Identification and quantification of cardiovascular structures from CCTA an end-to-end, rapid, pixel-wise, deep-learning method

Baskaran, L.; Maliakal, G.; Al'Aref, S.J.; Singh, G.; Xu, Z.R.; Michalak, K.; ... ; Min, J.K.

JACC: CARDIOVASCULAR IMAGING VOL. 13, NO. 5, 2020

PUBLISHED BY ELSEVIER

ORIGINAL RESEARCH

Identification and Quantification of Cardiovascular Structures From CCTA

An End-to-End, Rapid, Pixel-Wise, Deep-Learning Method

Lohendran Baskaran, MBBS,[a,b,c,*] Gabriel Maliakal, MSc,[a,*] Subhi J. Al'Aref, MD,[a,b] Gurpreet Singh, PhD,[a] Zhuoran Xu, MD, MSc,[a] Kelly Michalak, BA,[a] Kristina Dolan, BA,[a] Umberto Gianni,[a] Alexander van Rosendael, MD,[a] Inge van den Hoogen,[a] Donghee Han,[d] Wijnand Stuijfzand,[e] Mohit Pandey, MSc,[a] Benjamin C. Lee, PhD,[a] Fay Lin, MD,[a,b] Gianluca Pontone, MD, PhD,[f] Paul Knaapen, MD, PhD,[e] Hugo Marques, MD, PhD,[g] Jeroen Bax, MD, PhD,[h] Daniel Berman, MD,[d] Hyuk-Jae Chang, MD, PhD,[i] Leslee J. Shaw, PhD,[a,b] James K. Min, MD[a,b]

ABSTRACT

OBJECTIVES This study designed and evaluated an end-to-end deep learning solution for cardiac segmentation and quantification.

BACKGROUND Segmentation of cardiac structures from coronary computed tomography angiography (CCTA) images is laborious. We designed an end-to-end deep-learning solution.

METHODS Scans were obtained from multicenter registries of 166 patients who underwent clinically indicated CCTA. Left ventricular volume (LVV) and right ventricular volume (RVV), left atrial volume (LAV) and right atrial volume (RAV), and left ventricular myocardial mass (LVM) were manually annotated as ground truth. A U-Net–inspired, deep-learning model was trained, validated, and tested in a 70:20:10 split.

RESULTS Mean age was 61.1 ± 8.4 years, and 49% were women. A combined overall median Dice score of 0.9246 (interquartile range: 0.8870 to 0.9475) was achieved. The median Dice scores for LVV, RVV, LAV, RAV, and LVM were 0.938 (interquartile range: 0.887 to 0.958), 0.927 (interquartile range: 0.916 to 0.946), 0.934 (interquartile range: 0.899 to 0.950), 0.915 (interquartile range: 0.890 to 0.920), and 0.920 (interquartile range: 0.811 to 0.944), respectively. Model prediction correlated and agreed well with manual annotation for LVV ($r = 0.98$), RVV ($r = 0.97$), LAV ($r = 0.78$), RAV ($r = 0.97$), and LVM ($r = 0.94$) ($p < 0.05$ for all). Mean difference and limits of agreement for LVV, RVV, LAV, RAV, and LVM were 1.20 ml (95% CI: −7.12 to 9.51), −0.78 ml (95% CI: −10.08 to 8.52), −3.75 ml (95% CI: −21.53 to 14.03), 0.97 ml (95% CI: −6.14 to 8.09), and 6.41 g (95% CI: −8.71 to 21.52), respectively.

CONCLUSIONS A deep-learning model rapidly segmented and quantified cardiac structures. This was done with high accuracy on a pixel level, with good agreement with manual annotation, facilitating its expansion into areas of research and clinical import. (J Am Coll Cardiol Img 2020;13:1163–71) © 2020 by the American College of Cardiology Foundation.

From the [a]Dalio Institute of Cardiovascular Imaging, Weill Cornell Medicine, New York, New York; [b]Department of Radiology, New York-Presbyterian Hospital and Weill Cornell Medicine, New York, New York; [c]Department of Cardiovascular Medicine, National Heart Centre, Singapore; [d]Department of Imaging, Cedars-Sinai Medical Center, Cedars-Sinai Heart Institute, Los Angeles, California; [e]Department of Cardiology, Amsterdam UMC, Location VU University Medical Center, Amsterdam, the Netherlands; [f]Centro Cardiologico Monzino, IRCCS, Milan, Italy; [g]UNICA, Cardiac CT and MRI Unit, Hospital da Luz, Lisbon, Portugal; [h]Department of Cardiology, Heart Lung Center, Leiden University Medical Center, Leiden, the Netherlands; and the [i]Division of Cardiology, Severance Cardiovascular Hospital, Integrative Cardiovascular Imaging Center, Yonsei University College of Medicine, Seoul, South Korea. *Mr. Baskaran and Mr. Maliakal contributed equally to the content of this paper. The study was supported by the Dalio Institute of Cardiovascular Imaging. Dr. Lee has been a consultant for Cleerly. Dr. Bax has received speaker fees from Abbott Vascular and Boehringer Ingelheim. Dr. Min has received funding from the Dalio Foundation, National Institutes of Health, and GE Healthcare; has served on the scientific advisory board of Arineta and GE Healthcare; and has an equity interest in Cleerly. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose. Joseph Schoepf, MD, served as Guest Editor for this paper.

