Deep Learning-based Automated Aortic Area and Distensibility Assessment: the Multi-Ethnic Study of Atherosclerosis (MESA)

Vivek P. Jani1 · Nadja Kachenoura2 · Alban Redheuil2 · Gisela Teixido-Tura3 · Kevin Bouaou2 · Emilie Bollache2 · Elie Mousseaux4 · Alain De Cesare2 · Shelby Kutty1 · Colin O. Wu5 · David A. Bluemke6 · Joao A. C. Lima1 · Bharath Ambale-Venkatesh1

Received: 25 February 2021 / Revised: 21 September 2021 / Accepted: 15 October 2021 / Published online: 1 March 2022 © The Author(s) under exclusive licence to Society for Imaging Informatics in Medicine 2022

Abstract
This study details application of deep learning for automatic segmentation of the ascending and descending aorta from 2D phase-contrast cine magnetic resonance imaging for automatic aortic analysis on the large MESA cohort with assessment on an external cohort of thoracic aortic aneurysm (TAA) patients. This study includes images and corresponding analysis of the ascending and descending aorta at the pulmonary artery bifurcation from the MESA study. Train, validation, and internal test sets consisted of 1123 studies (24,282 images), 374 studies (8067 images), and 375 studies (8069 images), respectively. The external test set of TAAs consisted of 37 studies (3224 images). CNN performance was evaluated utilizing a dice coefficient and concordance correlation coefficients (CCC) of geometric parameters. Dice coefficients were as high as 97.55% (CI: 97.47–97.62%) and 93.56% (CI: 84.63–96.68%) on the internal and external test of TAAs, respectively. CCC for maximum and minimum and ascending aortic area were 0.969 and 0.950, respectively, on the internal test set and 0.997 and 0.995, respectively, for the external test. The absolute differences between manual and deep learning segmentations for ascending and descending aortic distensibility were 0.0194 × 10⁻⁴ ± 9.67 × 10⁻⁴ and 0.002 ± 0.001 mmHg⁻¹, respectively, on the internal test set and 0.44 × 10⁻⁴ ± 20.4 × 10⁻⁴ and 0.002 ± 0.001 mmHg⁻¹, respectively, on the external test set. We successfully developed a U-Net-based aortic segmentation and analysis algorithm in both MESA and in external cases of TAA.

Keywords Deep learning · U-Net · Cardiovascular disease · Coronary artery disease · Aortic distensibility · Aortic aneurysm

Introduction

Increased arterial stiffness is associated with aging and incident cardiovascular disease, namely stroke, ischemic heart disease, and heart failure [1–4]. A stiffened aorta provides less systolic cushioning resulting in increased systolic blood pressure, inducing left ventricular hypertrophy [2, 5]. Population- and community-based studies in patients have demonstrated that decreased aortic distensibility is an independent predictor of mortality and incident cardiovascular events [3, 5, 6]. Furthermore, direct measurements of central aortic stiffness by means of aortic distensibility have been shown to be an early marker of subclinical vascular alterations [5, 7]. Cardiovascular magnetic resonance (CMR) has the unique ability to simultaneously assess aortic stiffness and ventricular function [1]. Recently, convolutional neural networks (CNNs) have demonstrated remarkable performance for classification, segmentation, and prediction tasks.
in radiology and CMR-related tasks [8–13]. However, deep learning applications for aortic segmentation are limited by a lack of models trained on large cohorts of data required for robust, generalizable development and widespread use.

Previously, in the Multi-Ethnic Study of Atherosclerosis (MESA), the ArtFun software has been used for aortic analysis, which has been validated on a pulsatile phantom and in several human studies; however, manual intervention is still required [1]. This study details application of U-Net, a deep learning (DL)-based architecture, for automatic segmentation of the ascending and descending aorta from 2D PC-cine MRI for automatic aortic analysis, including completely automated quantification of aortic cross-sectional areas and aortic distensibility. We subsequently tested the efficacy of our model on a separate cohort of patients with thoracic aortic aneurysms (TAA) to demonstrate application to a clinically relevant pathology where aortic geometry is noticeably altered.

