over time [3]. This resulted in a dataset of 206 aortic CTA scans with diverse TBAD imaging findings acquired between 2007 and 2020.

In order to expedite the segmentation process, GT WAs for the CTA scans were generated in two phases as described in Fig. 2:

(1) Manual segmentation: A radiologist manually segmented 80 scans chosen randomly from the 206 scans; the segmented scans were randomly divided to form a test set and an initial training set with 40 scans in each set.

(2) Semi-automatic segmentation: The initial training set was used to train a 3D CNN ensemble, which automatically produced preliminary segmentations for the remaining 126 scans; these segmentations were subsequently corrected by the radiologist. The final dataset was then formed including 40 scans for testing and 166 scans for training.

Manual segmentation

Manual WA GT segmentations were created for 40 scans in the test set and 40 scans in the training set. The external aortic wall surface contours from the aortic valve to the common iliac arteries were annotated in every axial slice of each image volume by a radiologist with over 5 years of experience in aortic imaging to create the WA segmentations. The annotations were performed using 3D Slicer's Segment Editor module [36].
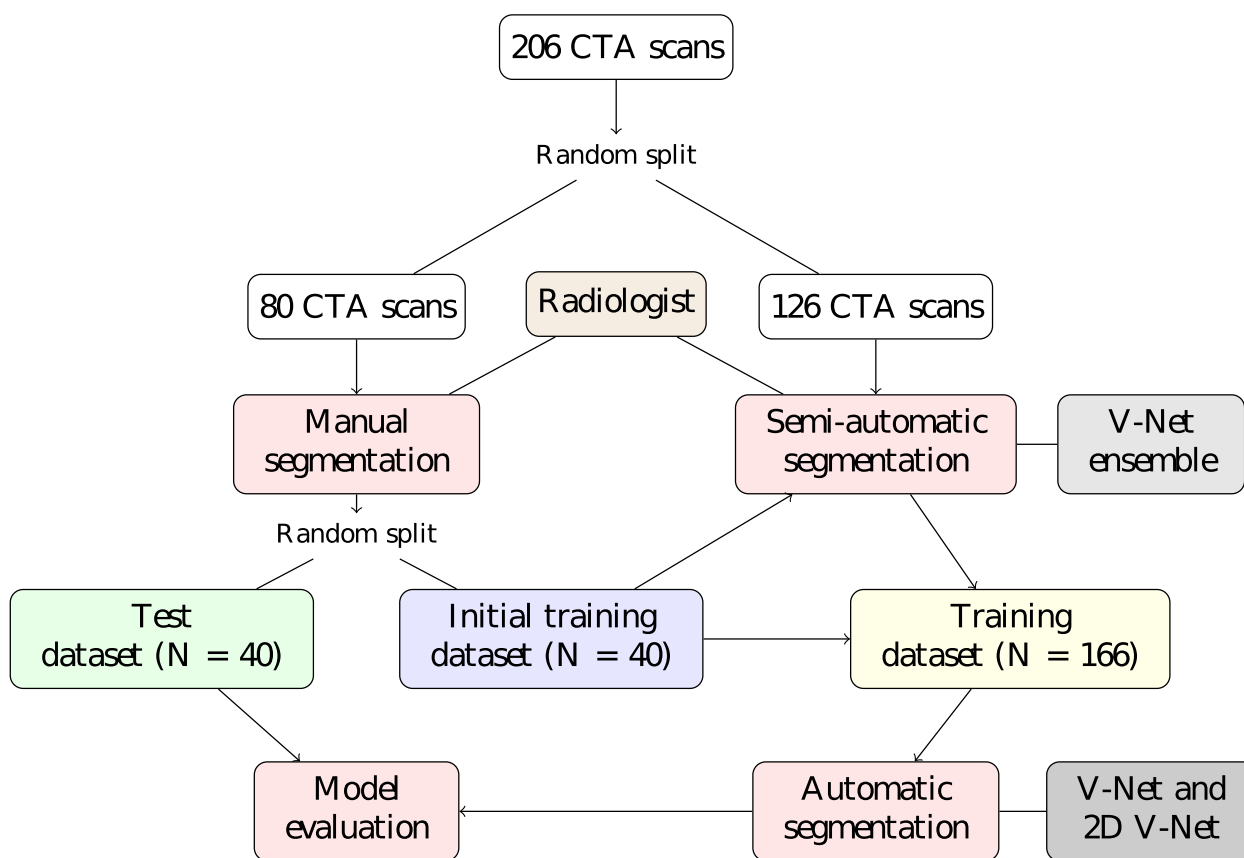
Fig. 2 Diagram of the study: 206 CTA scans from 206 patients with TBAD were split randomly to subsets of 80 and 126 scans. Ground truth WAs from the first subset were manually segmented by a radiologist using 3D Slicer. GT WAs of the second subset were generated via a semi-automatic segmentation process. In this process, initial training set, consisting of randomly chosen 40 scans from the first subset with manually segmented WAs, was utilized to train V-Net ensemble. The ensemble produced initial segmentations for the second subset. These segmentations were corrected by the radiologist and combined with the initial training set to form a training set of 166 studies. V-Net and 2D V-Net were trained using the training set and evaluated over the test set of 40 studies on tasks of automatic segmentation of WAs and extraction of the outer walls. *CTA* Computed tomography angiography, *TBAD* Stanford type B aortic dissection, *WA* Whole aorta

The outer aortic wall surface contours were drawn using a commercially available pen display (Wacom Cintiq Pro 16, Wacom Co. Ltd., Kazo, Saitama, Japan, or Microsoft Surface Pro 8, Microsoft Corporation, Redmond, WA, USA) to produce robust WA segmentations. When the aorta was surrounded by fat tissue, intensity-based drawing methods were used when appropriate to include the outermost aortic wall voxels (higher density than fat) to the WA segmentations. Coronal and sagittal reformats or multiplanar reconstruction was used to confirm precise annotation of the external aortic wall surface when necessary. Special attention was paid to drawing of aortic segments that run along with denser structures (*e.g.*, diaphragm, pleura, or inferior vena cava) to reliably exclude periaortic structures from the segmentations. Coronary, supraaortic, or visceral arteries were not included in the segmentations. Supraaortic arteries were segmented to the level of branching from the aorta in the axial plane.

At the coronary and visceral artery ostia, the WA segmentations were interpolated as a continuation of the outer aortic surface. Aortic wall calcifications were included in the WA segmentations. After completing the manual segmentation process, the 40 test set scans were re-examined, and their GTs were corrected if necessary. The average duration of manual and semi-automatic segmentations was estimated based on time used to complete 10 consecutive segmentations including file saving and opening of a new CTA scan for segmentation.

