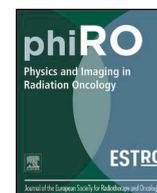




ELSEVIER

Contents lists available at ScienceDirect

# Physics and Imaging in Radiation Oncology

journal homepage: [www.elsevier.com/locate/phro](http://www.elsevier.com/locate/phro)

Original Research Article

## Cardio-pulmonary substructure segmentation of radiotherapy computed tomography images using convolutional neural networks for clinical outcomes analysis


 Rabia Haq<sup>a,\*</sup>, Alexandra Hotca<sup>b,\*</sup>, Aditya Apte<sup>a</sup>, Andreas Rimner<sup>b</sup>, Joseph O. Deasy<sup>a</sup>, Maria Thor<sup>a</sup>
<sup>a</sup> Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York NY 10017, USA

<sup>b</sup> Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York NY 10017, USA

## ARTICLE INFO

## Keywords:

 Semantic segmentation  
 Convolutional neural networks  
 Cardio-pulmonary  
 Clinical outcomes  
 Radiation therapy

## ABSTRACT

**Background and purpose:** Radiation dose to the cardio-pulmonary system is critical for radiotherapy-induced mortality in non-small cell lung cancer. Our goal was to automatically segment substructures of the cardio-pulmonary system for use in outcomes analyses for thoracic cancers. We built and validated a multi-label Deep Learning Segmentation (DLS) model for accurate auto-segmentation of twelve cardio-pulmonary substructures. **Materials and methods:** The DLS model utilized a convolutional neural network for segmenting substructures from 217 thoracic radiotherapy Computed Tomography (CT) scans. The model was built in the presence of variable image characteristics such as the absence/presence of contrast. We quantitatively evaluated the final model against expert contours for a hold-out dataset of 24 CT scans using Dice Similarity Coefficient (DSC), 95th Percentile of Hausdorff Distance and Dose-volume Histograms (DVH). DLS contours of an additional 25 scans were qualitatively evaluated by a radiation oncologist to determine their clinical acceptability. **Results:** The DLS model reduced segmentation time per patient from about one hour to 10 s. Quantitatively, the highest accuracy was observed for the Heart (median DSC = (0.96 (0.95–0.97))). The median DSC for the remaining structures was between 0.81 and 0.93. No statistically significant difference was found between DVH metrics of the auto-generated and manual contours (p-value  $\geq 0.69$ ). The expert judged that, on average, 85% of contours were qualitatively equivalent to state-of-the-art manual contouring. **Conclusion:** The cardio-pulmonary DLS model performed well both quantitatively and qualitatively for all structures. This model has been incorporated into an open-source tool for the community to use for treatment planning and clinical outcomes analysis.

### 1. Introduction

Lung cancer is currently the most commonly diagnosed cancer in the world, resulting in an estimated 1.7 million deaths in 2018 alone. Non-small-cell lung cancer (NSCLC) accounts for up to 85% of all lung cancers, with radiotherapy (RT) as the most commonly prescribed treatment option [1]. Various studies have shown that the dose to the cardio-pulmonary system is critical for survival following RT for NSCLC [2,3]. Following the results of the Radiation Therapy Oncology Group 0617 study, it was discovered that overall survival of NSCLC patients is linked to heart dose [4], with higher death rate associated with electrocardiographic changes at 6 months for patients receiving cardiac radiation doses greater than 63 Gy [5]. Others also discovered that poor survival is attributed to irradiation of particular constituents of the

cardio-pulmonary system, particularly the atria and the pericardium receiving average and mean to hottest dose greater than 45 and 55 Gy respectively [6].

Currently, segmentation of cardio-pulmonary structures other than the whole heart and lung is overlooked, and only these two organs are routinely defined as part of the treatment planning process. This process requires robust and accurate segmentations in order to maximize radiation to the tumor while sparing the normal tissue as much as possible. The introduction of a new set of structures puts requirements on both segmentation accuracy and segmentation time that would clinically result in an overhead of several hours of manual segmentation and contour refinement, demonstrating the need for a robust automatic segmentation method.