Methods

MESA Population Characteristics

The Multi-Ethnic Study of Atherosclerosis (MESA) is a prospective cohort study that evaluates both risk factors and mechanisms underlying cardiovascular disease progression and development among asymptomatic individuals [14]. A total of 6418 individuals without CVD aged 45 to 84 years, and identified as White, Black, Hispanic, or Chinese were recruited between 2000 and 2002 from 6 US field centers (Wake Forest University (WFU), Winston-Salem, NC; Columbia University (COL), New York, NY; Johns Hopkins University (JHU), Baltimore, MD; University of Minnesota, Twin Cities (UMN), Minneapolis, MN; Northwestern University (NWU), Chicago, IL; University of California, Los Angeles (UCLA), Los Angeles, CA). Of 5005 participants with CMR imaging, 1872 had 2D phase-contrast cine MRI for automatic aortic analysis, including completely automated quantification of aortic cross-sectional areas and aortic distensibility. No patients in this cohort had an aortic dissection. A total of 19 out of the 37 individuals had a bicuspid aortic valve (BAV). Three of these individuals had a dilated ascending aorta but did not meet the aneurysmal threshold. All remaining 18 patients with a tricuspid aortic valve had TAA. No patients in this cohort had an aortic dissection.

External Test Set MRI Imaging

TAA MRI images were acquired with 3 T whole-body scanners and gradient echo phase-contrast cine MRI with electrocardiographic gating for aortic flow and lumen evaluation. Ascending and descending aorta images were obtained in the transverse plane perpendicular to the aortic centerline at the level of main pulmonary artery bifurcation. Imaging parameters were as follows: repetition time, 10 ms; echo time, 1.9 ms; flip angle, 20°; field of view, 340 mm; slice thickness, 8 mm; matrix, 256×256; number of images, 20–60 for 1 cardiac cycle; encoding velocity, 150 cm/s; and bandwidth, 245 Hz/pixel.

External Test Set Characteristics

A total of 37 individuals (3224 images) from a cohort of TAA external to MESA from a single center were included, described in detail elsewhere [15]. Characteristics of the population with suitable aortic MRI are shown in Table 1. Central pressures that were recorded simultaneously to MRI acquisitions using SphygmoCor are shown in Table 1. A total of 19 out of the 37 individuals had a bicuspid aortic valve (BAV). Three of these individuals had a dilated ascending aorta but did not meet the aneurysmal threshold. All remaining 18 patients with a tricuspid aortic valve had TAA. No patients in this cohort had an aortic dissection.

Deep Learning Model and Training

Automated segmentation utilized the U-Net CNN architecture, described elsewhere [16]. Segmentation models, in which the ascending and descending aorta were identified as separate classes, were trained with and without training data augmentation, which included random zoom (zoom range 0.2), rotation (rotation range 45°), image shear (shear range 0.2), and horizontal/vertical shifts (20% in either direction). Images in the training set were also normalized with 0 mean pixel intensity and unit variance. 2D PC-cine MRI magnitude images from the MESA cohort were randomly assigned to non-overlapping training (60%), validation (20%), and internal test (20%) sets, as detailed in Fig. 1. Train, validation, and internal test sets consisted of 1123 studies (24,282...
images), 374 studies (8067 images), and 375 studies (8069 images), respectively. We further included an external test set of TAA patients, which consisted of 37 studies (3224 images). Each image from a given time series was an independent data set, and time series information was not considered by the CNN. Images from the same participant and time series were assigned to the same train, validation, or internal/external test group. Segmentation maps for training were manually generated as previously detailed using the ArtFun software by an expert reviewer with 21 years of experience as a magnetic resonance technologist and analyst [5, 7, 17]. Inter- and intra-reader reproducibility is provided in detail elsewhere [18]. Segmentation maps for the external test set were generated as described previously [15]. Both images and segmentation maps were resized to 256×256 with zero-padding, where necessary.