Figures 3 and 4 present examples of manual GT segmentations of two patients with different false lumen morphologies. In Fig. 3, false lumen thrombosis and pleural fluid have nearly similar intensity values, which complicates the extraction of the outer aortic surface (Fig. 3a and d). Coronal and sagittal reformats were used to confirm correct outer surface annotation. Figure 4 describes more simple segmentation task in a TBAD patient with

Kesävuori *et al. European Radiology Experimental*   (2023) 7:35

Page 5 of 13
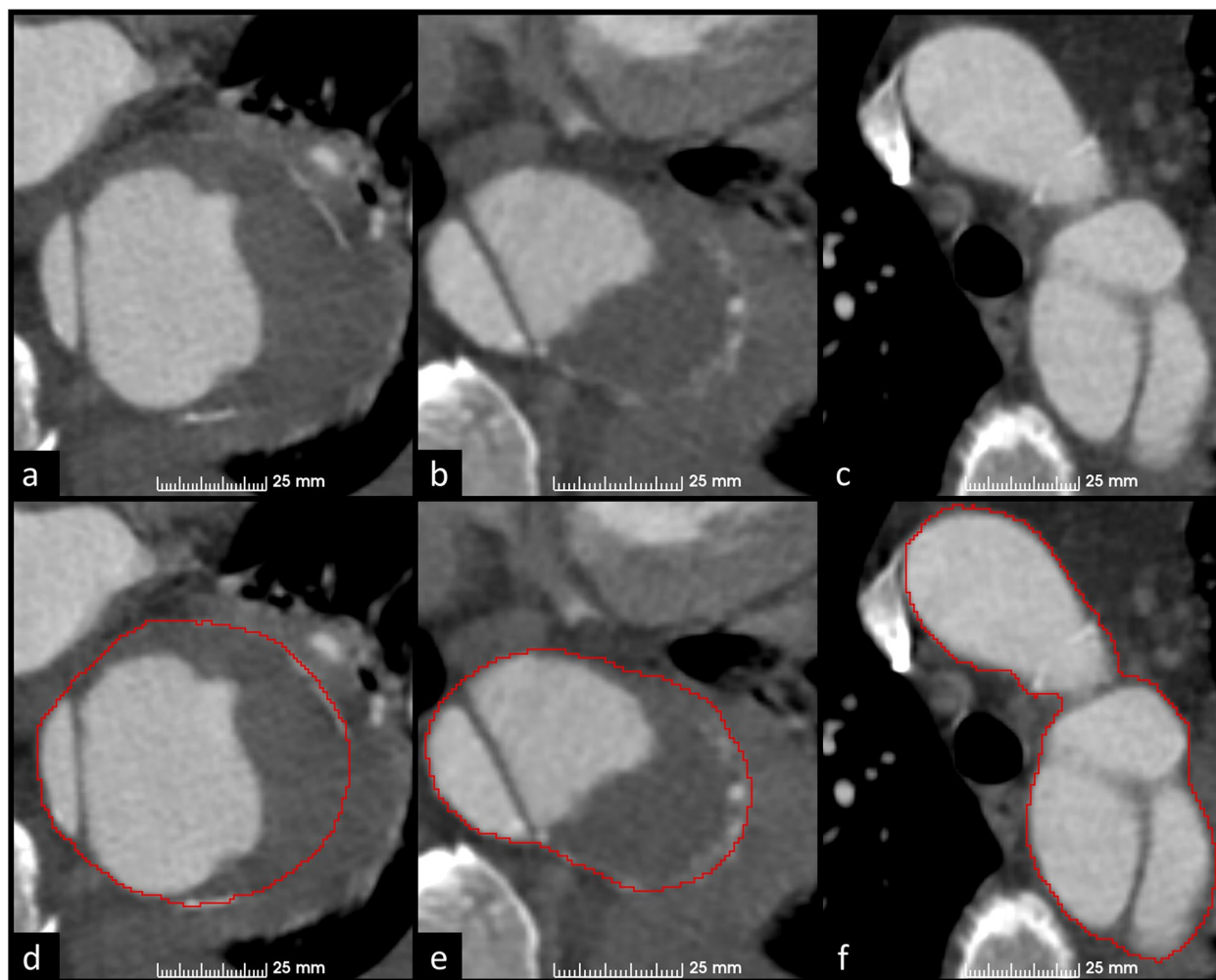
Fig. 3 Manual WA segmentations of a patient with pleural fluid and false lumen thrombosis at the level of descending aorta (**a**, **b**, **d**, and **e**) and aortic arch (**c**, **f**). Delineation of the outer aortic surface is obscured due to similar intensities of pleural fluid and false lumen thrombosis. Multiplanar reconstruction was used to precisely annotate the outer aortic surface margin. *WA* Whole aorta

open false lumen, but motion artifact in ascending aorta obscures the aortic surface margins and leads to imperfect outer surface annotation (Fig. 4b and e).

Semi-automatic segmentation

Forty manually created WA GT segmentations were used to form the initial training set. Then, five different V-Net models were trained by dividing the initial training set to five unique folds with each fold having 32 training and eight validation scans. Training was performed using MONAI, an open-source framework based on PyTorch [37, 38]. The architecture of each V-Net was the same and followed the default V-Net configuration given in the framework. This architecture was based on the one introduced by Milletari et al. [39] but included batch normalization before each parametric rectified linear unit activation function, dropout before each encoder and

decoder block, and another dropout applied on the forwarded fine-grained features. The dropout probability of both dropouts was 0.5.

The patch size was $192 \times 192 \times 48$ voxels. Prior to the extraction of the patches, the intensities of the volumes were scaled from [-50, 600] to [0, 1] with clipping the values outside the initial range to the output range. Spacings were normalized to $1.0 \times 1.0 \times 3.0$ mm. Batch size was 8 with 4 patches collected from two different scans. In each batch, the patches were randomly chosen, but so that half of the patches had centers as foreground voxels and the other half centers as background voxels. The optimizer was Adam and the learning rate 0.0001. The loss function of each model consisted of summation of Dice coefficient score (DCS) and binary cross entropy. The total number of iterations was 16,000. Validation was performed every 128 iterations, and the validation metric was average