ISSN 1936-878X/$36.00 https://doi.org/10.1016/j.jcmg.2019.08.025

ABBREVIATIONS AND ACRONYMS

CCTA = coronary computed tomography angiography

CMR = cardiac magnetic resonance

CVD = cardiovascular disease

LAV = left atrial volume

LVM = left ventricular myocardial mass

LVV = left ventricular volume

RAV = right atrial volume

RVV = right ventricular volume

The imaging of cardiac structures is essential in the diagnosis and prognosis of cardiovascular disease (CVD) and is a useful tool in monitoring progression and therapeutic response. Coronary computed tomography angiography (CCTA) provides rapid, noninvasive, isotropic whole-heart imaging with high-spatial resolution. Structural and functional parameters obtained from CCTA imaging correlate well with those from echocardiographic and cardiac magnetic resonance (CMR) (1,2). Although commercial software packages may provide semi-automated delineation and quantification of cardiac structures, substantial manual input is still required (3). This is time-consuming and may be operator-dependent.

Machine learning develops novel algorithmic strategies for the construction of inferential predictive models using large datasets, and its use in cardiac imaging is expanding (4). Within machine learning, deep learning is a subdomain that uses more sophisticated frameworks to perform automated feature (parameter) extraction. Using deep networks with many intermediate layers of artificial "neurons," deep learning can model complex relationships between inputs and outputs. In this way, deep learning can substantially outperform systems that are dependent on experts or that are hand-crafted (5). In this study, we developed and evaluated an automated, end-to-end, deep-learning model for the identification and quantification of cardiovascular structures from CCTAs.

METHODS

STUDY POPULATION. The study population consisted of a randomly selected convenience sample aggregation of 2 international, multicenter, prospective observational registries that were described in detail elsewhere (6,7). Briefly, patients who underwent clinically indicated CCTA were included, with prospectively collected history, risk factors, and symptoms at baseline. Inclusion criteria were patients undergoing clinically indicated CCTA and age 18 years or older. Exclusion criteria were known coronary artery disease, hemodynamic instability, inability to provide consent, pregnancy, known adult congenital heart disease, baseline irregular heart rhythm, heart rate ≥100 beats/min, systolic blood pressure ≤90 mm Hg, contraindications to beta-blockers or nitroglycerin or adenosine, and uninterpretable CCTA. Each site obtained local institutional review board or ethics board approval.

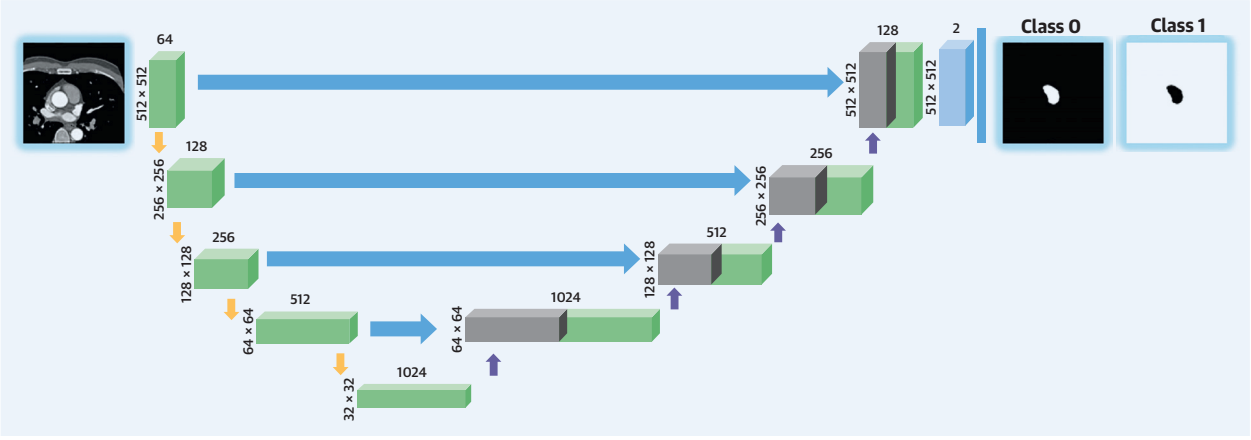
IMAGE ACQUISITION AND SEGMENTATION. Scans were performed using ≥64 detector row scanners. Image acquisition, image post-processing, and data interpretation were performed according to current guidelines (8,9). Images were obtained and reconstructed at 0.50-mm intervals. Digital Imaging and Communications in Medicine files for site CCTAs were transmitted to a core laboratory, where annotation of the cardiac structures was done by level III experienced technologists who were blinded to all other data. Five cardiac structures were annotated in total using Adobe Photoshop (Adobe Systems, San Jose, California): left and right ventricular volume (LVV, RVV); left and right atrial volume (LAV, RAV); and left ventricular myocardial mass (LVM). LVV and RVV were delineated by the left and right ventricular endocardial borders, respectively. These excluded papillary muscles and trabeculations, and followed the contours below the atrioventricular valve planes on a 3-dimensional isotropic voxel level. LAV and RAV measurements included the appendage, but excluded adjacent veins, and were delineated by identifying the endocardial borders. LVM was calculated as the left ventricular myocardial volume derived by the delineation of its endocardial and epicardial borders and multiplied with the specific gravity of myocardial tissue (assuming a tissue density of 1.05 g/ml) (3). These annotations, which were established and verified by board-certified cardiologists, were used as the "ground truth" for the deep-learning model.