| Characteristics | MESA All MESA participants (n = 1872) | Train (n = 1123) | Validation (n = 374) | Internal test (n = 375) | TAA External test TAA (n = 37) |
|-----------------|--------------------------------------|-----------------|---------------------|------------------------|-----------------------------|
| Number of Images | 40,418                               | 24,282          | 8067                | 8069                   | 3224                        |
| Age, y          | 64 ± 10                              | 64 ± 11         | 64 ± 10             | 63 ± 10                | 60 ± 15                     |
| Men, %          | 48                                   | 47              | 51                  | 51                     | 84                          |
| Height, cm      | 166 ± 10                             | 165 ± 10        | 166 ± 10            | 166 ± 10               | 175 ± 9                     |
| Weight, kg      | 76 ± 16                              | 76 ± 16         | 76 ± 16             | 78 ± 16                | 79 ± 16                     |
| BMI, kg/m²      | 27.6 ± 5.0                           | 27.5 ± 5.0      | 27.6 ± 4.9          | 27.8 ± 4.8             | 25.9 ± 4.7                  |
| Blood pressures |                                      |                 |                     |                        |                             |
| SBP, mmHg       | 128 ± 22                             | 127 ± 22        | 130 ± 22            | 130 ± 23               | 119 ± 15                    |
| DBP, mmHg       | 72 ± 10                              | 72 ± 10         | 72 ± 10             | 73 ± 11                | 84 ± 10                     |
| PP, mmHg        | 56 ± 18                              | 56 ± 18         | 57 ± 18             | 57 ± 18                | 35 ± 9.2                    |
| Heart rate, bpm | 64 ± 10                              | 64 ± 10         | 63 ± 10             | 64 ± 10                | 70 ± 15                     |
| Aortic geometric parameters | | | | | |
| AA maximum area (cm²) | 9.0 ± 2.1 | 9.0 ± 2.1 | 8.7 ± 1.9 | 9.0 ± 2.1 | 12.7 ± 6.1 |
| AA minimum area (cm²) | 8.3 ± 2.0 | 8.4 ± 2.0 | 8.1 ± 1.8 | 8.4 ± 2.0 | 11.5 ± 5.7 |
| AA distensibility (×10⁻³ mmHg⁻¹) | 1.5 ± 1.3 | 1.5 ± 1.3 | 1.6 ± 1.2 | 1.5 ± 1.1 | 3.4 ± 1.9 |
| DA maximum area (cm²) | 5.3 ± 1.3 | 5.3 ± 1.3 | 5.1 ± 1.3 | 5.4 ± 1.4 | 5.2 ± 2.7 |
| DA minimum area (cm²) | 4.8 ± 1.3 | 4.9 ± 1.3 | 4.7 ± 1.2 | 4.9 ± 1.3 | 4.5 ± 2.4 |
| DA distensibility (×10⁻³ mmHg⁻¹) | 1.8 ± 1.3 | 1.8 ± 1.3 | 1.8 ± 1.3 | 2.0 ± 1.5 | 5.3 ± 4.2 |

Fig. 1 Train, validation, and test group split and study design

© Springer
An un-weighted categorical cross-entropy loss function with Adam optimization was utilized for training [19]. Network hyperparameters were as follows: learning rate, $1 \times 10^{-5}$; batch size, 64; and epochs, 75. Hyperparameters were tuned utilizing empirical methods by maximizing the validation set accuracy, assessed from the Dice coefficient and loss function. Learning rate and epochs parameters were updated in one-log increments, while the batch size parameter was updated by multiples of two. Training time was 2 min 50 s per epoch (approximately 5 h for 75 epochs). Segmentation time was approximately 2 ms per image. The model was built and trained in Python (ver. 3.8.8) utilizing Keras (ver. 2.4.3) with a TensorFlow (ver. 2.4.1) backend [20–22]. Model training and evaluation were performed on a server workstation with 12 CPU cores, 32 GB RAM, and two NVIDIA graphics processing units (GPUs) each with 16 GB of video memory (NVIDIA [Santa Clara, CA, USA] Tesla P100).

Model Evaluation

Automated segmentation maps in the test set were evaluated relative to manual ground truth segmentation maps with the Dice coefficient [12, 23, 24]. Briefly, the Dice coefficient assesses the overlap between two different segmentation maps, considering only non-zero background pixels. For visualization, contours were obtained from segmentation maps utilizing the OpenCV image processing library in Python.