Kesävuori *et al. European Radiology Experimental*   (2023) 7:35

Page 6 of 13
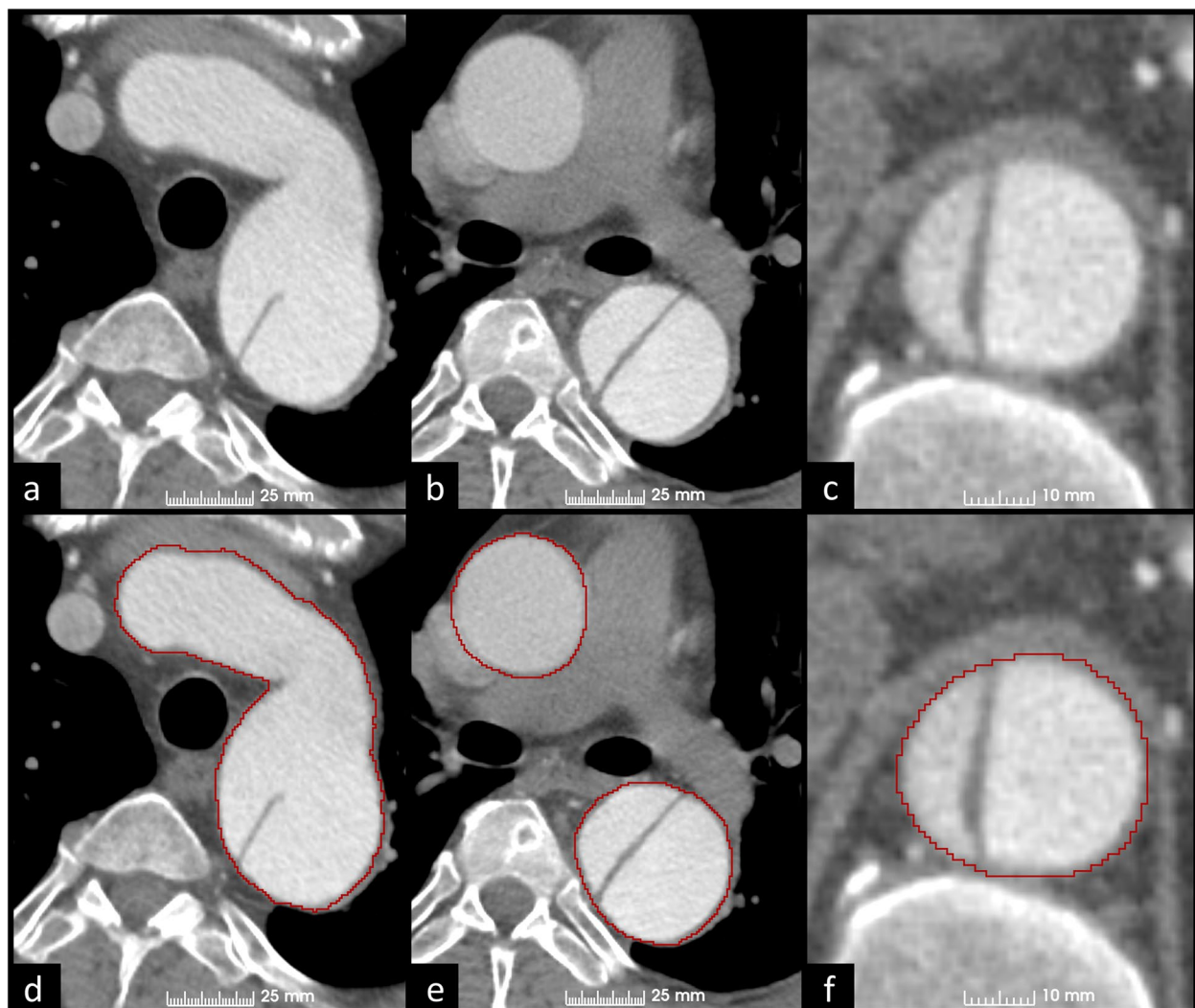
Fig. 4 Manual WA segmentation of a patient with TBAD at the level of aortic arch (**a**, **d**), ascending aorta (**b**, **e**), and diaphragm (**c**, **f**). Notice the motion artifact in the ascending aorta due to non-ECG-gated imaging obscuring the aortic surface margins and leading to imperfect outer surface annotation. *ECG* Electrocardiogram, *TBAD* Stanford type B aortic dissection, *WA* Whole aorta

DCS over all image volumes in the validation set. The model configuration with the highest validation metric was chosen as the final model.

The models performed inference on the 126 scans which were left out from the manual segmentation process. For each scan, five different predictions from each model were generated with sliding window inference technique using the same patch size as in training and 25% overlap. The predictions were binarized by applying the argmax operation on the output of the model. After the binarization, value 1 defined aortic voxels and value 0 other voxels.

Five different ensemble segmentations were formed using these binarizations: In the ensemble N, a voxel was marked as aorta voxel if it belonged to aorta at least in N of the binarized predictions. Thus, in the ensemble 1, the aorta was the combination (union) of all five predictions, ensemble 3 was the majority voting ensemble, and in the ensemble 5, only the aorta voxels that overlapped in all five predictions were marked to aorta. One ensemble segmentation of each scan was visually chosen to be manually corrected. In most cases, majority of voting ensemble was chosen, but occasionally, another ensemble with more suitable predictions was selected. The segmentations were preliminary corrected by a researcher with 1 year of experience on the manual aortic segmentation process and subsequently corrected by the radiologist with experience of aortic imaging using 3D Slicer's Segment Editor module. The corrections were performed with similar methods and robustness as manual segmentations.

Automatic segmentation

Two neural network models, 2D V-Net and V-Net, were trained for automatic segmentation of the WA. The input of the first consisted of 320×320 sized axial slices, whereas the input of the second is of $224 \times 224 \times 64$ sized patches. V-Net had the same architecture as discussed in semi-automatic segmentation, but the number of filters of V-Net started from 12 instead of 16 proposed in Milletari et al. [39]. 2D V-Net was simply a 2D version of the V-Net model used in the semi-automatic segmentation. The dropout probabilities for both models were set as 0.4. Both models had the same 136 scans used for training and 30 scans used for validation. The validation set was chosen randomly.

Before extracting the inputs, the spacings of image volumes were normalized similarly as in semi-automatic segmentation, but the scaling was performed from $[-500, 1,500]$ to $[0, 1]$. The batch size was 32 for 2D V-Net with 16 slices extracted from two scans. The size was 4 for V-Net with 2 patches extracted from two scans. In each batch, the choice of samples was performed similarly as in the semi-automatic segmentation. The optimizer and the loss function were also the same as used in the semi-automatic segmentation. The learning rate was 0.001 and halved after 81,600 iterations. The total number of iterations was 136,000 for 2D V-Net and 125,800 for V-Net. During training, axial rotations up to 45° were performed to every second sample of the batch randomly as augmentation. The validation process was the same as explained in the semi-automatic segmentation but performed after every 3,400 iterations for both 2D V-Net and V-Net.