In the clinical workflow, structure delineation is currently

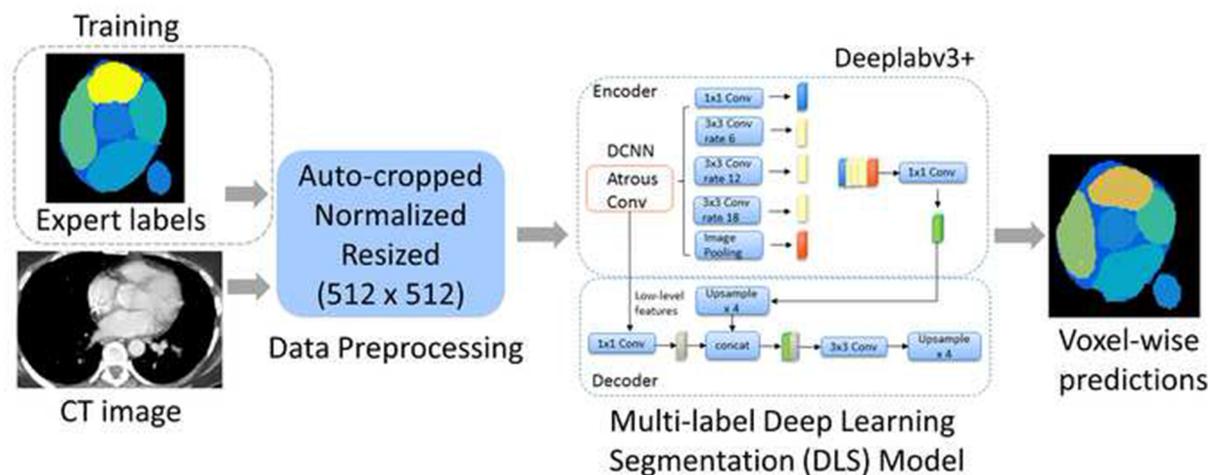
\* Corresponding author.

E-mail addresses: [haqr@mskcc.org](mailto:haqr@mskcc.org) (R. Haq), [hotcacha@mskcc.org](mailto:hotcacha@mskcc.org) (A. Hotca), [apte@mskcc.org](mailto:apte@mskcc.org) (A. Apte), [rimnera@mskcc.org](mailto:rimnera@mskcc.org) (A. Rimner), [deasyj@mskcc.org](mailto:deasyj@mskcc.org) (J.O. Deasy), [thorm@mskcc.org](mailto:thorm@mskcc.org) (M. Thor).

<https://doi.org/10.1016/j.phro.2020.05.009>

Received 25 February 2020; Received in revised form 21 May 2020; Accepted 29 May 2020

2405-6316/ © 2020 The Authors. Published by Elsevier B.V. on behalf of European Society of Radiotherapy & Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



**Fig. 1.** Schematic overview of the proposed deep learning multi-label segmentation scheme. The network is trained on 2D CT images that are auto-cropped around the lung region of interest, augmented and batch normalized for dense voxel-wise label prediction.

performed either manually or automatically through various clinically-approved segmentation frameworks [7,8]. Manual delineation is time-consuming and prone to inter- and intra-observer variability, resulting in poor repeatable performance. In contrast, fully automated segmentation methods produce faster results without requiring manual interaction. Multi-atlas based automatic methods propagate voxel-wise structure labels from a set of atlases onto a new patient scan, requiring large registration computation time [9]. In contrast, deep learning methods, such as Convolutional neural networks (CNNs) and encoder-decoder neural networks have been successfully implemented to perform various medical image segmentation tasks [10–13] in order to achieve robust and reproducible results with low computation time.

The aim of this study was to develop a robust and accurate multi-label Deep Learning Segmentation (DLS) framework for automated segmentation of cardio-pulmonary substructures to enable outcomes analyses of thoracic patients treated with RT as well as input to RT planning. Therefore, we evaluated the auto-generated contours against expert delineations and investigated their acceptability as input to outcomes analysis.

## 2. Materials and methods

Our approach utilized a deep neural network for 2D segmentation of contrast as well as non-contrast enhanced thoracic CT images. The network was trained to perform multi-label prediction of eight non-overlapping, contiguous substructures: the aorta, Left Atrium (LA), Right Atrium (RA), Left Ventricle (LV), Right Ventricle (RV), Inferior Vena Cava (IVC), Superior Vena Cava (SVC) and Pulmonary Artery (PA) [14]. Additionally, separate models were trained to segment the overlapping structures such as the heart, the atria, pericardium and ventricles. Output label predictions for the multi-label segmentation network and overlapping structures were combined for each input scan, resulting in auto-segmentation of a total of 12 cardio-pulmonary substructures. Resulting 2D segmentations for each structure were combined to generate a 3D contour and were then evaluated against expert clinical contours.