Quantification of Aortic Distensibility

Aortic geometric parameters evaluated in this study include maximum and minimum ascending and descending aortic area and ascending and descending aortic distensibility. Aortic strain was calculated as the difference of maximum and minimum aortic area divided by the minimum aortic area. Distensibility was quantified as strain over pulse pressure. Strain was quantified based on maximum and minimum area and not by phases of the cardiac cycle (e.g., systole and diastole). Manually derived and deep learning-derived segmentations may have resulted in different phases of systole and diastole, though not significantly. Quantitative aortic parameters were obtained from the ArtFun software by providing manual and deep learning-derived segmentation maps of the modulus images as inputs, which were then superimposed to velocity images acquired from PC-cine MRI [5, 17]. Aortic geometric parameters were quantified for two sets of contours for each participant, namely (A) contours defined by the user (manual) and (B) deep learning-derived contours (deep learning).

Outcomes

All-cause death, stroke, and all cardiovascular disease (CVD) were adjudicated as part of MESA and were used to assess whether deep learning-derived contours predicted outcomes. CVD outcomes represented a composite of death, stroke, and coronary heart disease (any of myocardial infarction, resuscitated cardiac arrest, definite angina, probable angina followed by revascularization, and death). Events were assessed by telephone interview every 6 to 9 months to assess all interim hospital admissions, outpatient diagnoses, and deaths. Two physicians reviewed all medical records for endpoint classification and assignment of event dates.

Statistical Analysis

Continuous variables are expressed as mean ± SD, unless otherwise specified. Categorical variables are expressed as percentages. The 95% confidence intervals for Dice coefficients were generated assuming a normal distribution. Agreement between geometric parameters calculated from deep learning and manual segmentation maps was assessed utilizing Bland–Altman analysis and concordance-correlation coefficients (CCC) [25]. Cox proportional hazard models were used to assess the association between aortic distensibility and body surface area (BSA) adjusted maximum aortic area with all CVD outcomes and death. All statistical analyses were performed using Stata, version 15.0 (Stata Corp LP, College Station, TX).

Results

Participant Characteristics

Demographic and relevant clinical parameters for both MESA participants and TAA patients with suitable 2D phase-contrast cine MRI of the aorta are presented in Table 1. Characteristics of the MESA population with suitable aortic MRI were 48% men, 34% White, 14% Chinese American, 30% African American, and 22% Hispanic, with mean age of 64 ± 10 years. Train, validation, and internal (MESA) test group characteristics were observed to be similar to population characteristics. No significant differences were observed in all other relevant clinical parameters between the train, validation, and internal (MESA) test groups.

Model Training and Internal Evaluation

Segmentation models with and without augmentation of the training data set were evaluated on an internal test set, randomly selected from MESA. Dice coefficients for these
models are shown in Supplemental Table 1 and Fig. 2 for both models. These data demonstrate that all models are in good agreement with manually segmented contours (Dice coefficients >95% for all). Of the two models, the segmentation model trained on non-augmented training data was utilized for further evaluation, as this model had the best performance and was better able to discern the ascending aorta (Dice coefficient 96.7%) and descending aorta (Dice coefficient 97.6%) compared with the model trained on augmented training data (ascending aorta Dice coefficient, 94.4%; descending aorta Dice coefficient, 95.9%). Representative contours generated from this model (non-augmented model) are shown in Fig. 3. To demonstrate generalizability, the model was evaluated across different sites in MESA, shown in Supplemental Table 2. These data demonstrated that the site/scanner of image acquisition did not appreciably affect model performance (Dice coefficient >95% for all sites).

To further evaluate model accuracy and investigate the coherence of the segmentation on images from the same time series, quantitative aortic parameters, namely time-resolved ascending and descending aorta area, maximum and minimum ascending aorta area, maximum and minimum descending aorta area, and ascending and descending aortic distensibility, were evaluated as well. Bland Altman plots for ascending and descending aortic areas and distensibility are shown in Figs. 4A–C and 5A–C, respectively. The mean difference and concordance correlation coefficients (CCC) between aortic parameters derived from manual and deep learning segmentations are shown in Table 2. These data demonstrate that the model had excellent performance for quantification of aortic areas (CCC >90%), and fair-to-good performance (CCC >60%) for distensibility assessment.