Model evaluation

Both 2D V-Net and V-Net performed inference on the scans in the test set using the same sliding window technique utilized in semi-automatic segmentation. The patch size of V-Net was $416 \times 416 \times 64$ and the dimensions of the axial slices of 2D V-Net 416×416. The evaluation metrics used were DCS, Hausdorff distance (HD, mm), normalized surface Dice (NSD), and mean surface distance (MSD, mm) [40]. The HD was calculated using 95th percentile of the distances. DCS is a measure of the overlap between the prediction and the GT segmentations, HD measures the maximum distance between the surfaces of segmentations, and MSD measures the average distance. NSD measures overlap between the surfaces of the prediction and GT. The aorta voxels which had at least one non-aorta voxel in the first-order 3D neighborhood were defined as surface voxels. The tolerance parameter of the NSD specifies the maximum difference in the boundary that is tolerated without penalty in the metric. For NSD, 1-mm tolerance level was selected. Results are reported as median values and interquartile ranges calculated over the studies. Wilcoxon signed-rank test for the zero median for the difference of paired samples of V-Net and 2D V-Net was conducted, and corresponding p-values were calculated.

Results

The datasets and demographic characteristics are presented in Table 1, and the performance scores of the models are presented in Table 2, which includes the medians and interquartile ranges of the evaluation

Table 1 Description of the datasets

	Training dataset ($n = 166$)	Test dataset ($n = 40$)	*p*-value*
Age, years	71 (59–79)	66 (53–74)	0.049
Male, *n* (%)	49 (30)	9 (23)	0.438
Hyperacute dissection (initial CT), *n* (%)	53 (32)	12 (30)	0.852
Acute or subacute dissection (1–90 days), *n* (%)	47 (28)	13 (33)	0.698
Chronic dissection (> 90 days), *n* (%)	66 (40)	15 (38)	0.858
Intraluminal or intramural thrombosis in at least two aortic zones**, *n* (%)	134 (81)	31 (78)	0.662
Classic dissection, *n* (%)	26 (62)	27 (68)	0.716
IMH, *n* (%)	60 (36)	13 (33)	0.716
CT-scanner vendor, *n* (%)			
Siemens	134 (81)	31 (78)	0.662
GE Medical Systems	26 (16)	7 (18)	0.812
Toshiba	6 (4)	2 (5)	0.654

Data is presented as counts (percentages) or median (interquartile range). *Fisher's exact test and Mann–Whitney *U*-test for categorical and continuous variables, respectively. **Lombardi et al. (J Vasc Surg 2020, reference [5]). *CT* Computed tomography, *IMH* Intramural hematoma

Table 2 Evaluation scores of V-Net and 2D V-Net over the test set

Model	NSD	HD (mm)	DCS	MSD (mm)
V-Net	0.90 (0.84–0.93)	**2.22** (1.68–3.56)	0.96 (0.95–0.97)	**0.43** (0.31–0.61)
2D V-Net	**0.92** (0.85–0.94)	2.36 (1.49–4.12)	0.96 (0.95–0.97)	0.44 (0.29–0.72)
p-value*	0.009	0.614	0.110	0.125

Each score is presented as median (interquartile range). With NSD and DCS, higher is better, while with MSD and HD, lower is better. The best scores are bolded. *Wilcoxon signed-rank test for the zero median for the difference of paired samples of V-Net and 2D V-Net. *DCS* Dice coefficient score, *HD* 95th percentile Hausdorff distance, *MSD* Mean surface distance, *NSD* Normalized surface dice

metrics computed over the test set with V-Net and 2D V-Net; *p*-values of each metric between the models are also illustrated. The median NSD scores of V-Net and 2D V-Net were 0.90 and 0.92, median HD scores 2.22 mm and 2.36 mm, median DCS 0.96 and 0.96, and median MSD scores 0.43 mm and 0.44 mm, respectively. Only the difference between NSD scores of the models was statistically significant. All metrics evaluating surface delineation, *i.e.*, NSD, HD, and MSD, have relatively high dispersion in values.