SPLITTING OF DATASET AND PREPROCESSING. The entire dataset containing 166 patients was split into 3 parts so that no 2 parts contained images from the same patient: training (70%, n = 132, 9,156 images); validation (20%, n = 34, 2,591 images); and testing (10%, n = 17, 1,477 images). The process of extraction of the ground truth from the annotated images was accomplished by an open source Python library known as "psd_tools" (10). The annotated slices were extracted from the annotated Photoshop files and were arranged according to the color coding

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the *JACC: Cardiovascular Imaging* author instructions page.

Manuscript received May 30, 2019; revised manuscript received August 8, 2019, accepted August 23, 2019.

CENTRAL ILLUSTRATION **Deep Learning Architecture**



Baskaran, L. et al. J Am Coll Cardiol Img. 2020;13(5):1163–71.

The deep-learning network architecture. Each **green block** consists of 2 consecutive sets of 3 × 3 Convolution layer, Rectified Linear Unit activation, and Batch Normalization. The number of feature maps in each layer is indicated at the **top of each box**, whereas the size of each feature map per layer is on the **left**. **Colored arrows** indicate separate operations as shown.

assigned for each label. Because the Digital Imaging and Communications in Medicine volumes were converted to have an isotropic voxel spacing of 0.625 × 0.625 × 0.625 mm, the extracted labels would also have the same isotropic resolution. As a preprocessing step, the images were windowed with a Hounsfield unit (HU) window (−300 to 500) so that all structures to be segmented were sufficiently visible. Each input image contained a pair of complementary label images. All the images and labels were then resized to 512 × 512 and passed to the model.

DEEP-LEARNING MODEL. The deep-learning architecture used was U-Net (11). This network was used for biomedical image segmentation in the past and demonstrated good performance on segmentation of organs in chest images (12). It consists of 4 layers in which the image is down sampled by a Conv3 × 3 layer that consists of 2 runs through a set of component operations (convolution with 3 × 3 kernel, rectified linear unit, and a batch normalization layer) (**Central Illustration**). The resultant feature maps are down sampled by one-half the resolution by a 2 × 2 Max-pool layer. After 4 layers of this, the feature maps are up sampled by transposed convolution using a kernel size of 2 and a stride of 2, followed by successive Conv3 × 3 blocks. The feature maps from the contracting path are concatenated with the up sampled feature maps of the expanding path. At the last layer, the feature maps are reduced from 128 to 2 using a Conv1 × 1 block that consist of a 1 × 1 convolutional kernel; pixel-wise probabilities for belonging to each class is obtained once this is passed to a Softmax layer. Five separate but similar networks were trained for each structure (i.e., 1 network per structure).

TRAINING STRATEGY AND MODEL EVALUATION. The images were randomly shuffled and passed to the network with a batch size of 4 at a resolution of 512 × 512. The network was made to output binary masks for 2 classes (i.e., the foreground and the background). Dice loss was used to train the network. The Dice loss was calculated by subtracting the mean Dice similarity score from 1 (13). An Adam optimizer was used with a learning rate of 0.001 to carry out training (14). The outputs were compared with the ground truth that contained complimentary images of the contour of interest using the Dice loss. The network with the lowest Dice loss on the validation set was selected and evaluated on the test set. Each model was trained for 50 epochs. The model with the best validation loss was chosen among these epochs. The data were shuffled every epoch.

The image-based performance metric was based on Dice loss, calculated by subtracting the mean Dice similarity score from 1 (13). The Dice similarity score quantifies the pixel-wise degree of similarity between the model-predicted segmentation mask and the ground truth, and ranges from 0 (no similarity) to 1 (identical). Mathematically, it can be expressed as follows:

TABLE 1 Baseline Characteristics (N = 166)

Age, yrs	61.14 ± 8.38
Female	48.795
Diabetes mellitus	32.53
Dyslipidemia	45.18
Hypertension	53.61
Smoker	15.06

Values are mean ± SD or %.

$$\text{Dice similarity coefficient} = \frac{(2 \cdot \text{True Positive})}{(2 \cdot \text{True Positive} + \text{False Positive} + \text{False Negative})}$$

STATISTICAL ANALYSIS. Statistical analysis was performed using R version 3.5.0 (R Core Team, R Foundation, Vienna, Australia). Continuous and normally distributed variables were expressed by their mean ± SD and categorical data by their numbers (percentages). Levels of agreement between the model prediction and ground truth (manual annotation) were assessed on the test set with the Bland-Altman difference against mean plot. Pearson correlation's test of volumes and mass between model prediction and ground truth were similarly assessed. Dice scores were summarized as medians and quartiles. Subgroup analysis of Dice scores by sex and age were compared by Wilcoxon's test. A p value of <0.05 was considered significant.

RESULTS

The dataset consisted of 166 patients, with 13,224 images and 3.466 billion pixels. The cohort mean age was 61.1 ± 8.4 years, and 49% were women (Table 1). There were no significant differences for patient characteristics among the training, validation, and test sets (Supplemental Table 1). Mean LVV was 47.46 ± 20.27 ml, RVV was 65.44 ± 18.78 ml, LAV was 43.35 ± 13.25 ml, and RAV was 44.74 ± 14.37 ml (Table 2). Mean LVM was 69.95 ± 21.35 g. There were no significant differences between sexes or between ages younger than 65 years and 65 years or older for all parameters.