To evaluate whether deep learning predicted aortic distensibility and BSA adjusted maximum aortic area predicted all CVD outcomes and death in MESA, we performed Cox regression. Results are summarized in Supplemental Table 3. We found that descending aortic distensibility from deep learning predicted all CVD events (HR 0.45, 95% CI: 0.27–0.77) and death (HR 0.69, 95% CI: 0.51–0.92). We further observed that BSA indexed maximum aortic area of both the ascending and descending from deep learning predicted all CVD events and death (hazard ratios in Supplemental Table 3). These findings were similar to what was observed from manually derived contours.

External Model Evaluation

Following evaluation on the internal test set derived from MESA, the models were evaluated on a test set of TAA patients external to MESA. Model performance, in terms of Dice coefficient (median Dice coefficient >90% for all) is shown in Fig. 2 with representative contours shown in Fig. 3. As with the internal data set, quantitative aortic parameters were evaluated for this external data set, with Bland Altman plots shown in Figs. 4D–F and 5D–F and CCC shown in Table 3. Similar to what was observed for the internal data set, the model had excellent performance for quantification of aortic areas (CCC >90%), and fair-to-good performance (CCC >60%) for quantification of distensibility (CCC >60% for ascending aorta). In particular, DL-derived descending aortic distensibility had a lower CCC (47%) compared with that from the internal MESA-derived data set.
This study evaluates the application of U-Net for automatic segmentation of the ascending and descending aorta and automatic quantification of aortic area and distensibility. Compared to other methods, which require manual user intervention, the deep learning-based method presented here is completely automatic and requires no manual intervention. Importantly, all deep learning models evaluated in this study were trained on the MESA data set, which is a large data set with a diverse population and evaluated on an external cohort of thoracic aortic aneurysm patients, which are known to have appreciably different aortic geometry. Segmentation maps from U-Net were found to be in close agreement with manual segmentation.
maps as shown by the high Dice coefficients for both test sets. Furthermore, aortic areas from deep learning segmentations were found to be in excellent agreement with those from manual segmentations. However, the model only had moderate correlation for quantification of aortic distensibility when compared to distensibility calculated from manual segmentations.

### Table 2

|                              | Manual mean (± SD) | DL mean (± SD) | Mean diff (± SD) (DL—manual) | CCC   |
|------------------------------|--------------------|----------------|-----------------------------|-------|
| **Ascending aorta**          |                    |                |                             |       |
| Maximum area (cm²)           | 9.0 ± 2.1          | 8.7 ± 1.9      | −0.3 ± 0.4                  | 0.97  |
| Minimum area (cm²)           | 8.4 ± 2.0          | 8.3 ± 1.9      | −0.1 ± 0.6                  | 0.95  |
| Distensibility (×10⁻³ mmHg⁻¹) | 1.5 ± 1.1          | 1.5 ± 1.9      | 0.2 ± 1.0                   | 0.63  |
| **Descending aorta**         |                    |                |                             |       |
| Maximum area (cm²)           | 5.4 ± 1.4          | 5.2 ± 1.3      | −0.3 ± 0.3                  | 0.96  |
| Minimum area (cm²)           | 5.0 ± 1.3          | 5.0 ± 1.3      | −0.03 ± 0.29                | 0.98  |
| Distensibility (×10⁻³ mmHg⁻¹) | 2.0 ± 1.5          | 1.8 ± 1.3      | −2.0 ± 1.0                  | 0.63  |

**Fig. 4** Bland Altman analysis ascending aortic quantitative parameters. Maximum and minimum ascending aorta area (cm²) along with aortic distensibility (mmHg⁻¹) were determined on both the internal MESA-derived test set and external thoracic aortic aneurysm (TAA) test set. Shown here are **A** maximum ascending aorta area, **B** minimum ascending aorta area, and **C** ascending aortic distensibility for the internal MESA test set and **D** maximum ascending aorta area, **E** minimum ascending aorta area, and **F** ascending aortic distensibility for the external TAA test set. Corresponding concordance correlation coefficients are found in Table 2.
Our model demonstrated strong performance (CCC: 0.93–0.99) for determination of aortic area parameters. Previously, in MESA, intraclass correlation coefficients (ICC) for intra- and inter-observer reproducibility ranged between 0.87–0.99 and 0.56–0.99 for all aortic parameters, respectively [18]. Similar deep learning-based analysis models to assess the aorta reported a mean Dice coefficient of 94.0% [26]. The deep learning method for aortic analysis presented here has comparable performance to manual methods and previously published deep learning models. Importantly, application of the deep learning model on a small cohort of TAA external to the MESA set demonstrated close agreement with manually derived segmentation maps as revealed by the high Dice coefficients and high correlations with low observed bias and variability. Despite the enlarged area of the ascending aorta in TAA, and with no TAA patients in the training data, the model was able to successfully segment the vessel and accurately quantify area in most cases.