In Fig. 5, we present the same axial slices as in Figs. 3 and 4 with GT, V-Net, and 2D V-Net segmentations. Both 2D V-Net and V-Net fail the discrimination between thrombosed false lumen and pleural fluid (Fig. 5b and c). Both CNNs have difficulties delineating ascending aorta
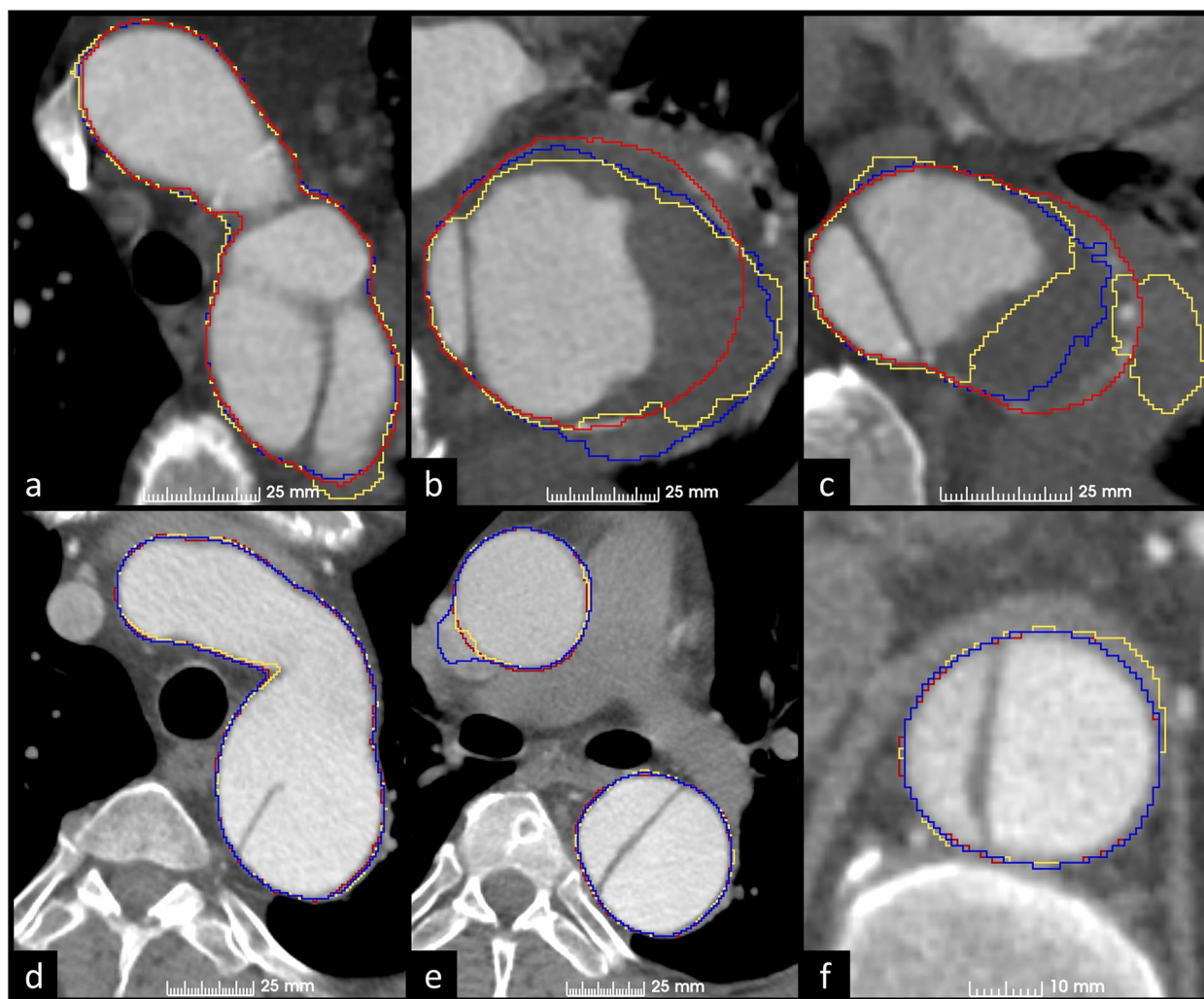
Fig. 5 Examples of automatic segmentations of 2D V-Net (blue) and V-Net (yellow) and the corresponding manual ground truth segmentations (red) on the same axial slices than in Figs. 3 and 4. Both 2D V-Net and V-Net falsely label pleural fluid as WA and omit part of false lumen thrombosis from WA segmentations (**a**, **b**, **c**). Both 2D V-Net and V-Net correctly omit atelectatic lung from WA segmentations (**d**). Motion artifact in the ascending aorta complicates automatic segmentation (**e**); note the correct automatic segmentation of the descending aorta at this level. V-Net labels part of the diaphragm as WA, while 2D V-Net correctly labels the aorta (**f**). *WA* Whole aorta
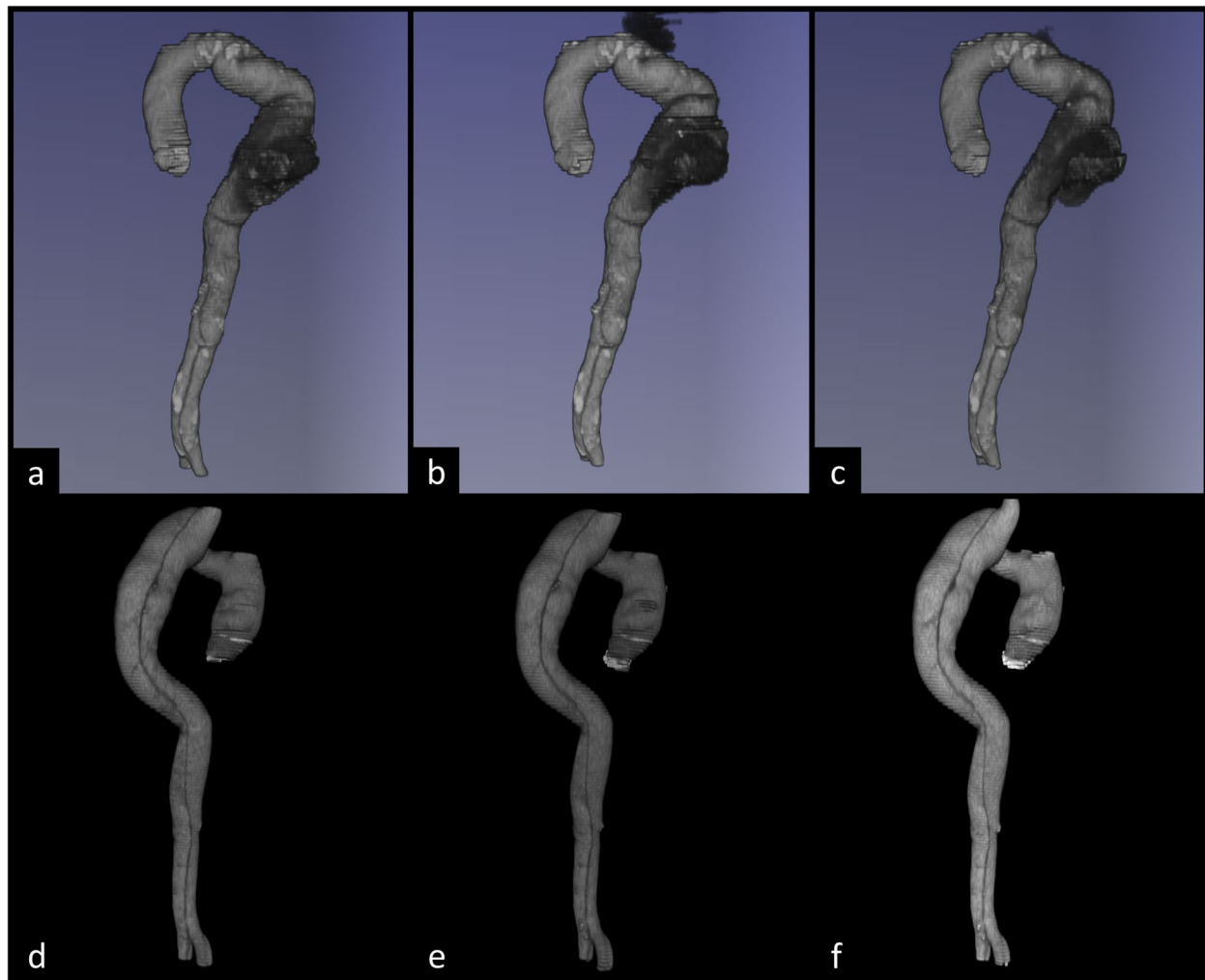
Fig. 6 Three-dimensional representations of ground truth (**a**, **d**), 2D V-Net (**b**, **e**), and V-Net (**c**, **f**) segmentations of two different patients (**a**–**c**, patient from Fig. 3; **d**–**f**, patient from Fig. 4). CNNs variably detect the border between thrombosed false lumen and pleural fluid (**b**, **c**). High agreement between ground truth and CNN segmentations (**d**–**f**). *CNN* Convolutional neural network

with motion artifact (Fig. 5e), whereas V-Net mistakenly adds part of the diaphragm as WA (Fig. 5f). NSD scores over the scan of the first row (Fig. 5a, b, c) were 0.79 with V-Net and 0.73 with 2D V-Net. NSD score over the scan presented in the second row (Fig. 5d, e, f) was 0.96 with both V-Net and 2D V-Net. In Fig. 6, we present the GT, 2D V-Net, and V-Net segmentations in 3D.

