

		Number of patients					
		Training		Validation		Testing	
		ES	LA	ES	LA	ES	LA
	Total number	25	30	6	4	20	20
T-stage	T1	17	6	6	2	15	3
	T2	4	6			3	4
	T3	1	7			1	7
	T4	2	11		2	1	5
	TX						1
	unknown	1					
N-stage	N0	22	5	6		17	5
	N1		3			2	4
	N2	1	17		3	1	8
	N3	1	3		1		3
	NX		2				
	unknown	1					
M-stage	M0	22	26	6	3	18	18
	M1	2	4		1	2	2
	unknown	1					
Prescription	3x20Gy	10		3		8	
	4x15Gy	5					
	5x12Gy	1					
	8x7.5Gy	9		3		12	
	24x2.75Gy		15		2		10
	30x1.5Gy		2		1		2
	33x2Gy		9		1		6
	35x2Gy		4				1

Supplementary table S1: Patient characteristics

The model had a vgg16 backbone as provided by the same module with the previously discussed pretrained encoderweights. The decoder part had a 1024, 512, 256, 128, 64 structure regarding the amount of convolution filters with randomized starting values for the weights. The start of the model used three convolution filters to go from a 256x256x1 image to a 256x256x3 image, as the pretrained encoders were based on RGB images.

This basically cloned the 256x256 image so there were three of them. The activation function was a softmax and the optimizer was an Adam with a linear rectifier of 0.0001. There was no use of ensemble techniques.

OAR	$\hat{p} > 0.5$	$\hat{p} > 0.9$	$\hat{p} > 0.99$	$\hat{p} > 0.999$
All OARs	0.824	0.822	0.819	0.815
Heart	0.910	0.909	0.909	0.908
Esophagus	0.720	0.713	0.706	0.697
Lungs	0.977	0.976	0.974	0.971
Trachea	0.840	0.839	0.836	0.834
Spinal cord	0.796	0.792	0.788	0.783
Main left bronchus	0.754	0.752	0.750	0.746
Main right bronchus	0.775	0.773	0.770	0.767

Supplementary table S2: Average Dice score coefficients between the auto-segmented and manually delineated OAR for different thresholds of \hat{p} .

OAR	Mean HD95 ± STD [mm]	HD95 from previous studies on auto-segmentation [mm]	HD95 from previous studies on inter-observer variability [mm]
Lungs	6.9 ± 21.8	8.0 [39]	0.2 [44]
Trachea	11.8 ± 20.6	0.2 [23] 2.1 [42]	
Esophagus	11.6 ± 12.6	0.3 [23] 9.3 [39] 3.4 [42] 4.5 [16]	
Heart	17.6 ± 15.7	0.2 [23] 8.0 [39] 2.0 [42] 4.6 [16]	0.8 [43]
Spinal cord	29.6 ± 29.6	8.7 [39] 1.2 [16]	0.8 [39] 0.8 [43]
Main left bronchus	19.4 ± 18.1		
Main right bronchus	16.8 ± 38.4		

Supplementary table S3: 95th percentile Hausdorff distance between the manual and automatic contours for each OAR and a summary of average Hausdorff distances for automatic segmentation methods (including CNN methods) of thoracic OARs found in literature.