Our model demonstrated moderate correlation for DL-estimated aortic distensibility with that determined from manual segmentations for both the ascending and descending aorta. In particular, model performance was lowest for descending aortic distensibility on the external test set of TAAs. Bland Altman analysis reveals no significant trends in variance or significant positive or negative bias, though there was a larger difference observed for larger distensibility, which is a limitation of our model. There could be several reasons for this observation. Distensibility is a ratio, and as such, accuracy suffers if either the maximum or minimum aortic area is not accurately quantified. In our analysis, the maximum and minimum aortic areas were based purely on aortic geometry and not on the corresponding systolic

Fig. 5 Bland Altman analysis descending aortic quantitative parameters. Maximum and minimum descending aorta area (cm²) along with aortic distensibility (mmHg⁻¹) were determined on both the internal MESA-derived test set and external thoracic aortic aneurysm (TAA) test set. Shown here are A maximum descending aorta area, B minimum descending aorta area, and C descending aortic distensibility for the internal MESA test set and D maximum descending aorta area, E minimum descending aorta area, and F descending aortic distensibility for the external TAA test set. Corresponding concordance correlation coefficients are found in Table 2.
and diastolic phases of the cardiac cycle. As a result, maximum and minimum areas are taken from different phases of the cardiac cycle for manual- and deep learning-derived analyses. While not significantly affecting quantification of aortic area, this approach affects distensibility quantification more. Despite this limitation, the results from our model agree with previous efforts to assess aortic distensibility in MESA—ICC for inter-class variability for aortic distensibility for MESA was 0.56 between two independent expert users. Consequently, compared to two independent users, deep learning-derived distensibility was comparable, likely because there was significant variability in pulse pressure between patients and not just strain. We suspect that higher resolution imaging and further model training will allow for distensibility to be calculated with higher accuracy. Despite these limitations, we nonetheless observed that current distensibility measures and BSA adjusted maximum aortic area from deep learning predicted all cardiac events and death.

Data augmentation did not appear to improve model performance on the internal data set, as assessed by median and IQR of the dice coefficient, likely because the MESA dataset used for training was large and diverse enough to give strong results for a relatively simple segmentation task. We hypothesize that the neural network likely used spatial relationships between the ascending and descending aorta for segmentation. We further suspect that in pathologies that alter anatomy, thoracic aortic aneurysm in this case, spatial relationships between the ascending and descending aorta may be less valuable. Importantly, on the external data set, we found that augmentation did improve model performance, as assessed by the IQR of the dice coefficient (Fig. 2). Furthermore, these results are consistent with what is expected for deep learning model training, namely that data augmentation improves model generalizability to external data sets.

Errors in segmentation by our deep learning model were observed in both the internal and external test sets. The training data set in the present study utilized only magnitude images, and phase images were not utilized for training. As such, poorly segmented cases in the internal MESA test data set were primarily cases where manual segmentation was performed utilizing both the magnitude and phase images. Incorporation of phase images would likely improve network performance in these cases. Errors in the external TAA test set, however, were primarily isolated to areas of filling defects in the ascending aorta, well described in TAA, and due to abnormal descending aorta shapes, not present in the training set. Of these, the descending aorta segmentations were the primary source of error. We suspect that transfer learning with additional cases of TAA in the training set will likely reduce segmentation errors.

A major strength of the CNN trained in this study is the utilization of a training data set from multiple sites and multiple vendors. Studies by Bernard et al. have demonstrated that a neural network trained on a heterogeneous training set have increased generalizability [27]. Consequently, the site and vendor generalizability of this CNN are important for effective clinical implementation. The model was able to successfully segment the ascending and descending aorta despite slice selection, anatomic variation, and noise on both an internal and an external data set of TAA. Additionally, time series information was not considered by the network for segmentation. The primary advantage of this approach is expansion of training set size, increasing probability for convergence utilizing traditional deep learning methodologies.