On average, the duration of manual segmentation of one scan ranged from 45 to 75 min, and correction of a semi-automatic segmentation lasted between 20 and 40 min. The segmentation times included loading of the scan and saving the segmentation using 3D Slicer. On occasional poor automatic segmentations, the duration of semi-automatic segmentation was comparable to manual segmentation times. The training time of V-Net and 2D V-Net was 160 h and 16 h, respectively, using Tesla V100 GPU. The inference times of the models were on average less than 1 min.

Discussion

This study presented a dataset of 206 CTA scans from 206 patients with fully segmented GT WAs of TBAD including the aortic wall. The study proposed deep learning-based semi-automatic segmentation process to expedite the segmentation process and evaluated the use of 2D- and 3D-based CNNs for segmentation of WA and the extraction of the outer aortic wall surface. Semi-automatic segmentation process was approximately twice as fast as manual segmentation process. 2D and 3D CNNs reached 0.92 and 0.90 median NSD scores over the test set of 40 scans, respectively. The results obtained do not justify automation of WA segmentation but suggest that CNNs can be helpful in semi-automatic segmentation of WA and outer aortic wall extraction.

In this study, we introduce a large and robustly segmented WA dataset with diverse TBAD imaging findings.

Previously, Cao et al. [31] presented a dataset of 276 CTA scans with WA, true lumen, and false lumen GT labels. Their dataset comprised of TBAD patients undergoing TEVAR (thoracic endovascular aortic repair), and the segmentations were produced to meet the requirements for its planning and standard TBAD measurements. The authors segmented WA GTs in intervals and used fill-between-slices tool to fill the gaps between slices. We approached the segmentation process from a different perspective. Our aim was to produce accurate annotations of the outer surface of the aorta in every axial slice in a patient population with diverse set of aortic dissection imaging findings to enable robust dimension measurements of a dissected aorta in the future. The WA dataset of Bratt et al. [32] consisted of over 5,000 scans with various aortic pathologies, including aortic dissections, but the authors explicitly mentioned that the GTs were not originally generated for dimension measurements. Sieren et al. [35] studied automatic segmentation of WAs of healthy and diseased aortas but did not describe how accurately they aimed to delineate the aorta's outer surface. Krissian et al. [33] aimed to carefully segment WAs in their work but segmented only five scans. Wobben et al. [34] reported high DCSs of TL, FL, FL thrombosis, and aorta segmentations using deep learning using preprocessed imaging data for CNN training. The data consisted of 80×80 mm cropped multiplanar reconstructions of the aorta centered along and perpendicular to the aortic centerline. Yu et al. [22] and Yao et al. [23] discussed results on WA segmentation, but in these works, the wall of the aorta was not included in the WA segmentations.

In previous studies by Sieren et al. [35] and Bratt et al. [32], WAs were manually segmented in 30 min on average. In our material, manual segmentation time of one scan was on average 1 h despite using 3-mm axial slices in contrast to Sieren et al. [35] who used 1-mm axial slices in the segmentation process. Sieren et al. [35] and Bratt et al. [32] used semi-automatic tools in their segmentation processes, which may explain the difference in the segmentation times.

We opted not to use semi-automatic tools in the manual segmentation process, since this could have resulted in segmentation imprecisions. Instead, we opted to segment manually 80 scans of our dataset and use deep learning-based semi-automatic segmentation process in the creation of GTs for the remaining 126 scans. The aim of this process was to reduce the manual annotation time used in the creation of GTs. In the process, initial segmentations were first generated via an ensemble of V-Nets and then corrected manually. The use of deep learning as an aiding tool in the creation of GTs has been studied previously [41]. We chose to use

the ensemble to effectively utilize the 40 manually segmented GTs of the initial training set. Via ensembling, all 40 studies of the set could be utilized for both training and validation. The ensemble achieved an average of 0.94 DCS over the ensemble validation sets. In large parts of the aorta, the initial segmentations followed the outer aortic margin correctly, and manual corrections were performed only in the areas with imprecise initial segmentations. The process decreased the segmentation time approximately by half but maintained the quality of the annotations. The test set was segmented manually without using semi-automatic process to ensure that the test set was independent of the training set.

Previous studies have evaluated the performance of automatic WA segmentation methods primarily with DCS metrics, with reported mean DCSs of 0.91 to 0.96 [31–33, 35]. The datasets in these studies have varied in terms of aortic pathology, WA segmentation accuracy, and area of the aorta used for DCS calculation. In only one study, by Cao et al. [31], the dataset has comprised solely of patients with aortic dissection where 0.93 DCS was reached. In our study, 2D V-Net and V-Net achieved 0.96 median DCSs, which is a good agreement considering the variability of TBAD imaging findings in our dataset and high occurrence of motion artifacts at the level of the aortic root and ascending aorta. The performance of the models on the task of extraction of outer aortic surface was quantified using NSD, HD, and MSD scores. With NSD, we selected a tolerance threshold of 1 mm, because contemporaneous 1-mm segmentation errors in opposite sides of the aorta would lead to a 2-mm error in dimension measurements. Such error was considered unacceptable for clinical use.

Median NSD scores of 2D V-Net and V-Net were 0.92 and 0.90. The difference between the scores was deemed statistically significant, while no significant difference was observed between the other scores. Consequently, V-Net was unable to outperform 2D V-Net in this study, although the CTA scans used were inherently 3D. In addition to our study, at least Bratt et al. [32] have reported the applicability of 2D CNNs on the task of segmenting 3D WAs. The usage of 2D CNNs is intriguing since they are significantly faster to train than the 3D CNNs. Through the example shown in Fig. 5, we showed that segmentations with NSD scores as high as 0.96 could include unacceptable mistakes. Thus, the median NSD scores of models are yet too low to justify automation of the outer aortic wall extraction. Most major segmentation mistakes by both CNNs were related to unsuccessful delineation between false lumen thrombosis or IMH and periaortic structures with near similar intensity values, such as pleural fluid,

Kesävuori *et al. European Radiology Experimental*   (2023) 7:35

Page 11 of 13

diaphragm, or large hiatal hernias. On the other hand, the models can be very useful in semi-automatic setting. Sieren et al. [35] reported a 0.99-mm median MSD score on the WA border extraction over a subgroup of scans with patients suffering from aortic dissection. The same score over our test set was 0.44 mm with 2D V-Net and 0.43 mm with V-Net. While the datasets are different, this result comparison highlights the suitability of V-Net type CNN architectures for outer aortic surface extraction.

Evaluation of aortic growth during follow-up and imaging-based risk assessment for aortic rupture after aortic dissection is based on dimension measurements in several areas of the aorta and lumina [1, 3]. The aim of this work was to produce a robust automatic CNN-based method for WA segmentation and to evaluate the segmentation task in detail. The CTA scans and the WA segmentations included variable presentations of aortic, true lumen, and false lumen morphologies, intraluminal thrombosis, and IMH. In the future, we aim to combine our WA segmentation approach with lumen segmentations and proceed to compute the measurements.

