# AI-driven Quality Insurance for organ-at-risk annotations in radiation therapy breast clinical trials

Authors : S. Rivera (CR), A. Lombard, D. Pasquier, S. Wong ,…, A. Lamrani-Ghaouti, N. Bonnet, N. Paragios, C. Martineau-Huynh, A. Ruffier

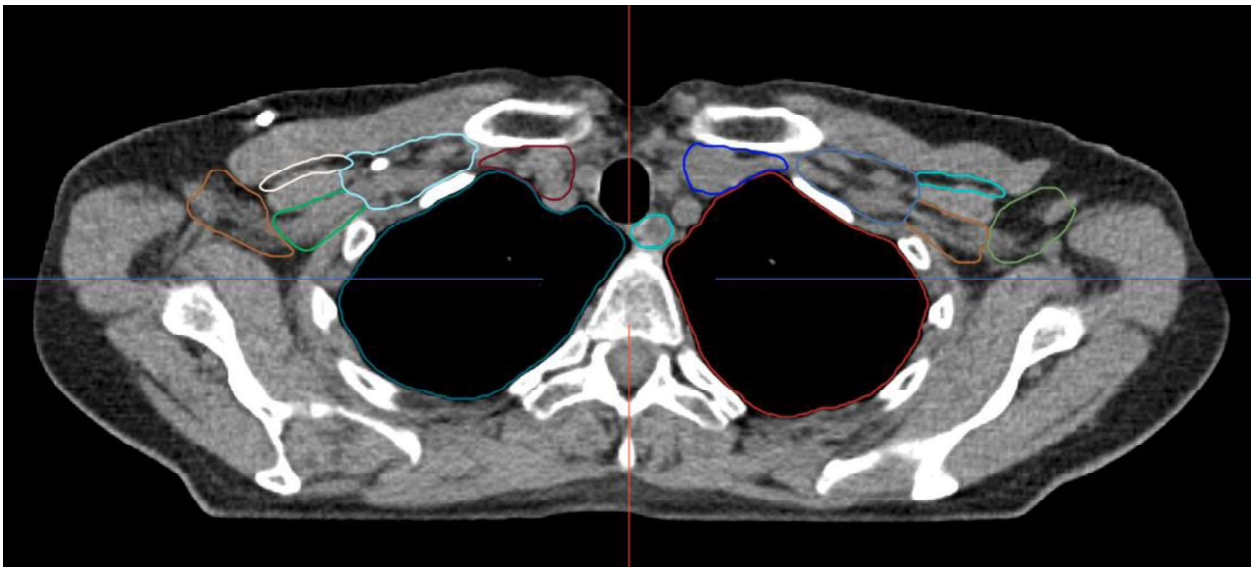
**ABSTRACT**

Purpose/Objective: Clinical trials in radiation therapy inherit strong uncertainties on their outcomes due to significant inter/intra-user variability with respect to annotations. 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 the trial's outcome. In this work, we use artificial intelligence towards the development of a systematic, scalable and bias-free tool for quality control assessment of the annotation step.

Material/Methods: ART-Plan is a CE-marked solution for automatic annotation of 65+ organs at risk in radiation therapy harnessing anatomically preserving deep learning ensemble networks. In this study, ART-Plan was re-trained using 256 patients from 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. Data from seven investigating cancer sites compose HYPOG-01. These annotations were used to amend the ART-Plan pre-trained ensemble network towards the development of the quality insurance annotation tool. The derived solution was compared with human annotations on an independent set of 50 patients from HYPOG-01.

Results: Median Dice Similarity (MDS) and Mean contour distance (MCD) between clinical & deep learning contours were assessed. Organs with a training MDS ≥ 0.65 and MCD ≤ 2mm were included to the quality control protocol (coronary artery & brachial plexus were excluded). The spinal cord was included despite low MDS due to variability of practices (z-axis start/end point). Acceptance criteria were set for testing as follows: MDSC ≥ 0.8 * MDSC_TR & MCD ≤ 1.2*MSHD_TR. Quantitative/qualitative results on the testing set are appended:

| | MDSC (range) | MCD (range) | 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 |
| 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 |
| Lymph node (LD) L3 | 0.74 ±0.10 | 1.00 ±1.16 | 0 |
| LD L1 | 0.72 ±0.10 | 2.10 ±1.16 | 0 |
| Larynx | 0.72 ±0.20 | 1.46 ±1.37 | +1 |
| LD L4 | 0.70 ±0.10 | 1.39 ±0.57 | +1 |
| Interpectoral LD | 0.66 ±0.10 | 1.24 ±0.82 | -1 |
| LD L2 | 0.66 ±0.15 | 1.49 ±1.03 | -2 |



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 clinical acceptable control annotation tool. Prospective evaluation in the 30 HYPOG-01 clinical trial is ongoing. Large scale deployment in breast radiotherapy trials and daily routine could lead to treatment standardization and systematization of contours quality assessment in trials involving radiotherapy ensuring a higher reliability of the results, while saving medical expert time .