For all 5 structures (LVV, RVV, LAV, RAV, LVM), a combined overall median Dice score of 0.9246 (interquartile range: 0.8870 to 0.9475) was achieved on the validation set. Comparisons among the original image, manual annotation, and model prediction are shown in Figure 1. The median Dice scores for LVV, RVV, LAV, RAV, and LVM were 0.938 (interquartile range: 0.887 to 0.958), 0.927 (interquartile range: 0.916 to 0.946), 0.934 (interquartile range: 0.899 to 0.950), 0.915 (interquartile range: 0.890 to 0.920), and 0.920 (interquartile range: 0.811 to 0.944), respectively (Table 3). For LVV, although there was no significant difference in model prediction performance between the scores for male (0.923) and female (0.953) patients (p = 0.200), there was a difference between scores for ages younger than 65 years (0.958) and 65 years or older (0.834) (p = 0.002). For RVV, there were no significant differences between sexes (men = 0.916; women = 0.927) or age groups (age younger than 65 years = 0.927, age 65 years or older = 0.926). For the LAV, there were no significant differences in Dice scores between sex- or age-based subgroups (men = 0.918, women = 0.943; age younger than 65 years = 0.945, age 65 years or older = 0.918), nor were there differences for the RAV (men = 0.892, women = 0.917; age younger than 65 years = 0.915, age 65 years or older = 0.900) or the LVM (men = 0.855, women = 0.938; age younger than 65 years = 0.935, age 65 years or older = 0.850) (p > 0.05 for all). Automated segmentation for all 5 structures

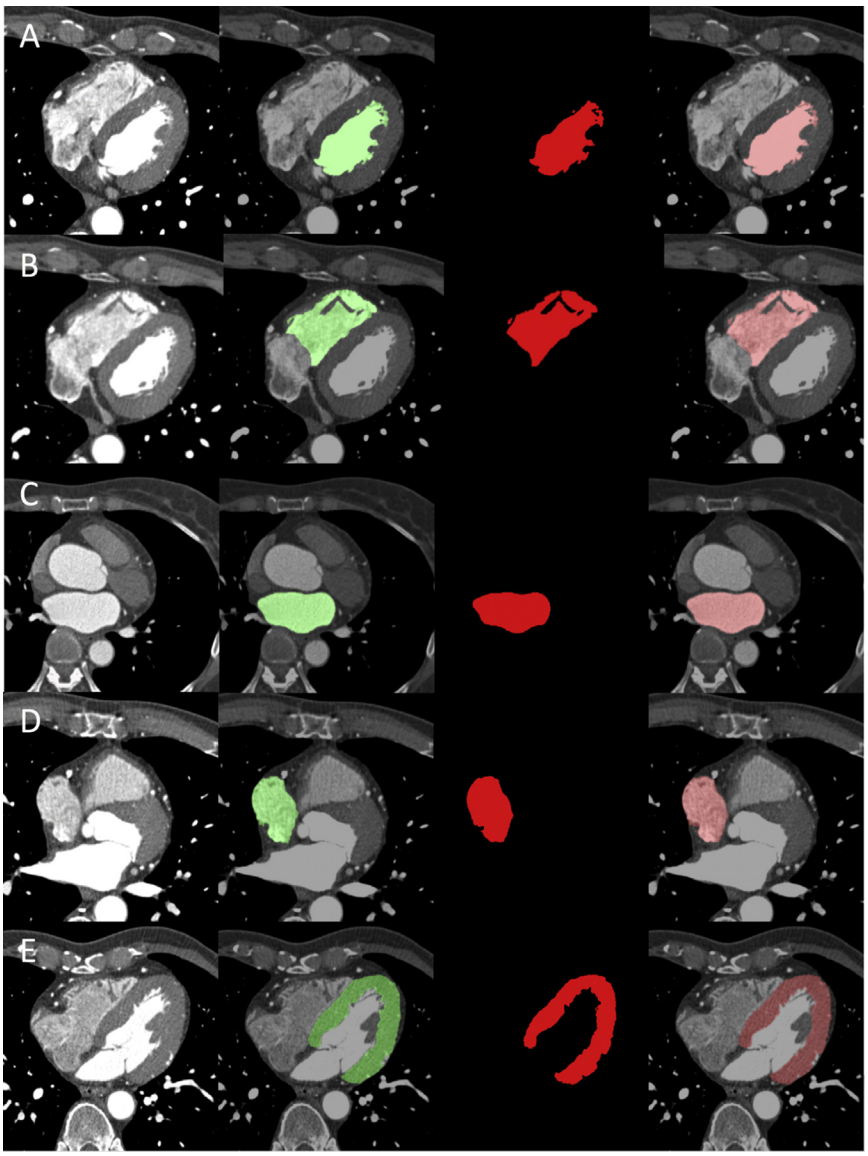
TABLE 2 Ground Truth Measurements

Structure/Category		p Value
LVV (ml)		
Overall	47.46 ± 20.27	—
Male	46.65 ± 15.53	0.815
Female	48.36 ± 24.51	
Age <65 yrs	39.01 ± 13.38	0.200
Age ≥65 yrs	54.96 ± 22.30	
RVV (ml)		
Overall	65.44 ± 18.78	—
Male	63.91 ± 14.94	0.749
Female	67.16 ± 22.20	
Age <65 yrs	59.78 ± 12.38	0.481
Age ≥65 yrs	70.46 ± 21.82	
LAV (ml)		
Overall	43.35 ± 13.25	—
Male	40.60 ± 12.92	0.423
Female	46.44 ± 12.93	
Age <65 yrs	42.59 ± 11.87	0.963
Age ≥65 yrs	44.03 ± 14.33	
RAV (ml)		
Overall	44.74 ± 14.37	—
Male	41.78 ± 11.47	0.541
Female	48.08 ± 16.43	
Age <65 yrs	41.37 ± 10.41	0.277
Age ≥65 yrs	47.74 ± 16.57	
LVM (g)		
Overall	69.95 ± 21.35	—
Male	67.76 ± 15.77	0.963
Female	72.40 ± 26.03	
Age <65 yrs	62.83 ± 18.34	0.277
Age ≥65 yrs	76.27 ± 21.84	

Values are mean ± SD.