### 2.1. Experimental datasets

Experimental data consisted of 241 treatment planning CT scans of patients previously treated with RT for non-small cell lung cancer at our institute [15]. The study went under the Institutional Review Board (IRB) protocol 16–142 and all patients were given written informed consents. This data consisted of contrast and non-contrast enhanced scans of varying imaging quality and resolution across different

scanners, with a kilovoltage peak (kVp) range of 120–140 applied during image acquisition for all patients. Potential image artifacts included motion artifacts and streak artifacts introduced due to the presence of calcification. Manual expert delineations of the twelve cardio-pulmonary structures for each patient scan were performed either by an expert radiation oncologist or a physicist [16]. The pericardium expert delineation also included the thymus substructure. These manual delineations were considered as the gold standard and were subsequently used for model training, validation and testing. 193 CT scans, pertaining to 80% of the total data cohort, were utilized for model training, 24 CT scans (10%) were used for model validation and the remaining 24 CT scans (10%) were used for hold-out testing. The average image resolution of the scans was  $1 \text{ mm} \times 1 \text{ mm} \times 2.5 \text{ mm}$ . These 3D CT scans were auto-cropped around the lungs in the superior-inferior as well as the anterior-posterior directions to extract the volume of interest around the heart substructures. 2D axial slices for each patient image volume were resized to  $512 \times 512$  to achieve size harmonization for the dataset and were then normalized between 0 to 255 resulting in a total of 10,284 training images. No spatial resampling was performed in order to mitigate any image and contour interpolation bias. All images were similarly preprocessed for training, validation and subsequent model testing.

During network training, expert labels along with the input training images were used as model input, which were similarly preprocessed and then used to train the DLS model. An additional dataset of 25 RT planning thoracic CT scans, for which no expert contours were available, was used for qualitative evaluation by a radiation oncologist to rate the auto-generated contour acceptability for clinical use.

### 2.2. Network architecture

A schematic overview of our data training and inference approach is presented in Fig. 1. 2D image slices of the CT scans cropped around the lungs were resized to  $512 \times 512$  and normalized as part of the pre-processing step. This image was then passed on to the network for multi-label prediction of all pixels as either being part of the image background or as one of the structure target segmentations. All images corresponding to a CT scan were segmented using the network and the output segmentation labels were stacked together to generate a 3D segmentation.

Our approach leveraged the deep neural network architecture of [17]. The Deeplab encoder-decoder network architecture with atrous separable convolutions consists of spatial pyramid pooling that encodes multi-scale contextual information to capture spatial anatomical information of contiguous structures. Dense feature maps extracted in the

last encoder network path consisted of detailed semantic information. The decoder network was able to robustly recover structure boundaries through bi-linear upsampling at a factor of 4 while applying atrous convolutions to reduce features before semantic labeling. The network input data was augmented per batch and consisted of random cropping, random horizontal and vertical flipping and rotation by ten degrees. The resulting automated-segmented 2D axial images were stacked back together to generate 3D segmentations without further post-processing. We trained the network using ResNet-101 [18] as the encoder network backbone with learning rate = 0.01 using “policy” learning rate scheduler [19], crop size =  $513 \times 513$ , batch size = 8, loss = cross-entropy, output stride = 16 for 50 epochs for dense label prediction. Our approach has been implemented using the Pytorch DL framework, with training, validation and testing performed using a Nvidia GeForce GTX 1080Ti GPU.

We also investigated the performance of various network loss functions and their influence on correct multi-label prediction. We trained our network with various segmentation losses on the same architecture backbone to account for varying structure sizes and class imbalance during training and determine the efficacy of modifying label prediction probabilities during back propagation for multi-label segmentation. The validation dataset was utilized to finetune hyperparameters, evaluate the training loss after every epoch for each model, and consequently evaluate the model steady state where a minimal validation loss was achieved for optimal model selection for each investigated loss function. The network was trained using cross entropy (CE), Multi-class Dice Loss (M-DSC), Generalized Dice Loss (G-DSC) [20] and a weighted combination of (0.5 M-DSC + 0.5CE). Cross entropy loss can be described as

$$L(\chi; \theta) = - \sum_{x \in \chi} \log p(t_i | x_i; \theta), \quad (1)$$

where  $X$  denotes the input images,  $p(t_i | x_i; \theta)$  is the pixel probability of the target class  $x_i \in \chi$  that is being predicted with network parameters  $\theta$ .

The Dice Similarity Coefficient (DSC) compares the pixel-wise overlap between the ground truth segmentation against the predicted segmentation labels, and has been widely used as a loss function during network training. The G-DSC loss is the average DSC loss for all target segmentation classes for each image averaged across a training batch, and can be calculated as

$$L_{G-DSC} = 1 - \frac{2 \sum_{n \in N} s_n p_n}{\sum_{n \in N} s_n + \sum_{n \in N} p_n} \quad (2)$$

where  $s$  is the ground truth,  $p$  is predicted output of the network and  $n \in N$  is the number of pixels. The G-DSC loss optimizes the probability of achieving maximal surface overlap between the expert versus predicted labels averaged across all structure segmentations. However, the DSC score is biased towards large structures by definition as it accounts for the total pixel overlap between the ground truth and target segmentations. Therefore, multi-class DSC Loss is introduced in order to reduce target class imbalance for smaller structures during training by individually calculating the DSC for each target class within an image, and then averaging over all target classes for each training image within a batch during each iteration. M-DSC loss is thus calculated as