This is a retrospective study with limitations. Although the GT segmentations in this study were performed by a radiologist experienced in aortic imaging, the outer wall surface annotations in the datasets may occasionally be imperfect. Additionally, we did not evaluate GT segmentation quality or agreement between radiologists. Albeit we included a large diversity of aortic dissection CTA scans with varying morphologic findings, imaging parameters, and image quality, a larger training set would have produced a more thorough set of aortic dissection imaging findings. Therefore, rare findings in the test dataset not presented at the training set were not adequately labeled by the final neural network.

In conclusion, we introduced a novel, diverse dataset of WA segmentations and evaluated deep learning-based approaches on the automatic generation of the segmentations. We adduced the difficulty of manual WA segmentation process and illustrated that the use of deep learning as a part of a semi-automatic segmentation pipeline can speed up the process without compromising segmentation quality. We reached promising results on the automatic segmentation task using both 3D- and 2D-based CNN models.

Abbreviations

2D	Two dimensional
3D	Three dimensional
CNN	Convolutional neural network
CT	Computed tomography
CTA	Computed tomography angiography
DCS	Dice coefficient score
GT	Ground truth
HD	Hausdorff distance
IMH	Intramural hematoma
MSD	Mean surface distance
NSD	Normalized surface dice
TBAD	Stanford type B aortic dissection
WA	Whole aorta

Authors' contributions

RK, conceptualization, data curation, data analysis, methodology, resources, visualization, original draft preparation, and manuscript review and editing. TK, conceptualization, data curation, data analysis, methodology, software, visualization, original draft preparation, and manuscript review and editing. ES, conceptualization, data curation, data analysis, methodology, software, original draft preparation, and manuscript review and editing. PR, conceptualization, resources, and manuscript review and editing. SS, conceptualization, methodology, project administration, software, and manuscript review and editing. MK, conceptualization, resources, methodology, project administration, and manuscript review and editing. All the authors have approved the manuscript.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The Helsinki University Hospital's ethical committee approved this retrospective study, and patients' informed consent was waived.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Radiology, HUS Medical Imaging Center, Helsinki University Hospital and University of Helsinki, FI-00290 Helsinki, Finland. [2]Department of Cardiac Surgery, Heart and Lung Center, Helsinki University Hospital and University of Helsinki, Helsinki, Finland. [3]Department of Physics, University of Helsinki, Helsinki, Finland.

References

1. Erbel R, Aboyans V, Boileau C et al (2014) 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 35:2873–2926. https://doi.org/10.1093/eurheartj/ehu281

2. Hiratzka LF, Bakris GL, Beckman JA et al (2010) 2010 ACCF/AHA/AATS/ACR/ASA/SCA/SCAI/SIR/STS/SVM guidelines for the diagnosis and management of patients with thoracic aortic disease: a report of the American College of Cardiology Foundation/American Heart Association task force on practice guidelines, American Association for Thoracic Surgery, American College of Radiology, American Stroke Association, Society of Cardiovascular Anesthesiologists, Society for Cardiovascular Angiography and Interventions, Society of Interventional Radiology, Society of Thoracic Surgeons, and Society for Vascular Medicine. Circulation 121:e266-369. https://doi.org/10.1161/CIR.0b013e3181d4739e

3. Fleischmann D, Afifi RO, Casanegra AI, et al (2022) Imaging and surveillance of chronic aortic dissection: a scientific statement from the American Heart Association. Circ Cardiovasc Imaging 15:e000075. https://doi.org/10.1161/HCI.0000000000000075

4. Goldstein SA, Evangelista A, Abbara S et al (2015) Multimodality imaging of diseases of the thoracic aorta in adults: from the American Society of Echocardiography and the European Association of Cardiovascular Imaging: endorsed by the Society of Cardiovascular Computed Tomography and Society for Cardiovascular Magnetic Resonance. J Am Soc Echocardiogr 28:119–182. https://doi.org/10.1016/j.echo.2014.11.015

5. Lombardi JV, Hughes GC, Appoo JJ et al (2020) Society for Vascular Surgery (SVS) and Society of Thoracic Surgeons (STS) reporting standards for type B aortic dissections. J Vasc Surg 71:723–747. https://doi.org/10.1016/j.jvs.2019.11.013

6. Singh K, Jacobsen BK, Solberg S et al (2003) Intra- and interobserver variability in the measurements of abdominal aortic and common iliac artery diameter with computed tomography. The Tromsø study. Eur J Vasc Endovasc Surg 25:399–407. https://doi.org/10.1053/ejvs.2002.1856

7. Mora C, Marcus C, Barbe C et al (2014) Measurement of maximum diameter of native abdominal aortic aneurysm by angio-CT: reproducibility is better with the semi-automated method. Eur J Vasc Endovasc Surg 47:139–150. https://doi.org/10.1016/j.ejvs.2013.10.013

8. Quint LE, Liu PS, Booher AM et al (2013) Proximal thoracic aortic diameter measurements at CT: repeatability and reproducibility according to measurement method. Int J Cardiovasc Imaging 29:479–488. https://doi.org/10.1007/s10554-012-0102-9

9. Dugas A, Therasse É, Kauffmann C et al (2012) Reproducibility of abdominal aortic aneurysm diameter measurement and growth evaluation on axial and multiplanar computed tomography reformations. Cardiovasc Intervent Radiol 35:779–787. https://doi.org/10.1007/s00270-011-0259-y

10. Shimada I, Rooney SJ, Farneti PA et al (1999) Reproducibility of thoracic aortic diameter measurement using computed tomographic scans. Eur J Cardiothorac Surg 16:59–62. https://doi.org/10.1016/S1010-7940(99)00125-6

11. Lu T-LC, Rizzo E, Marques-Vidal PM et al (2010) Variability of ascending aorta diameter measurements as assessed with electrocardiography-gated multidetector computerized tomography and computer assisted diagnosis software. Interact Cardiovasc Thorac Surg 10:217–221. https://doi.org/10.1510/icvts.2009.216275

12. Cayne NS, Veith FJ, Lipsitz EC et al (2004) Variability of maximal aortic aneurysm diameter measurements on CT scan: significance and methods to minimize. J Vasc Surg 39:811–815. https://doi.org/10.1016/j.jvs.2003.11.042

13. Willemink MJ, Mastrodicasa D, Madani MH et al (2023) Inter-observer variability of expert-derived morphologic risk predictors in aortic dissection. Eur Radiol 33(2):1102–1111. https://doi.org/10.1007/s00330-022-09056-z