LAV = left atrial volume; LVM = left ventricular myocardial mass; LVV = left ventricular volume, RAV = right atrial volume; RVV = right ventricular volume.

FIGURE 1 **Model Prediction**



Comparisons **(left to right)** between the original image, ground truth manual annotation **(green)** deep learning model mask **(red)** and prediction overlaid on the image **(red)** are shown for the: **(A)** left ventricular volume (LVV), **(B)** right ventricular volume (RVV), **(C)** left atrial volume (LAV), **(D)** right atrial volume (RAV), and **(E)** left ventricular myocardial mass (LVM).

took 13.13 s/patient, at 0.124 s/slice, whereas manual segmentation took approximately 1 h/patient. Intra-reader agreement was high with correlation coefficients for the LVV, RVV, LAV, RAV, and LVM of 0.999, 0.999, 0.998, 0.985, and 0.993, respectively. Correlation coefficients for inter-reader agreement for LVV, RVV, LAV, RAV, and LVM were 0.999, 0.995, 0.980, 0.978, and 0.993, respectively.

Overall, the model prediction for all structures correlated well with no significant differences compared with manual annotation ground truth. LVV as predicted by the model correlated well ($r = 0.98$; $p < 0.05$), with a difference in measurement of 1.19 ± 4.12 ml or $5.0 \pm 13.0\%$ ($p = 0.35$), as did RVV prediction ($r = 0.97$; $p < 0.05$), with a volume difference of 0.78 ± 4.60 ml or $1.3 \pm 9.0\%$ ($p = 0.85$). LAV correlated marginally less well ($r = 0.78$; $p < 0.05$), but there was still no significant difference between the model prediction and manual annotation (3.75 ± 8.80 ml or $8.0 \pm 17.0\%$; $p = 0.28$). RAV correlated

TABLE 3 Model Performance

Structure/Category		p Value
LVV		
Overall	0.938 (0.887-0.958)	—
Male	0.923 (0.742-0.942)	0.200
Female	0.953 (0.922-0.960)	
Age <65 yrs	0.958 (0.942-0.965)	0.002
Age ≥65 yrs	0.834 (0.712-0.926)	
RVV		
Overall	0.927 (0.916-0.946)	—
Male	0.916 (0.741-0.934)	0.236
Female	0.927 (0.925-0.956)	
Age <65 yrs	0.927 (0.916-0.946)	0.541
Age ≥65 yrs	0.926 (0.683-0.940)	
LAV		
Overall	0.934 (0.899-0.950)	—
Male	0.918 (0.493-0.945)	0.167
Female	0.943 (0.93-0.952)	
Age <65 yrs	0.945 (0.928-0.950)	0.167
Age ≥65 yrs	0.918 (0.467-0.940)	
RAV		
Overall	0.915 (0.890-0.920)	—
Male	0.892 (0.687-0.920)	0.114
Female	0.917 (0.914-0.943)	
Age <65 yrs	0.915 (0.914-0.920)	0.423
Age ≥65 yrs	0.900 (0.661-0.938)	
LVM		
Overall	0.920 (0.811-0.944)	—
Male	0.855 (0.580-0.935)	0.093
Female	0.938 (0.915-0.947)	
Age <65 yrs	0.935 (0.90-0.945)	0.139
Age ≥65 yrs	0.850 (0.558-0.935)	

Values are median (25th to 75th quartiles).
Abbreviations as in Table 2.

better ([$r = 0.97$; $p < 0.05$]; difference: 0.97 ± 3.52 ml or $2.0 \pm 8.0\%$; $p = 0.31$]). LVM correlated well ($r = 0.94$; $p < 0.05$) with a difference of 6.72 ± 7.85 g or $10.0 \pm 11.0\%$ ($p < 0.05$). Bland-Altman and linear regression plots are shown (**Figure 2**), with the mean difference and limits of agreement for LVV, RVV, LAV, RAV, and LVM being 1.20 ml (95% CI: −7.12 to 9.51), −0.78 ml (95% CI: −10.08 to 8.52), −3.75 ml (95% CI: −21.53 to 14.03), 0.97 ml (95% CI: −6.14 to 8.09), and 6.41 g (95% CI: −8.71 to 21.52), respectively.