$$L_{M-DSC} = \frac{1}{K} - \sum_{k \in K} \frac{2 \sum_{n \in N} s_n^k p_n^k}{\sum_{n \in N} s_n^k + \sum_{n \in N} p_n^k} \quad (3)$$

where  $k \in K$  is the number of target classes corresponding to the number of structure labels being predicted by the model. The final implemented loss function is a sum of the multi-class DSC loss and the cross-entropy loss to determine whether equally combining both the pixel-level classification accuracy along with surface segmentation quality would result in improved label prediction. This combined loss is calculated as

$$L_{(0.5M-DSC+0.5CE)} = 0.5L_{(M-DSC)} + 0.5L_{(CE)} \quad (4)$$

A model for segmenting the non-overlapping eight structures encompassing the vessels and the chambers was initially trained using different loss functions to compare and select the best model performance. The superior performing network model was then used for segmenting all twelve structures. Additional models were separately trained for the overlapping structures, resulting in a total of five models that segment all twelve cardio-pulmonary structures.

### 2.3. Model evaluation

We quantitatively evaluated the auto-generated DLS contours of 24 patients against expert delineations using two sets of metrics: by comparing the geometric evaluation metrics of the two contours and by calculating the difference between the dose-volume histograms (DVHs). A radiation oncologist qualitatively evaluated the DLS contours of an additional 25 CT scans, for which no expert contours were available, to determine the number of slices in need of adjustments for each of the non-overlapping substructures.

The quantitative evaluation of the segmentation accuracy of the DLS contours against expert contours consisted of the DSC and the 95th Percentile Hausdorff Distance (HD95 (mm)). The DSC ranges from 0 to 1 and captures the overlap between two contours A and B. The HD95 is the 95th percentile of the measure of the largest surface distance between the contours being compared. A higher DSC and a lower HD95 value indicates higher contour agreement and, thus, higher segmentation accuracy.

Further, we extracted eight DVH metrics for the auto-generated DLS and the expert labeled contours pertaining to six structures that were found to be associated with high likelihood of heart toxicity after radiation therapy [21]. These metrics were minimum or average dose-volumes received by the atria, the left atrium, the pericardium, the SVC and the ventricles. In addition, the relative percentage volume proportion receiving dosage greater than and equal to the specified dosage was compared between the two sets of contours for the heart and the left atrium. We compared the DVH values using the Wilcoxon signed-rank test to determine any statistically significant difference between the two sets of metrics (significance level set at the five percent level).

### 3. Results

Fig. 2. (a) compares the DSC evaluation metric for multi-label substructure segmentations between the four implemented network training losses. The largest performance difference was observed for the long tubular structures, such as the aorta ( $0.81$  (G-DSC + CE)  $\leq$  DSC  $\leq$   $0.93$  (CE)) and the IVC ( $0.67$  (G-DSC)  $\leq$  DSC  $\leq$   $0.81$  (CE)). However, all implemented losses performed well for the larger chamber structures, with the highest segmentation agreement achieved for the LV where image motion artifacts were least present within the chambers during image acquisition.

Fig. 2. (b) displays the DSC Score results for the 24 hold-out test CT images for all 12 substructures segmented using the CE loss. The highest segmentation accuracy was observed for the heart (median DSC = 0.96, median HD95 = 3.5 mm), while the remaining structures achieved median accuracy between ( $0.81 \leq$  DSC  $\leq$   $0.94$ ) and ( $6.0$  mm  $\leq$  HD95  $\leq$   $3.0$  mm), with the lowest HD95 surface distance accuracy observed for the Aorta (Supplementary Fig. S1). Fig. 3 displays the contour results comparing the DLS contours against expert contours for two randomly selected patients, with the worst performing segmentation results presented in supplementary figure S2.

No statistically significant difference was discovered between any of the calculated DVH metrics (Fig. 4) for all investigated structures ( $0.69 \leq$   $p$ -value  $\leq$   $0.98$ ), demonstrating high similarity between the extracted dosimetric values for the two sets of contours.

Overall, the presented model reduced substructure segmentation time from about one hour of manual contouring to approximately ten