No aortic deep learning segmentation models have been trained on large cohorts of data, analogous to ImageNet, required for both robust model development and generalizability [26, 28]. In one study, Bratt et al. reported uniformly successful automated aortic flow determination utilizing
U-Net segmentation [26]. However, the CNN was trained on 150 patients, who underwent clinical PC-CMR, an appreciably smaller training set compared to the model presented. To our knowledge, this is the first study to have evaluated deep learning segmentation methods for automated aortic analysis on a large, diverse population like MESA. It is our objective that the weights from the deep learning model trained in our study may be used as an initialization for future vascular segmentation tasks.

Limitations

Our study has several limitations. Despite the large training data set, several limitations result from utilizing the MESA cohort, though many are addressed by inclusion of an external test set of TAAs. Participants in the MESA cohort are mostly normal without incident CVD. As the MESA cohort is relatively elderly (mean age, 64 years at baseline), our model may have limitations in patients with congenital heart disease, which present earlier in life. As suggested, weights from the network trained in the present study can be applied to various clinically relevant cases utilizing transfer learning. Segmentation maps utilizing deep learning fail to account for sub-pixel areas; therefore, small geometric differences in segmentation remain. The strength of our study is the large training data set from MESA derived from multiple sites and vendors, allowing for more generalizable weights, as seen through successful segmentation of the aorta from the TAA cohort.

Conclusion

The present study demonstrates uniformly successful DL-based ascending and descending aortic segmentation and excellent performance for quantification of aortic area in both an internal MESA test set and an external test set consisting of a cohort of TAA patients. Aortic distensibility from our model demonstrated moderate correlation with that derived from manual segmentations. It is our objective that the weights from the deep learning model trained in our study may be used as an initialization for future vascular segmentation tasks. Future studies may involve training and evaluating expansions of the proposed model on images of specific clinical pathologies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10278-021-00529-z.

Acknowledgements The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

Author contribution VJ, JL, and BAV conceived and designed the study. VJ, NK, AR, GTT, KB, and ADC performed measurements and manual segmentations. NK and AR provided the ArtFUN software. VJ and BAV performed statistical analysis. VJ, SK, JL, and BAV wrote the manuscript. SK, CW, and DB provided additional assistance drafting the manuscript. All authors revised the manuscript critically for important intellectual content, and all authors read and approved the final version to be published.

Funding This research was supported by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, and N01-HC-95169 from the National Heart, Lung, and Blood Institute, and by grants UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420 from the National Center for Advancing Translational Sciences (NCATS).

Availability of Data and Materials The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

Declarations

Ethics Approval and Consent to Participate All participants gave informed consent for the study protocol, which was approved by the institutional review boards of all MESA field centers and the CMR reading center. MESA field center IRB numbers (WFU—IRB00008492, COL—IRB00002973; JHU—00001656; UM—IRB0000438; NWU—IRB00005003; UCLA—0000172).

Consent for Publication All authors have provided consent for publication.

Competing Interest There is no potential conflict of interest, real or perceived, by the authors. The views expressed in this manuscript are those of the authors and do not necessarily represent the view of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