DISCUSSION

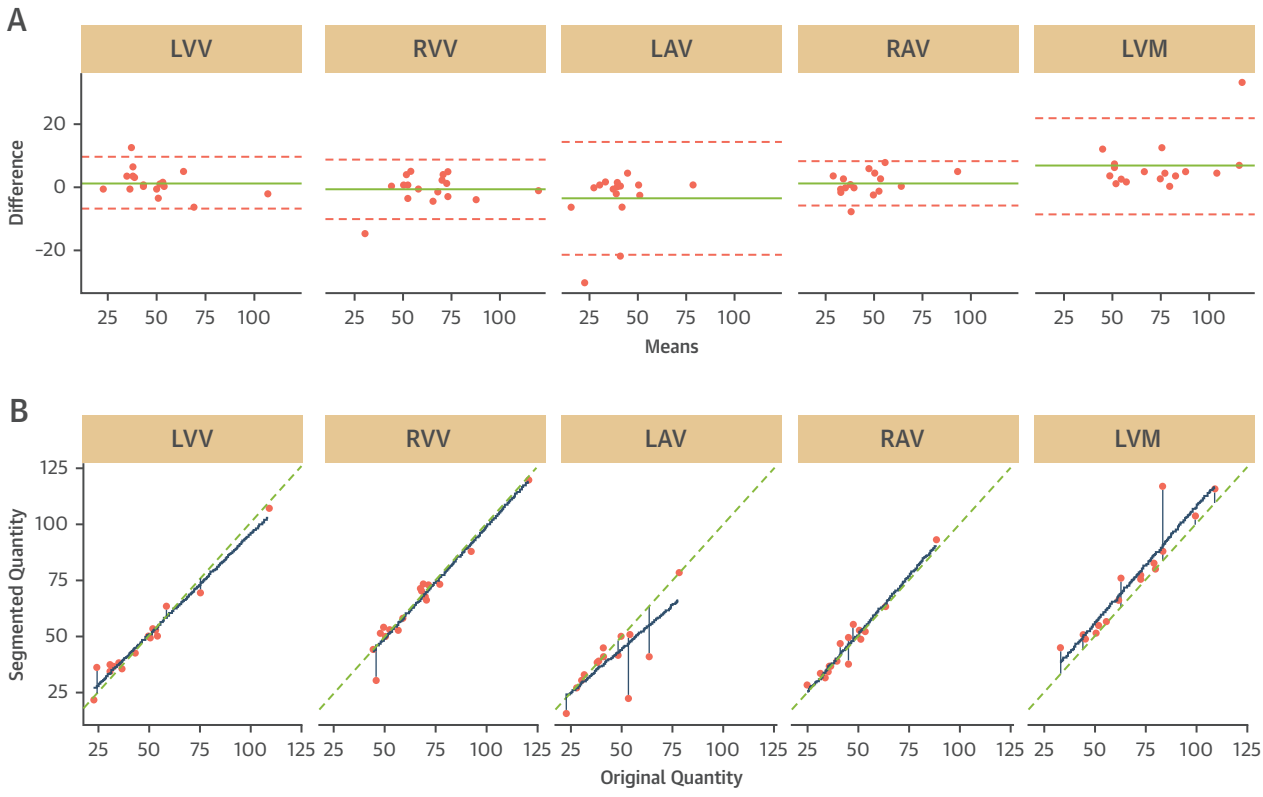
In this study, a deep-learning model provided rapid, end-to-end, automated, pixel-wise identification of the 4 cardiac chambers and LV wall in CCTAs, with quantification of practically relevant parameters. The model exhibited high accuracy and agreed well with manual annotation across sex- and age-stratified subgroups in this multicenter, international cohort.

On a pixel level, this model was able to identify cardiac structures with high accuracy, as reflected by high Dice scores. Although a previous study using CCTA had a Dice score of 0.92, it used a thresholding, rather than a deep-learning method, to segment the LAA (15). A deep-learning method to segment the same structure obtained a superior Dice score of 0.95, as did one that segmented the LVM, with a score of 0.91 (16,17). However, in contradistinction to these studies, the present study identified multiple structures. Although another study used deep learning for multiple structures with comparable accuracies to the present model, accuracy was evaluated on a lower resolution patch-level basis, with a single patch equivalent to 961 pixels (18). This might result in a large difference when measuring volumes of structures. The present model's pixel-based accuracy translated with high correlation and narrow limits of agreement for practically meaningful measurements of cardiac structures. Although different in approach, a 3-dimensional, mesh-based segmentation model of the same structures performed more rapid segmentation in 4 s (19). A similar study segmented 2 additional structures in 20 patients and obtained a Dice score of 0.90 (20).

To have practical import, this model was evaluated with regard to volume and mass quantification of cardiac structures, with high correlation coefficients and low measurement differences. These margins err comparable or markedly narrower than those found in other deep-learning studies. Compared with the present study, a previous study using deep learning to quantify LVV and LVM in CMR showed larger differences, with wider limits of agreement (21). Another deep-learning CMR study of 20 patients showed comparable or lower agreement of LVV and LAV (22). A larger deep-learning CMR study tested on 196 patients found higher correlation than the present study ($r = 0.99$), but with significant underestimation of LVV and LVM, and with wider limits of agreement in the study by Tao et al. (23). In echocardiography, a larger study on LVV obtained a comparable correlation of 0.95 and absolute mean error of 9.5 ml (24). A nuclear myocardial perfusion feasibility study on 56 patients also obtained a comparable correlation of 0.91 (Dice score: 0.93) when segmenting the LVM (25).

