ESTRO

Abstract

Title

Improvement of a deep learning based automatic delineation model using anatomical criteria

Authors

Thomas Brion[1], Eleni Karamouza[2], Louis De Vitry[3], Aurélien Lombard[3], Thais Roque[3], Nikos Paragios[3], Guillaume Auzac[1], Assia Lamrani-Ghaouti[4], Naïma Bonnet[4], Elaine Limkin[1], Marjolaine Ung[1], Sophie Bockel[1], David Pasquier[5], Stéphanie Wong[6], HypoG-01 trialists[4], Samir Achkar[1], Sofia Rivera[7,1]

Authors Affiliations

[1]Institut Gustave Roussy, Radiation Oncology, Villejuif, France; [2]Institut Gustave Roussy, Biostatistics and Epidemiology, Villejuif, France; [3]Therapanacea, Artificial Intelligence, Paris, France; [4]Unicancer, Unitrad, Kremlin Bicêtre, France; [5]Centre Oscar Lambret, Radiation Oncology, Lille, France; [6]Assistance Publique des Hôpitaux de Marseille, Radiation Oncology, Marseille, France; [7]Université Paris - Saclay, INSERM U1030, Villejuif, France

Purpose or Objective

Automatic delineation models have been recently developed to reduce overall contouring time and to improve consistency in delineation process, but some OARs remain very challenging for these models to delineate accurately. In this study, we aim to demonstrate a clinically relevant improvement of an existing automatic delineation model using a blended method which combines anatomical criteria with an AI learning based approach.

Materials and Methods

ART-Plan^{TM} is a CE-marked solution for automatic delineation of OARs harnessing a unique combination of anatomically preserving and deep learning annotation concept. A first model of brachial plexus delineation, trained using 336 training cases from anonymized patients' dataset of the HypoG-01 trial, was deemed clinically unsuited, mostly due to high variability in manual contours used for training despite the use of delineation guidelines. The following 3 modifications were made to improve the clinical relevance of this model. First, the training dataset was reduced after 3 independent experts reviewed each of the 336 cases according to anatomical criteria (5 roots from C5 to T1, each starting from corresponding vertebral foramen, contour in-between the anterior and middle scalene muscles, distal part of the plexus reaching the medial border of pectoralis minor), leading to a better-quality dataset of 176 cases. Second, post-processing based on Hounsfield Unit values was added to avoid bone or air overlaps. Third, the probability threshold for the network output was lowered from 0.5 to 0.3. The first and the updated models were compared with one another and with manual delineation (MD) on an independent validation dataset from 19 HypoG-01 anonymized patients (38 brachial plexus). Blind qualitative evaluation was performed by 5 independent experts, scoring each contour as A/clinically acceptable, B/clinically acceptable after minor corrections, C/not acceptable. Interexpert agreement was evaluated by Fleiss' Kappa (k). Volumes and Dice similarity coefficients (DSC) between MD and both deep learning models were used for quantitative evaluation.

Results

Major improvement of clinically acceptable contours between the first and the updated model was observed (scores A+B=70.5% vs 0%, $p<0.0001$ and 70.5% vs 0.5%, $p<0.0001$ for right and left plexus respectively)(Fig. 1). For the first model, all experts agreed on all contour scoring but 1. For the updated one, interexpert agreement was good ($k=0.646$, $p<0.001$ and $k=0.671$, $p<0.001$ for right and left brachial plexus respectively). Mean DSC was significantly improved ($p= 0.0006$). Heterogeneity of brachial plexus volumes (in cc) was higher for MD than for deep learning models (Tab. 1).