References

1. Noda C, Ambale Venkatesh B, Ohyama Y, et al.: Reproducibility of functional aortic analysis using magnetic resonance imaging: the MESA. Eur Heart J Cardiovasc Imaging 2016; 17:909–917.
2. Mattace-Raso FUS, Van Der Cammen TJM, Hofman A, et al.: Arterial stiffness and risk of coronary heart disease and stroke: The Rotterdam Study. Circulation 2006; 113:657–663.
3. Liu CY, Chen D, Bluemke DA, et al.: Evolution of aortic wall thickness and stiffness with atherosclerosis: Long-term follow up from the multi-ethnic study of atherosclerosis. Hypertension 2015; 65:1015–1019.
4. Cavalcante JL, Lima JAC, Redheuil A, Al-Mallah MH: Aortic stiffness: Current understanding and future directions. J Am Coll Cardiol 2011; 57:1511–1522.
5. 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.
6. Ohyama Y, Ambale-Venkatesh B, Noda C, et al.: Aortic Arch Pulse Wave Velocity Assessed by Magnetic Resonance Imaging as a Predictor of Incident Cardiovascular Events: The MESA (Multi-Ethnic Study of Atherosclerosis). Hypertension 2017; 70:524–530.
7. Redheuil A, Yu WC, Wu CO, et al.: Reduced ascending aortic strain and distensibility: Earliest manifestations of vascular aging in humans. Hypertension 2010; 55:319–326.
8. Bello GA, Dawes TJW, Duan J, et al.: Deep-learning cardiac motion analysis for human survival prediction. Nat Mach Intell 2019; 1:95.
9. Tao Q, Yan W, Wang Y, et al.: Deep Learning–based Method for Fully Automatic Quantification of Left Ventricle Function from Cine MR Images: A Multivendor, Multicenter Study. Radiology 2018; 290:81–88.
10. Prevedello LM, Erdal BS, Ryu JL, et al.: Automated Critical Test Findings Identification and Online Notification System Using Artificial Intelligence in Imaging. Radiology 2017; 285:923–931.
11. Weston AD, Kortiatis P, Kline TL, et al.: Automated Abdominal Segmentation of CT Scans for Body Composition Analysis Using Deep Learning. Radiology 2018:181432.
12. Liu F, Jang H, Kijowski R, Bradshaw T, McMillan AB: Deep Learning MR Imaging–based Attenuation Correction for PET/ MR Imaging. Radiology 2017; 286:676–684.
13. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S: Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. IEEE Trans Med Imaging 2016; 35:1207–1216.
14. Folsom AR, Diez Roux A V, O’Leary D, et al.: Multi-Ethnic Study of Atherosclerosis: Objectives and Design. Am J Epidemiol 2002; 156:871–881.
15. Pascaran AF, Houriez S, Craiem D, et al.: Comprehensive assessment of local and regional aortic stiffness in patients with tricuspid or bicuspid aortic valve aortopathy using magnetic resonance imaging. Int J Cardiol 2021; 326:206–212.
16. Ronneberger O, Fischer P, Brox T: U-net: Convolutional networks for biomedical image segmentation. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2015; 9351:234–241.
17. Dogui A, Redheuil A, Lefort M, et al.: Measurement of aortic arch pulse wave velocity in cardiovascular MR: comparison of transit time estimators and description of a new approach. J Magn Reson Imaging 2011; 33:1321–1329.
18. Noda C, Ambale Venkatesh B, Ohyama Y, et al.: Reproducibility of functional aortic analysis using magnetic resonance imaging: the MESA. Eur Hear J - Cardiovasc Imaging 2016; 17:909–917.
19. Kingma DP, Ba J: Adam: A method for stochastic optimization. arXiv Prepr arXiv14126980 2014.
20. Abadi M, Barham P, Chen J, et al.: Tensorflow: A system for large-scale machine learning. In 12th {USENIX} Symp Oper Syst Des Implement ( {OSDI} 16); 2016:265–283.
21. Chollet F: keras. 2015.
22. Van Rossum G, Drake Jr FL: Python Tutorial. Volume 620. Centrum voor Wiskunde en Informatica Amsterdam; 1995.
23. Dice LR: Measures of the amount of ecologic association between species. Ecology 1945; 26:297–302.
24. Sørensen T: A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. Biol Skr 1948; 5:1–34.
25. Steichen TJ, Cox NJ: A note on the concordance correlation coefficient. Stata J 2002; 2:183–189.
26. Brett A, Kim J, Pollie M, et al.: Machine learning derived segmentation of phase velocity encoded cardiovascular magnetic resonance for fully automated aortic flow quantification. J Cardiovasc Magn Reson 2019; 21:1.
27. Bernard O, Lalande A, Zotti C, et al.: Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? IEEE Trans Med Imaging 2018; 37:2514–2525.
28. Biasioli L, Hann E, Lukaschuk E, et al.: Automated localization and quality control of the aorta in cine CMR can significantly accelerate processing of the UK Biobank population data. PLoS One 2019; 14:e0212272.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.