AI-driven quality assurance for delineation in radiotherapy breast clinical trials

Purpose or Objective

Clinical trials in radiotherapy may be subject to important uncertainties due to significant inter/intra-user variability with respect to delineation. The lack of systematic review– in particular for academic trials – and the absence of gold standard for the delineation could have a tremendous impact on trial outcomes. In this work we use artificial intelligence towards the development of a systematic, scalable and bias-free tool for quality control assessment of the delineation step.

Material and Methods

ART-Plan is a CE-marked solution for automatic delineation of 65+ ROI in radiotherapy harnessing anatomically preserving deep learning ensemble networks. In this study, ART-Plan was retrained using 256 patients from 7 investigating centers of the HYPOG-01 phase III randomized trial. HYPOG-01 data inclusion was done using a strict verification protocol. Delineations on random initial samples were assessed and evaluated according to the ESTRO breast contouring guidelines. These delineations were used to amend the ART-Plan pre-trained ensemble network towards the development of the quality assurance delineation tool. The derived solution was compared with manual delineations on an independent set of 50 patients from HYPOG-01 (Fig 1).

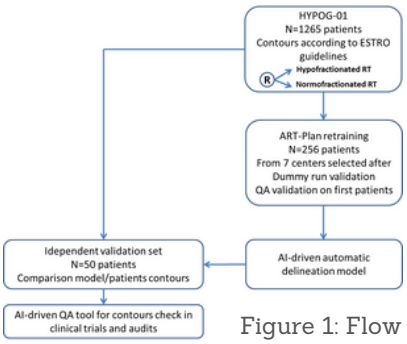

Figure 1: Flow Chart

	MDS	MCD (mm)	Average Volume difference (cc)
Lungs	0.97 ±0.02	0.51 ±0.26	+20
Liver	0.95 ±0.01	0.77 ±0.46	0
Heart	0.91 ±0.03	1.35 ±0.52	+60
Humeral heads	0.90 ±0.05	0.68 ±0.38	+1
CTV Breast/ Chest wall	0.89 ±0.06	1.48 ±0.60	+40
Esophagus	0.79 ±0.05	0.72 ±0.42	-1
Spinal cord	0.76 ±0.09	1.98 ±2.00	+19
Thyroid	0.75 ±0.09	0.78 ±0.37	0
CTVn Level 3	0.74 ±0.10	1.00 ±1.16	0
CTVn Level 1	0.72 ±0.10	2.10 ±1.16	0
Larynx	0.72 ±0.20	1.46 ±1.37	+1
CTVn Level 4	0.70 ±0.10	1.39 ±0.57	+1
CTVn Interpectoral	0.66 ±0.10	1.24 ±0.82	-1
CTVn Level 2	0.66 ±0.15	1.49 ±1.03	-2

Table 1: Quantitative results of the testing set

Results

Median Dice Similarity (MDS), mean contour distance (MCD) and average volume difference between clinical & deep learning contours were assessed (Table 1) as well as qualitative results (Fig 2).

Organs with a training MDS (MDS_TR) $\geqslant$ 0.65 and MCD_TR $\leqslant$ 2mm were included in the quality control tool developed.

Coronary artery & brachial plexus were excluded due to low MDS_TR.

The spinal cord was included despite low MDS_TR due to variability of practices (cranio-caudal axis: start/end point).

Acceptance criteria were set for testing as follows:

MDS $\geqslant$ 0.8 * MDS_TR

& MCD $\leqslant$ 1.2*MCD_TR

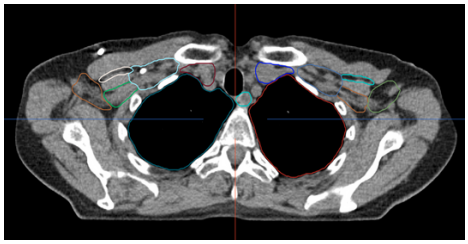

Figure 2: Automatic delineation of OAR and target volumes without editing

Conclusion

An anatomically preserving ensemble neural network retrained on high quality contours coming from a multi-center clinical trial could lead to the development of a clinically acceptable quality control delineation tool. Prospective evaluation in the 30 HYPOG-01 investigating centers is ongoing. Large scale development in breast radiotherapy trials and daily routine audits could lead to treatment standardization and systematization of contours quality assessment in trials involving radiotherapy improving reliability of the results, while saving medical expert time.

Authors: S Rivera1, A Lombard2, D Pasquier3, S Wong-Hee-Kam4, E Limkin1, G Auzac1, J Blanchecotte5, ME Chand-Fouché6, A Lamrani-Ghaouti7, N Bonnet7,N Paragios2, C Martineau-Huynh2, E Ullmann2, A Ruffier8, E Deutsch1

Institutions: 1Institut Gustave Roussy, radiotherapy, Villejuif, France; 2Therapanacea, artificial intelligence, Paris, France; 3Oscar Lambret, radiotherapy, Lille, France; 4Assistance Publique des Hôpitaux de Marseille, Radiotherapy, Marseille, France; 5Institut de cancérologie de l'ouest site Paul Papin, radiotherapy, Angers, France; 6Antoine Lacassagne, radiotherapy, nice, France; 7UNICANCER, Unitrad, Kremlin Bicêtre, France; 8Institut interrégional de cancérologie- centre Jean-Bernard, Radiotherapy, Le Mans, France.