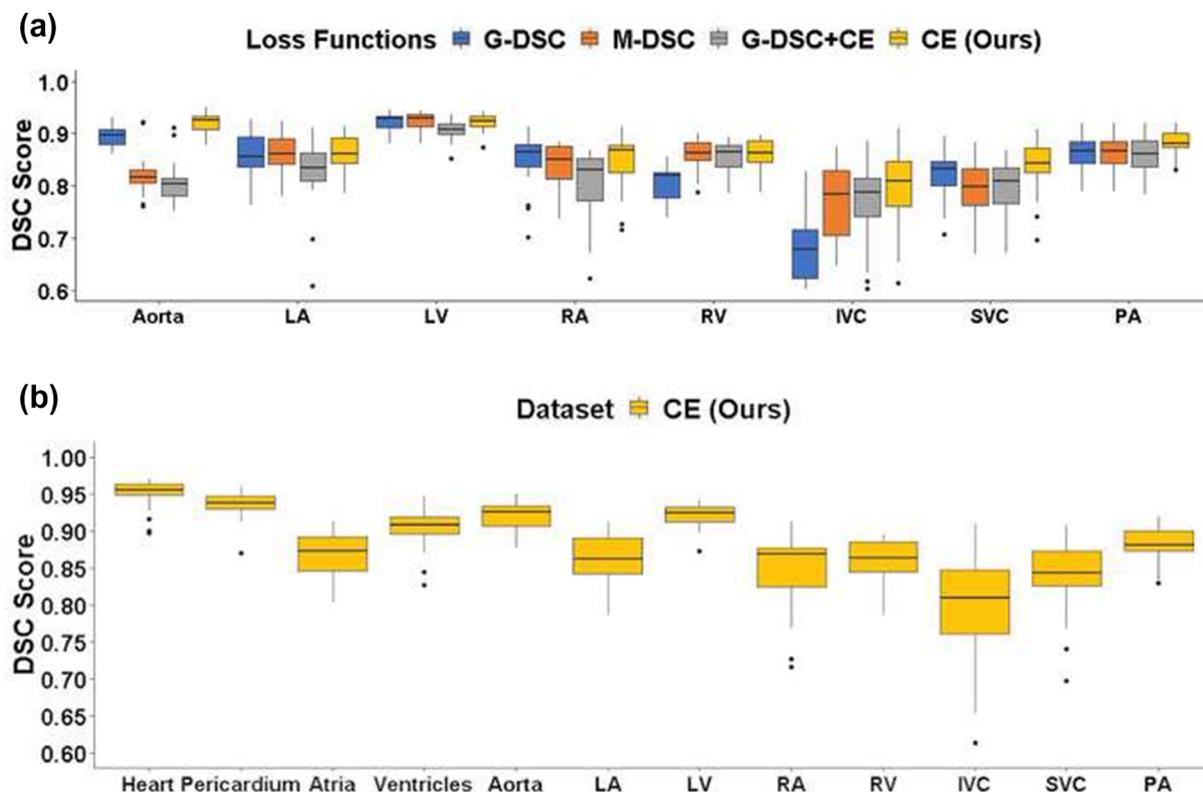


Fig. 2. (a) Multi-label segmentation comparison of eight cardio-pulmonary substructures between various network training loss configurations for 24 thoracic CT patients using the DSC evaluation metric. All training losses were implemented using the same network architecture and hyperparameters. (b) Dice Similarity Coefficient (DSC) Score results comparing the auto-generated DLS contours against expert contours for 12 cardio-pulmonary substructures using the CE network loss. LA: Left Atrium, LV: Left Ventricle, RA: Right Atrium, RV: Right Ventricle, IVC: Inferior Vena Cava, SVC: Superior Vena Cava, PA: Pulmonary Artery.

seconds per patient. Further qualitative contour scoring criteria and comparison for an additional patient cohort of 25 CT scans are presented in the [Supplementary Table S1](#) and [Figure S3](#) respectively. A

radiation oncology expert concluded that overall 85% of the auto-generated contours were acceptable for clinical use without requiring any adjustments.

