

Evaluation of ART-PlanTM autocontouring software for head and neck radiotherapy: A blinded assessment

1/3

Tom Young^{1,2}, Victoria Butterworth¹, Sarah Misson-Yates¹, Mary Leil^{1,2}, Anthony Kong^{1,2}, Imran Petkar^{1,2}, Miguel Reis Ferreira^{1,2}, Delali Adjogatse^{1,2}, Andrew King³, Teresa Guerrero Urbano^{1,2}

1 Guy's and St Thomas' NHS Foundation Trust, Radiotherapy, London, United Kingdom.

2 King's College London, School of Cancer & Pharmaceutical Sciences, London, United Kingdom.

3 King's College London, School of Biomedical Engineering Imaging Sciences, London, United Kingdom.

Purpose/Objective:

Contouring of organs-at-risk (OARs) and target volumes is a key task within the radiotherapy (RT) workflow but is time-consuming and subject to both inter- and intra-observer variability (1). Sub-optimal contouring has been shown to affect survival and toxicity outcomes (2). Software using Artificial Intelligence (AI) to automatically delineate OARs and elective nodal regions has been developed, with several commercial solutions now available.

The National Institute for Clinical Excellence (NICE) is supporting the use of these technologies to aid delineation, but has highlighted the need for real-world data to demonstrate quality and time-savings by using autocontouring software (3). We evaluated a commercially available autocontouring software, TheraPanacea's ART-PlanTM, to contour Head and Neck Cancer (HNC) radiotherapy OARs and elective nodal regions within the real-world setting of a busy NHS radiotherapy department.

Material/Methods:

RT treatment planning data was reviewed for 60 HNC patients previously treated with radical intent at our centre in 2022. Cases were randomly selected to reflect a typical radical HNC RT workload in terms of primary tumour site and radical treatment setting (both definitive and post-operative RT). ART-Plan was used to generate AI contours for OARs and elective nodal regions for these patients. AI-generated contours were then compared to clinician-delineated contours that had been used clinically and undergone peer review. OARs assessed were parotid glands, spinal cord, brainstem, optic chiasm, optic nerves, retinas, lenses, pituitary gland, oral cavity, larynx and pharyngeal constrictors. A geometric metric, the volumetric Dice Similarity Coefficient (vDSC), was calculated to quantitatively compare clinician-generated and AI-generated contours. Furthermore, a blinded assessment was performed. Five experienced HNC consultant Clinical Oncologists reviewed both clinician-generated (previously clinically used) and AI-generated contours in a blinded manner for a subset of 10 patients with a range of different primary tumour sites. For each contour, clinicians were asked to select the contour colour they overall considered to be better, or select "no difference" if they felt there was no significant difference between contours.

Evaluation of ART-PlanTM autocontouring software for head and neck radiotherapy: A blinded assessment

2/3

Results:

Median vDSC between clinician-generated and AI-generated contours ranged from 0.23 to 0.86. For elective nodal regions, vDSC was 0.76. For OARs vDSC values were: Left parotid 0.84, right parotid 0.83, spinal cord 0.80, brainstem 0.83, optic chiasm 0.23, left optic nerve 0.50, right optic nerve 0.54, optic pathway 0.37, left retina 0.87, right retina 0.88, lens 0.70, right lens 0.72, pituitary 0.52, oral cavity 0.86, pharyngeal constrictors 0.30, and larynx 0.45. When optic chiasm and nerves were combined to form the optic pathway as one structure, vDSC was 0.37.

For the blinded assessment, a total of 121 contours were assessed (Table 1). When considering all contour assessments, 31.4% of assessments preferred clinician-generated contours, 32.9% preferred AI-generated contours, and 35.7% felt there was no difference between the two contours. For specific structures, clinician-generated contours were the most frequently preferred choice for larynx (69%), oral cavity (73%) and pharyngeal constrictors (84%). AI-contours were the most frequent choice for optic chiasm (69%), elective nodal regions (53%), and brainstem (42%).

	% Clinician contour preferred	% AI- generated contour preferred	% No Difference
Larynx	69%	21%	11%
Left Parotid	36%	34%	30%
Right Parotid	38%	30%	33%
Spinal cord	30%	34%	36%
Brainstem	36%	42%	22%
Optic chiasm	29%	69%	3%
Left Optic nerve	20%	38%	43%
Right Optic nerve	12%	34%	54%
Left Retina	18%	8%	75%
Right Retina	9%	14%	77%
Left Lens	5%	13%	82%
Right Lens	3%	19%	79%
Pituitary	34%	28%	38%
Oral Cavity	73%	23%	3%
Pharyngeal constrictors	84%	12%	4%
Prophylactic Nodal levels	40%	53%	7%
Overall	31.4%	32.9%	35.7%

Table 1. Clinician preference for clinician-generated versus AI-generated contours for different HNC radiotherapy structures (highest preference highlighted).

For the 121 contours assessed, all five clinicians were unanimous in their agreement for 33 (27.3%) contours and at least four clinicians agreed for 60 of the contours (49.6%). For individual clinicians, the median preference for clinician-generated contours was 34.6% (range 26.3-40%), AI-contours was 35.2% (15.3-39.2%) and no difference was 35.6% (22-49%).

Evaluation of ART-PlanTM autocontouring software for head and neck radiotherapy: A blinded assessment

Conclusion:

Our results show vDSC in keeping with previous values reported for commercially available Deep Learning contouring softwares for most structures investigated (4), although poorer vDSC results were seen for optic chiasm and pharyngeal constrictors in particular.

The blinded assessment showed that in 69% of cases, AI contours were either preferred to or judged not be different to clinician-generated contours that had previously been used clinically for real-world treatment. While AI-generated optic chiasm contours were more frequently preferred to clinician-generated contours, the opposite was seen for pharyngeal constrictors as well as larynx and oral cavity. This finding is likely explained by differences between the local delineation protocol for these structures and that used to delineate them for model training.

Our finding that there was disagreement by at least two of the five experts for 51.6% of contour assessments is likely to be representative of previously reported inter-observer variability (1), and highlights a potential benefit of standardised autocontouring solutions to aid delineation.

Our findings suggest that ART-Plan would not worsen current standards of contouring for most OARs and elective nodal regions and warrants prospective qualitative and quantitative evaluation to assess its impact on the radiotherapy pathway.

References:

- [1] Franzese C et al., "Enhancing Radiotherapy Workflow for Head and Neck Cancer with Artificial Intelligence: A Systematic Review", J Pers Med. 2023 Jun 2;13(6):946. doi: 10.3390/jpm13060946. PMID: 37373935; PMCID: PMC10301548.
- [2] Segedin B et al., "Uncertainties in target volume delineation in radiotherapy - are they relevant and what can we do about them?", Radiol Oncol (2016) 50(3):254-625. doi: 10.1515/raon-2016-0023
- [3] National Institute of Clinical Excellence: Artificial intelligence technologies to aid contouring for radiotherapy treatment planning: early value assessment, published 27 September 2023. Available from: <https://www.nice.org.uk/guidance/htel1>
- [4] Doolan PJ et al., "A clinical evaluation of the performance of five commercial artificial intelligence contouring systems for radiotherapy", Front Oncol. 2023;13:1213068. Published 2023 Aug 4. doi:10.3389/fonc.2023.1213068