Deep learning auto contouring of OAR for HN radiotherapy: a blinded evaluation by clinical experts

Purpose or Objective

Contouring is one of the most time-consuming steps in the radiotherapy workflow. The accuracy of high precision image-guided delivery techniques is hampered by potential deviation in target and normal tissue volume delineation. Artificial intelligence can accelerate organs-at-risk (OAR) delineation and homogenize volume definition. This study aims at evaluating a commercial solution that explores an ensemble of anatomically preserving deep-learning-based networks in two radiotherapy sites with expertise in head-and-neck cancers.

Material and Methods

ART-Plan is a CE-marked solution for automatic annotation of OAR harnessing a unique combination of anatomically preserving and deep learning annotation concept. In average 6,600 samples were used for training per organ after data augmentation. Evaluation of the software was performed in two phases. In phase I, 100 patients were retrospectively selected in centres 1 and 2. For each patient, ART-Plan was used to generate full annotation of 15 OAR. Using a random selection, the contours generated from ART-Plan were blended with the ones corresponding to the clinical expert. For the whole cohort, 50% of the structures were the ones produced from the ART-Plan and the remaining ones from the expert in a random manner. Each contour was then scored by 5 experts, as A/clinically acceptable, B/clinically acceptable after minor corrections, C/not acceptable. The second phase of evaluation refers to the time gain between a fully manual delineation and one targeting to correct the outcomes of automatic contouring. This was done for 50 patients (25 patients from each centre) with respect to a full annotation of the 15 structures that were considered also in phase I.

Results

96% of all manual contours were classified as clinically relevant (75% and 21% for A and B categories respectively). Values were equal to 98% for automatic contouring (56 % and 39 % for A and B respectively). Spinal cord and oral cavity obtained better scores for automatic contouring than for manual contouring (77 % and 89 % of score A for spinal cord and oral cavity versus 65 % and 64 % for manual contouring). On the contrary, optical nerves and mandibular glands were more difficulty delineated by the automatic solution. Inter-observer variability was high between experts. Average consensus for phase I was 63% between experts ranging from 53% to 77%. The time observed to correct the automated contours was significantly inferior to the time required to generate contours fully manually. In average, 2 minutes were needed to correct the contours after auto-segmentation versus 30 minutes for manual delineation.

Conclusion

This is the first blinded, multicentric, and random back to back evaluation of an automated engine for delineation in HN tumours. The results are highly promising suggesting that this deep-learning based method should contribute to provide clinically acceptable OAR delineation. Further evaluation is on-going to quantify the dosimetric impact of the variations observed.

Authors: Vincent Grégoire1, Pierre Blanchard2, Albina Allajbej1, Claire Petit2, Nicolas Milhade1, France Nguyen2, Sofia Bakkar2, Geoffroy Boulle2, Edouard Romano2, Wael Zrafi2, Aurélien Lombard3, Eugénie Ullmann3, Nikos Paragios3, Eric Deutsch2,4, Charlotte Robert4,5.

Institutions:

1. Gustave Roussy Cancer Campus – Paris-Saclay University, Department of radiotherapy, Villejuif, France.
2. Léon Bérard Cancer Center, Department of radiotherapy, Lyon, France.
3. Therapanacea, Paris, France.
4. Molecular radiotherapy and innovative therapeutics, INSERM UMR1030, Gustave Roussy Cancer Campus, Université Paris Saclay, Villejuif, France
5. Gustave Roussy-CentraleSupélec-TheraPanacea, Center of Artificial Intelligence in Radiation Therapy and Oncology, Gustave Roussy Cancer Campus, Villejuif, France,