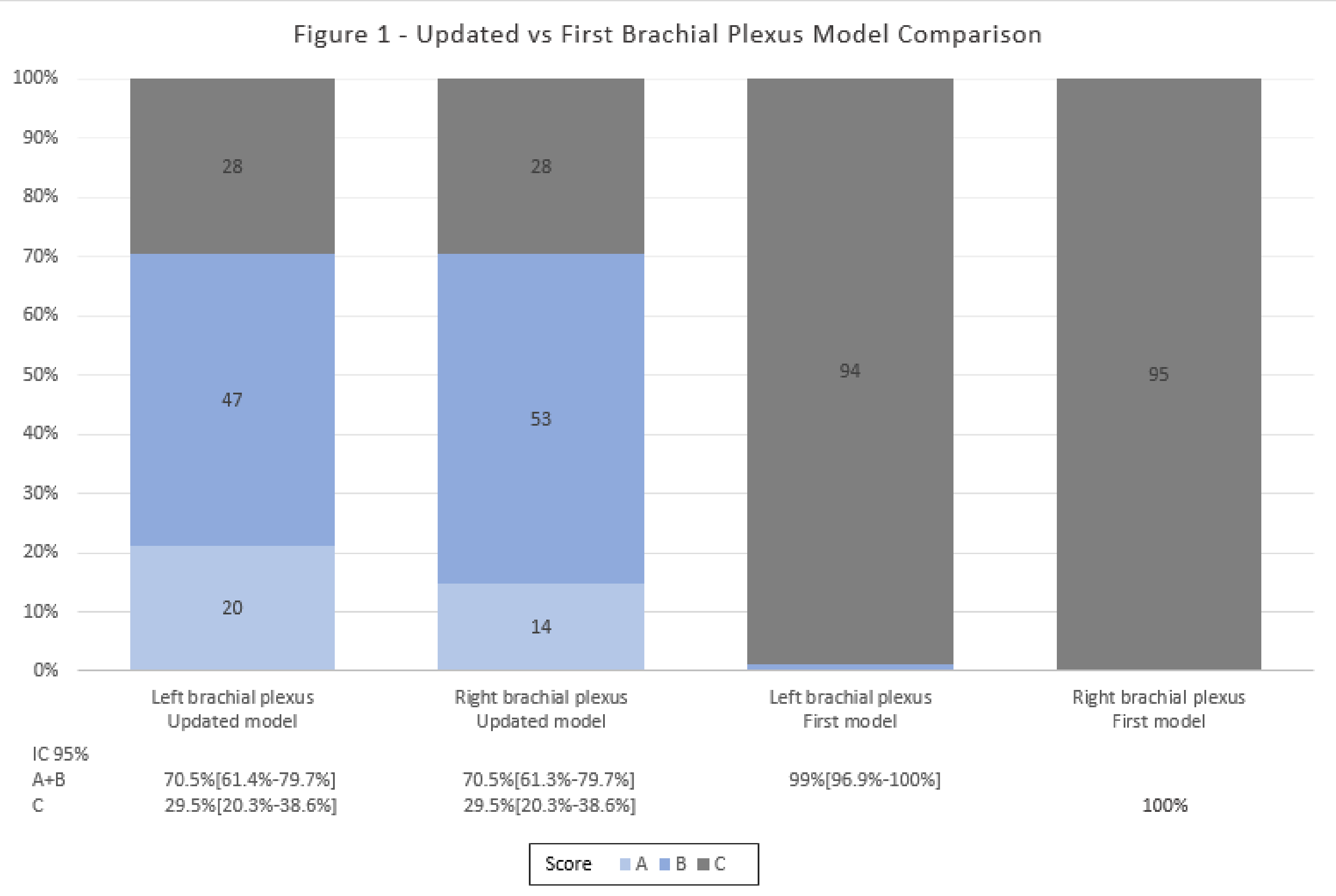

Table 1 - Volumes and Dice Similarity Coefficients

n = 19		Volumes (cc)		DSC	
Manual delineation (MD)	Mean (sd)	10.47 (7.45)			
	Median (Q1 ; Q3)	7.74 (5.53 ; 12.24)			
	Min ; Max	3.71 ; 36.53			
First model	Mean (sd)	6.71 (2.10)	p= 0.0446*	0.25 (0.11)	
	Median (Q1 ; Q3)	6.07 (5.04 ; 8.46)		0.26 (0.14; 0.31)	
	Min ; Max	2.52 ; 9.94		0.05 ; 0.42	
Updated model	Mean (sd)	15.33 (4.27)	p= 0.0160*	0.31 (0.10)	p= 0.0006**
	Median (Q1 ; Q3)	15.15 (12.31 ; 18.06)		0.30 (0.24 ; 0.35)	
	Min ; Max	5.71 ; 23.80		0.16 ; 0.54	

* p-value vs MD

** p-value vs First model

Conclusion

Both our deep learning based models decrease brachial plexus delineation variability. Combining our first learning based model with anatomical criteria leads to model improvement with a high rate of clinical acceptability. Such an approach could be used to improve automatic delineation tools for other complex OARs.