14. Liu C-Y, Chen D, Bluemke DA et al (2015) Evolution of aortic wall thickness and stiffness with atherosclerosis: long-term follow up from the Multi-Ethnic Study of Atherosclerosis (MESA). Hypertension 65:1015–1019. https://doi.org/10.1161/HYPERTENSIONAHA.114.05080

15. Rosero EB, Peshock RM, Khera A et al (2011) Sex, race, and age distributions of mean aortic wall thickness in a multiethnic population-based sample. J Vasc Surg 53:950–957. https://doi.org/10.1016/j.jvs.2010.10.073

16. Subasić M, Loncarić S, Sorantin E (2005) Model-based quantitative AAA image analysis using a priori knowledge. Comput Methods Programs Biomed 80:103–114. https://doi.org/10.1016/j.cmpb.2005.06.009

17. Duquette AA, Jodoin P-M, Bouchot O, Lalande A (2012) 3D segmentation of abdominal aorta from CT-scan and MR images. Comput Med Imaging Graph 36:294–303. https://doi.org/10.1016/j.compmedimag.2011.12.001

18. Freiman M, Esses SJ, Joskowicz L, Sosna J (2010) An iterative model-constrained graph-cut algorithm for abdominal aortic aneurysm thrombus segmentation. In: 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. pp 672–675

19. Lareyre F, Adam C, Carrier M et al (2019) A fully automated pipeline for mining abdominal aortic aneurysm using image segmentation. Sci Rep 9:13750. https://doi.org/10.1038/s41598-019-50251-8

20. López-Linares K, Aranjuelo N, Kabongo L et al (2018) Fully automatic detection and segmentation of abdominal aortic thrombus in post-operative CTA images using deep convolutional neural networks. Med Image Anal 46:202–214. https://doi.org/10.1016/j.media.2018.03.010

21. Martínez-Mera JA, Tahoces PG, Carreira JM et al (2013) A hybrid method based on level set and 3D region growing for segmentation of the thoracic aorta. Comput Aided Surg 18:109–117. https://doi.org/10.3109/10929088.2013.816978

22. Yu Y, Gao Y, Wei J et al (2021) A three-dimensional deep convolutional neural network for automatic segmentation and diameter measurement of type B aortic dissection. Korean J Radiol 22:168–178. https://doi.org/10.3348/kjr.2020.0313

23. Yao Z, Xie W, Zhang J, et al (2021) ImageTBAD: a 3D computed tomography angiography image dataset for automatic segmentation of type-B aortic dissection. Front Physiol 12:732711. https://doi.org/10.3389/fphys.2021.732711

24. Fetnaci N, Łubniewski P, Miguel B, Lohou C (2013) 3D segmentation of the true and false lumens on CT aortic dissection images. In: Three-Dimensional Image Processing (3DIP) and Applications 2013. SPIE, pp 176–190

25. Hahn LD, Mistelbauer G, Higashigaito K, et al (2020) CT-based true- and false-lumen segmentation in type B aortic dissection using machine learning. Radiol Cardiothorac Imaging 2:e190179. https://doi.org/10.1148/ryct.2020190179

26. Kovács T, Cattin P, Alkadhi H et al (2006) Automatic segmentation of the vessel lumen from 3D CTA images of aortic dissection. In: Handels H, Ehrhardt J, Horsch A et al (eds) Bildverarbeitung für die Medizin 2006. Springer, Berlin, Heidelberg, pp 161–165

27. Xiaojie D, Meichen S, Jianming W, et al (2016) Segmentation of the aortic dissection from CT images based on spatial continuity prior model. In: 2016 8th International Conference on Information Technology in Medicine and Education (ITME). pp 275–280

28. Li Z, Feng J, Feng Z et al (2019) Lumen segmentation of aortic dissection with cascaded convolutional network. In: Pop M, Sermesant M, Zhao J et al (eds) Statistical Atlases and Computational Models of the Heart. Springer International Publishing, Cham, Atrial Segmentation and LV Quantification Challenges, pp 122–130

29. Lee N, Tek H, Laine AF (2008) True-false lumen segmentation of aortic dissection using multi-scale wavelet analysis and generative-discriminative model matching. In: Medical Imaging 2008: Computer-Aided Diagnosis. SPIE, pp 878–888

30. Mastrodicasa D, Codari M, Bäumler K et al (2022) Artificial intelligence applications in aortic dissection imaging. Semin Roentgenol 57:357–363. https://doi.org/10.1053/j.ro.2022.07.001

31. Cao L, Shi R, Ge Y, et al (2019) Fully automatic segmentation of type B aortic dissection from CTA images enabled by deep learning. Eur J Radiol 121:108713. https://doi.org/10.1016/j.ejrad.2019.108713

32. Bratt A, Blezek DJ, Ryan WJ et al (2021) Deep learning improves the temporal reproducibility of aortic measurement. J Digit Imaging 34:1183–1189. https://doi.org/10.1007/s10278-021-00465-y

33. Krissian K, Carreira JM, Esclarin J, Maynar M (2014) Semi-automatic segmentation and detection of aorta dissection wall in MDCT angiography. Med Image Anal 18:83–102. https://doi.org/10.1016/j.media.2013.09.004

34. Wobben LD, Codari M, Mistelbauer G, et al (2021) Deep learning-based 3D segmentation of true lumen, false lumen, and false lumen thrombosis in type-B aortic dissection. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp 3912–3915

35. Sieren MM, Widmann C, Weiss N, et al (2022) Automated segmentation and quantification of the healthy and diseased aorta in CT angiographies using a dedicated deep learning approach. Eur Radiol 32:690–701. https://doi.org/10.1007/s00330-021-08130-2

36. Fedorov A, Beichel R, Kalpathy-Cramer J et al (2012) 3D Slicer as an image computing platform for the quantitative imaging network. Magn Reson Imaging 30:1323–1341. https://doi.org/10.1016/j.mri.2012.05.001

37. Paszke, A., Gross, S., Massa, F., et al (2019). PyTorch: an imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems 32 (pp. 8024–8035). Curran Associates, Inc. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

38. The MONAI Consortium (2020) Project MONAI. Zenodo. https://doi.org/10.5281/zenodo.4323059

39. Milletari F, Navab N, Ahmadi S-A (2016) V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). pp 565–571

40. Nikolov S, Blackwell S, Zverovitch A, et al (2021) Clinically applicable segmentation of head and neck anatomy for radiotherapy: deep learning algorithm development and validation study. J Med Internet Res 23:e26151. https://doi.org/10.2196/26151

41. Budd S, Robinson E, Kainz B (2021) A survey on active learning and human-in-the-loop deep learning for medical image analysis. Med Image Anal 71:102062. https://doi.org/10.1016/j.media.2021.102062

Publisher's Note