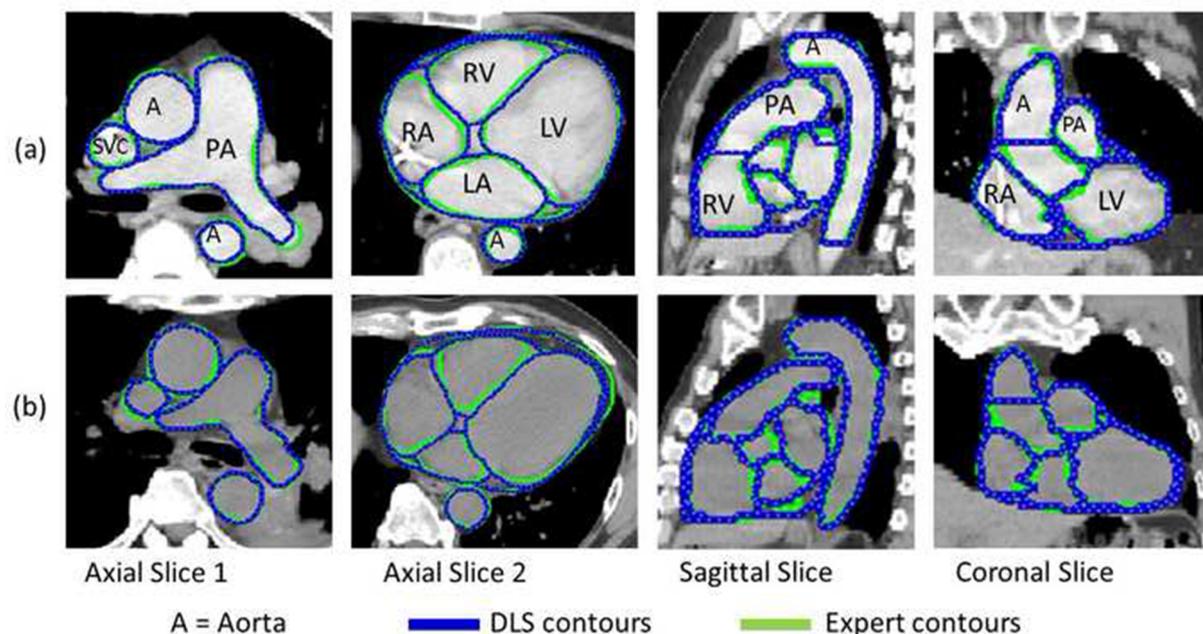
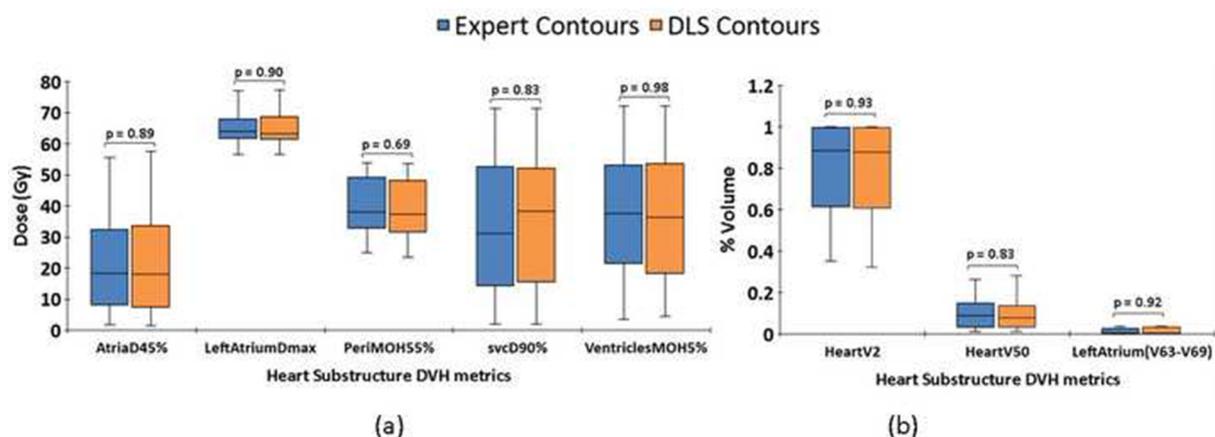


Fig. 3. Comparison of the auto-generated DLS contours (depicted in blue) against the expert delineations (depicted in green) for two patients (a) and (b) in axial, sagittal and coronal plane views. The Aorta, PA and SVC are visible in Axial Slice 1, whereas the four chambers: LA, LV, RA and RV, and the Aorta are visible in Axial Slice 2. A: Aorta, PA: Pulmonary Artery, SVC: Superior Vena Cava, LA: Left Atrium, LV: Left Ventricle, RA: Right Atrium, RV: Right Ventricle. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 4.** Dose volume histogram (DVH) comparison between expert contours and automated DLS contours for heart substructures receiving (a) Dose to the hottest percentage volume (Gy), and (b) Percentage volume receiving greater than and equal to the specified dose (% volume). Statistical comparison between the metrics was performed using the Wilcoxon signed rank-sum test. DX% = Minimum dose to the hottest X% volume. MOHX = Average dose of the hottest X% volume. VX = relative volume proportion receiving dosage = X Gy. Peri: Pericardium, svc: Superior Vena Cava.

#### 4. Discussion

Findings of the RTOG 0617 study highlighted the risk of radiation-induced heart toxicity for patients treated with radiation therapy for lung cancer. This signifies that irradiation to the heart should be a critical factor during radiation therapy treatment. This study introduces a robust automatic segmentation method for auto-contouring of cardiac substructures to facilitate further clinical outcomes analysis.