In previous non-deep–learning studies, correlation coefficients were lower, not only between modalities, but between differing measurement methods within the same modality (26). Furthermore, the agreement between the present prediction model and the ground truth was comparable to inter- and intra-reader agreements in multiple other manual studies (3,27). A deep-learning model reproduced the same result

FIGURE 2 **Model Agreement and Correlation With Manual Annotation**

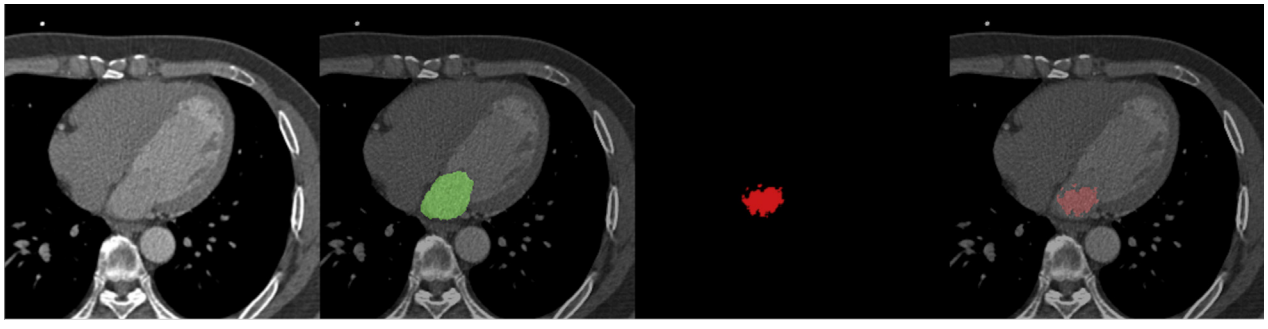


(**A**) Bland-Altman and **(B)** linear regression plots for the differences and agreement between the deep learning–predicted segmented quantities and the original ground truth manually annotated quantities. Abbreviations as in Figure 1.

every time, giving a hypothetical intra-reader agreement of 1.0.

This could benefit clinical workflow because it raises the future possibility of a deep-learning "second reader." Double reading reduces variability and interpretative error, and results in clinically impactful decision changes, but is still sparsely implemented due to time and cost issues. The ability to segment structures may also allow computer-aided detection to evaluate abnormal cardiac structure morphology.

FIGURE 3 **Incorrect Model Prediction**



An example of incorrect LAV prediction by the model **(left to right)**: the original image, ground truth manual annotation **(green)**, deep-learning model mask **(red)**, and prediction overlaid on the image **(red)**. Abbreviation as in Figure 1.

In addition, the rapid 13s throughput could allow future integration into clinical workflows with minimal disruption, helping to reduce workload burden for the imaging clinician and minimizing fatigue (28). As such, this model tentatively portends clinical integration that increases interpretive speed at potentially low cost, while reducing sources of interpretive error.

STUDY LIMITATIONS. A limitation of this study was the cohort size. Although the study design included a hold-out 10% test set that was never seen until the final evaluation, this test set only included 17 patients. This hold-out set consisted of 1,477 images, which was more than amply acceptable for medical image-based, deep-learning applications (29). The high Dice score attested to the model's robustness between these 2 groups, but overfitting could not be ruled out. Nevertheless, the model did make incorrect predictions (**Figure 3**). Furthermore, LVV wall prediction was significantly better in the younger than 65 years age group. Because the training, validation, and test sets were not significantly different, the difference in Dice score might perhaps be attributable to a higher prevalence of comorbidities that might have affected image quality (e.g., atrial fibrillation, respiratory disorders). This was outside the scope of the present study. A larger training and testing cohort and external validation will address these issues and improve real-world model performance. Integration of additional structures into the model (e.g., valves) will increase the applicability of the model. In addition, the scan timing of these images did not optimize contrast opacification of the RVV and RAV. Last, for image extraction and ease of integration into model architecture, annotation was done using Photoshop, instead of clinically conventional image annotation software, a gap that needs to be bridged for clinical integration.

CONCLUSIONS

An end-to-end, deep-learning model was able to detect and segment cardiac structures from CCTA images with high accuracy when evaluated on a pixel level, as well as compared with manual measurement values. This was done in a rapid manner, potentiating expansion into areas of research and clinical import.

ADDRESS FOR CORRESPONDENCE: Dr. Lohendran Baskaran, Weill Cornell Medical College and the Dalio Institute of Cardiovascular Imaging, 413 East 69th Street, Suite 108, New York, New York 10021. E-mail: lob2008@med.cornell.edu.

PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: We developed and evaluated a rapid deep-learning method to segment and measure the 4 cardiac chambers and the left ventricular myocardium with good correlation and agreement to manual methods.

TRANSLATIONAL OUTLOOK: Manual measurement of cardiac structures is laborious. Deep learning allows rapid and robust performance of this task, allowing quicker and larger datasets and measurements to be gathered for research and clinical use.

REFERENCES

1. Lee SC, Ko SM, Song MG, Shin JK, Chee HK, Hwang HK. Morphological assessment of the aortic valve using coronary computed tomography angiography, cardiovascular magnetic resonance, and transthoracic echocardiography: comparison with intraoperative findings. Int J Cardiovasc Imaging 2012;28 Suppl 1: 33–44.

2. Greupner J, Zimmermann E, Grohmann A, et al. Head-to-head comparison of left ventricular function assessment with 64-row computed tomography, biplane left cineventriculography, and both 2- and 3-dimensional transthoracic echocardiography: comparison with magnetic resonance imaging as the reference standard. J Am Coll Cardiol 2012;59:1897–907.

3. Fuchs A, Mejdahl MR, Kühl JT, et al. Normal values of left ventricular mass and cardiac chamber volumes assessed by 320-detector computed tomography angiography in the Copenhagen General Population Study. Eur Heart J Cardiovasc Imaging 2016;17:1009–17.