The non-overlapping substructures were more challenging to auto-segment due to the smaller volume of interest versus the overlapping structures. In addition, low anatomical-boundary contrast as well potential image artifacts introduced during image acquisition due to calcification were also present. Our experiments demonstrated that pixel-wise target class loss calculation using CE resulted in equal or superior multi-label segmentation predictions for all structures when compared against M-DSC, G-DSC and a weighted combination of (0.5CE + 0.5G-DSC). Fig. 2. (a) results demonstrate that the significant improvement in accuracy for the aorta and the long tubular structures justified selection of the CE loss function as optimal model selection. This may be because the DSC evaluation metric penalizes smaller volumetric structures during network training, which is mitigated by using the structure-specific M-DSC training loss. Geometric quantitative results using the CE loss function showed high agreement between the automatically generated and expert contours, with higher agreement for larger structures, such as the heart, the pericardium and the chambers. Our results are comparable to, or better than, the state-of-the-art deep learning [22,23] and multi-atlas [24,25] segmentation methods. Dormer et al. [23] (avg. DSC  $\leq$  0.87) used only 10 CT images to train a 3D model, which is insufficient to capture image heterogeneity with fidelity. The framework presented in [22] required paired registered non-contrast CT and MRI cardiac images (avg. DSC  $\leq$  0.88 for all chambers/vessels), whereas the multi-atlas methods required large patient-to-atlas registration time and were dependent on the similarity of the selected atlases.

Automatic deep learning segmentation methods are sensitive to domain adaptation and are dependent on the variability introduced within the training dataset. On the other hand, our qualitative contour analysis highlighted flexibility within acceptable segmentations of the heart due to the presence of motion artifacts in the CT images, due to which, sometimes the auto-generated contours were more accurate than the manual delineations. According to clinical contour guidelines, the IVC should not be contoured 0.6 mm below the last contoured image slice of the heart in the axial plane. However, due to lack of input constraints during initial network training, our model continued to

segment the IVC because of the presence of the substructure edges beyond the heart contour. Additionally, although no contour discontinuity was reported during qualitative evaluation, the model did not incorporate any contour continuity constraints during training and inference. Moreover, these auto-generated contours have not yet been validated and approved for clinical use. This highlighted the consideration towards additional requirements during network training and evaluation for generating clinically acceptable auto-segmentations.

Geometric agreement between auto-generated and manual contours is not easily translatable to clinical applicability. Therefore, we compared DVH metrics between the two sets of contours to determine whether small changes in structure volume produces any significant changes in the median DVH for larger structures such as the heart and the encompassing chambers as well as the maximum dose to the larger vessel SVC [6,26]. The DVH quantitative results showed no statistical significance between the automatically generated and manual contours (p-value  $\geq$  0.69 for all structure doses), demonstrating the efficacy of the auto-generated contours for use in heart toxicity outcomes analysis. We have applied our approach to auto-segment an additional 283 treatment planning CT scans to study heart toxicity outcomes for thoracic cancer in an effort towards improving radiotherapy treatment outcomes.

In conclusion, we have proposed a model for auto-segmentation of cardio-pulmonary substructures from contrast and non-contrast enhanced CT images. The proposed model reduced substructure segmentation time for a new patient from about one hour of manual segmentation to approximately ten seconds. We tested our approach by comparing resulting contours against expert delineation. The developed cardio-pulmonary segmentation models have being integrated into deep learning tools within the open-source CERR [27] platform.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgments

This research was partially supported by NCI R01 CA198121 and the MSK Cancer Center Support Grant/Core Grant (NIH P30 CA008748).

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.phro.2020.05.009>.

## References

- [1] Vojtiek R. Cardiac toxicity of lung cancer radiotherapy. *Reports Practical Oncol Radiother* 2020;25(1):13–9.
- [2] Dess RT, Sun Y, Matuszak MM, Sun G, Soni PD, Bazzi L, et al. Cardiac events after radiation therapy: combined analysis of prospective multicenter trials for locally advanced non-small-cell lung cancer. *J Clin Oncol* 2017;35:1395–402.
- [3] McWilliam A, Kennedy J, Hodgson C, Osorio EV, Faivre-Finn C, Herk M, et al. Radiation dose to heart base linked with poorer survival in lung cancer patients. *Eur J Cancer* 2017;85:106–13.
- [4] Bradley JD, Paulus R, Komaki R, Masters G, Blumenschein G, Schild S, et al. Standard-dose versus high-dose conformal radiotherapy with concurrent and consolidation carboplatin plus paclitaxel with or without cetuximab for patients with stage IIIA or IIIB non-small-cell lung cancer (RTOG 0617): a randomised, two-by-two factorial phase 3 study. *Lancet Oncol* 2015;16:187–99.
- [5] Vivekanandan S, Landau D, Counsell N, Warren D, Khwanda A, Rosen SD, et al. The impact of cardiac radiation dosimetry on survival after radiation therapy for non-small cell lung cancer. *Int J Radiat Oncol* 2017;99:51–60.
- [6] Thor M, Deasy J, Hu C, Choy H, Komaki RU, Masters G, et al. The role of heart-related dose-volume metrics on overall survival in the RTOG 0617 clinical trial. *Int J Radiat Oncol Biol Phys* 2018;102:S96.
- [7] MIM software; 2003. Accessed: 01–20–2020. URL:<https://www.mimsoftware.com/>.
- [8] Varian Eclipse Smart Segmentation; 1999. Accessed: 01–20–2020. URL:<https://www.varian.com/products/radiosurgery/treatment-planning/eclipse>.
- [9] Haq R, Berry SL, Deasy JO, Hunt MA, Veeraraghavan H. Dynamic multiatlas selection-based consensus segmentation of head and neck structures from CT images. *Med Phys* 2019;46(12):5612–22.
- [10] Isensee F, Petersen J, Klein A, Zimmerer D, Jaeger PF, Kohl S, et al. nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation. *CoRR*. 2018;abs/1809.10486. Available from: URL:<http://arxiv.org/abs/1809.10486>.
- [11] Oktay O, Schlemper J, Folgoc LL, Lee MCH, Heinrich MP, Misawa K, et al. Attention U-Net: Learning Where to Look for the Pancreas. *CoRR*. 2018;abs/1804.03999. Available from: URL:<http://arxiv.org/abs/1804.03999>.
- [12] Jin Q, Meng Z, Sun C, Wei L, Su R. RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans. *CoRR*. 2018;abs/1811.01328. Available from: URL:<http://arxiv.org/abs/1811.01328>.
- [13] Oktay O, Ferrante E, Kamnitsas K, Heinrich M, Bai W, Caballero J, et al. Anatomically Constrained Neural Networks (ACNNs): application to cardiac image enhancement and segmentation. *IEEE Trans Med Imaging* 2018;37(2):384–95.
- [14] Feng M, Moran J, Koelling T, Chughtai A, Chan JL, Freedman L, et al. Development and validation of a heart atlas to study cardiac exposure to radiation following treatment for breast cancer. *Int J Radiat Oncol Biol Phys* 2010;79(1):10–8.
- [15] Hotca A, Thor M, Deasy JO, Rimner A. Dose to the cardio-pulmonary system and treatment-induced electrocardiogram abnormalities in locally advanced non-small cell lung cancer. *Clin Transl Radiat Oncol* 2019;19:96–102.
- [16] Kong FM, Quint L, Machtay M, Bradley J. Atlases for Organs at Risk (OARs) in Thoracic radiation Therapy; 2015. Online; accessed 18 February 2020. Available from: URL: <https://www.rtog.org/LinkClick.aspx?fileticket=qlz0qMZxfQs%3d&tabid=361>.
- [17] Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. *ECCV* 2018.
- [18] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016. p. 770–8.
- [19] Liu W, Rabinovich A, Berg AC. ParseNet: Looking Wider to See Better. *CoRR*. 2015;abs/1506.04579. Available from: URL:<http://arxiv.org/abs/1506.04579>.
- [20] Milletari F, Navab N, Ahmadi S. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *CoRR*. 2016;abs/1606.04797. Available from: URL:<http://arxiv.org/abs/1606.04797>.
- [21] Thor M, Deasy JO, Hu C, Gore E, Bar-Ad V, Robinson E, et al. Modeling the impact of cardio-pulmonary irradiation on overall survival in NRG Oncology trial RTOG 0617. *Clinical Cancer Research*. 2020; Available from: URL: <https://doi.org/10.1158/1078-0432.CCR-19-2627>.
- [22] Morris ED, Ghanem AI, Dong M, Pantelic MV, Walker EM, Glide-Hurst CK. Cardiac substructure segmentation with deep learning for improved cardiac sparing. *Med Phys* 2020;47(2):576–86.
- [23] Dormer J, Ma L, Halicek M, Reilly CM, Schreiber E, Fei B. Heart chamber segmentation from CT using convolutional neural networks. *Proc SPIE Int Soc Opt Eng*. 2018.
- [24] Luo Y, Xu Y, Liao Z, Gomez D, Wang J, Jiang W, et al. Automatic segmentation of cardiac substructures from noncontrast CT images: accurate enough for dosimetric analysis? *Acta Oncologica* 2019;58(1):81–7.
- [25] Zhuo R, Liao Z, Pan T, Milgrom SA, Pinnix CC, Shi A, et al. Cardiac atlas development and validation for automatic segmentation of cardiac substructures. *Radiother Oncol* 2017;122:66–71.
- [26] Darby SC, Cutter DJ, Boerma M, Constine LS, Fajardo LF, Kodama K, et al. Radiation-related heart disease: current knowledge and future prospects. *Int J Radiat Oncol Biol Phys* 2010;76(3):656–65.
- [27] Apte AP, Iyer A, Thor M, Pandya R, Haq R, Jiang J, et al. Library of deep-learning image segmentation and outcomes model-implementations. *Physica Medica* 2020;73:190–6.