4. Al'Aref SJ, Anchouche K, Singh G, et al. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. Eur Heart J 2019;40:1975–86.

5. Hinton G. Deep learning—a technology with the potential to transform health care. JAMA 2018; 320:1101–2.

6. Rizvi A, Hartaigh BÓ, Knaapen P, et al. Rationale and design of the CREDENCE trial: computed TomogRaphic evaluation of atherosclerotic DEtermiNants of myocardial IsChEmia. BMC Cardiovasc Disord 2016;16:190.

7. Lee S-E, Chang H-J, Rizvi A, et al. Rationale and design of the Progression of AtheRosclerotic PlAque DetermIned by Computed TomoGraphic Angiography IMaging (PARADIGM) registry: a comprehensive exploration of plaque progression and its impact on clinical outcomes from a multicenter serial coronary computed tomographic angiography study. Am Heart J 2016;182: 72–9.

8. Leipsic J, Abbara S, Achenbach S, et al. SCCT guidelines for the interpretation and reporting of coronary CT angiography: a report of the Society of Cardiovascular Computed Tomography Guidelines Committee. J Cardiovasc Comput Tomogr 2014;8:342–58.

9. Abbara S, Blanke P, Maroules CD, et al. SCCT guidelines for the performance and acquisition of coronary computed tomographic angiography: a report of the Society of Cardiovascular Computed Tomography Guidelines Committee: Endorsed by the North American Society for Cardiovascular Imaging (NASCI). J Cardiovasc Comput Tomogr 2016;10:435–49.

10. Python package for reading Adobe Photoshop PSD files. Available at: psd-tools/psd-tools. Accessed March 10, 2019.

11. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. New York: Springer International Publishing, 2015; p 234–41.

12. Javaid U, Dasnoy D, Lee JA. Multi-organ segmentation of chest CT images in radiation oncology: comparison of standard and dilated UNet. In: Blanc-Talon J, Helbert D, Philips W, Popescu D, Scheunders P, editors. Advanced Concepts for Intelligent Vision Systems. New York: Springer International Publishing, 2018; p 188–99.

13. Sudre CH, Li W, Vercauteren T, Ourselin S, Cardoso MJ. Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. Available at: https://arxiv.org/abs/1707.03237. Accessed April 2, 2019.

14. Kingma DP, Ba J. Adam: a method for stochastic optimization. Available at: https://arxiv.org/abs/1412.6980. Accessed April 5, 2019.

15. Leventić H, Babin D, Velicki L, et al. Left atrial appendage segmentation from 3D CCTA images for occluder placement procedure. Comp Biol Med 2019;104:163–74.

16. Jin C, Feng J, Wang L, et al. Left atrial appendage segmentation using fully convolutional neural networks and modified three-dimensional conditional random fields. IEEE J Biomed Health Inform 2018;22:1906–16.

17. Zreik M, Lessmann N, van Hamersvelt RW, et al. Deep learning analysis of the myocardium in coronary CT angiography for identification of patients with functionally significant coronary artery stenosis. Med Image Anal 2018;44:72–85.

18. Dormer JD, Ma L, Halicek M, Reilly CM, Schreibmann E, Fei B. Heart chamber segmentation from CT using convolutional neural networks. Proc SPIE Int Soc Opt Eng 2018:10578.

19. Zheng Y, Barbu A, Georgescu B, Scheuering M, Comaniciu D. Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. IEEE Trans Med Imaging 2008;27:1668–81.

20. Mortazi A, Burt J, Bagci U. Multi-planar deep segmentation networks for cardiac substructures from MRI and CT. Available at: https://arxiv.org/abs/1708.00983. Accessed March 10, 2019.

21. Curiale AH, Colavecchia FD, Mato G. Automatic quantification of the LV function and mass: a deep learning approach for cardiovascular MRI. Comp Methods Progr Biomed 2019;169:37–50.

22. Narang A, Mor-Avi V, Prado A, et al. Machine learning based automated dynamic quantification of left heart chamber volumes. Eur Heart J Cardiovasc Imaging 2019;20:541–9.

23. 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;290:81–8.

24. Leclerc S, Smistad E, Pedrosa J, et al. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. IEEE Trans Med Imaging 2019;38:2198–210.

25. Wang T, Lei Y, Tang H, et al. A learning-based automatic segmentation and quantification method on left ventricle in gated myocardial perfusion SPECT imaging: a feasibility study. J Nucl Cardiol 2019 Jan 28 [E-pub ahead of print].

26. Rheinheimer S, Reh C, Figiel J, Mahnken AH. Assessment of right atrium volume by conventional CT or MR techniques: which modality resembles in vivo reality? Eur J Radiol 2016;85:1040–4.

27. Maffei E, Messalli G, Martini C, et al. Left and right ventricle assessment with cardiac CT: validation study vs. Cardiac MR. Eur Radiol 2012;22:1041–9.

28. Waite S, Scott J, Gale B, Fuchs T, Kolla S, Reede D. Interpretive error in radiology. Am J Roentgenol 2016;208:739–49.

29. Cho J, Lee K, Shin E, Choy G, Do S. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? Available at: https://arxiv.org/abs/1511.06348. Accessed April 6, 2019.

KEY WORDS coronary computed tomography angiography, deep learning, quantification

APPENDIX For a supplemental table, please see the online version of this paper.