Evaluating the efficacy of an artificial intelligence (AI) autocontouring algorithm for abdominal organs at risk (OARs) delineated during liver stereotactic body radiotherapy (SBRT) planning

Matthew Lee[a,b], Han Cheng[a], Minxuan Yuan[a], Olivia Chohan[a,b], Charles-Antoine Collins-Fekete[a], Maria Hawkins[a,b], Douglas Brand[a,b]

a. Department of Medical Physics and Bioengineering, University College London, London
b. Department of Clinical Oncology, University College London Hospital NHS Foundation Trust, London

BACKGROUND

- During radiotherapy planning, clinicians contour the cancer target and OARs on patient cross-sectional imaging. This is predominantly done manually and can be time-consuming and subjective.
- Artificial intelligence models permit automated OAR segmentation, offering time-savings and greater reproducibility.
- Commercial auto-contouring tools are being rapidly adopted for OAR contouring, although comparative data with humans is sparse in some sites including the upper abdomen. Development of data for open-source algorithms might help support resource constrained healthcare system.
- Our aim is to compare the performance of an open-source algorithm against human contours for upper abdominal OARs.

METHODS

Patients treated with SBRT for hepatocellular carcinoma were identified from the local database between May 2020 and July 2023. 65 CT scans were pseudonymised and extracted in DICOM format. Manual segmentations by consultant clinical oncologists were obtained as ground-truth references. Totalsegmentator (version 2.3.0, University Hospital Basel) was the chosen open-source autocontouring software.[1] Organs autocontoured for comparison were the duodenum, stomach, left and right kidneys and liver. Given potential variability in delineation of the border of stomach/duodenum, a combined structure was also created.

Outputted AI OARs were truncated cranio-caudally to match the human contour (since humans typically only contour OARs near the target). Dice similarity coefficient (DSC), Hausdorff, and 95th percentile Hausdorff distance were calculated. Python v3 was used for all analyses.

RESULTS

Participant demographic distribution is listed in **table 1**. 51/65 (78.5%) of participants were male. Median age was 71 years (interquartile range 9.4). **Table 2** summarises the contouring performance metrics of the TotalSegmentator algorithm.

Age Distribution / Years	
Median	71
Mean	69
Interquartile Range	9.4
Gender Distribution / %	
Male	78.5
Female	21.5
Cause of Chronic Liver Disease / %	
Viral	36.4
Non-alcoholic steatohepatitis	34.1
Mixed	20.5
Other/Unknown	9.1

Organ	Mean DSC (± 95% CI)	Mean Haudorff (± 95% CI)	Mean 95th percentile Hausdorff (± 95% CI)
Duodenum	0.41± 0.059	63.67± 9.40	50.50 ± 9.11
Stomach	0.89 ± 0.012	29.36 ± 5.14	11.41± 3.67
Duodenum & Stomach	0.88 ± 0.010	25.01 ± 3.74	6.07 ± 1.17
Left Kidney	0.92± 0.007	13.86 ± 0.97	5.99 ± 0.77
Right Kidney	0.92 ± 0.130	15.32 ± 2.19	6.69 ± 1.84
Liver	0.94 ± 0.005	20.11 ± 3.30	4.77 ± 1.63

Table 1 (left): *Participant demographic distribution.*
Table 2 (above): *Dice, Mean Hausdorff, and 95th percentile Hausdorff scores for each organ for the Totalsegmentator algorithm in comparison to human contour ground truth.*

Performance was acceptable (DSC ≥ 0.88) on all organs except duodenum. Duodenal performance was greatly improved by combination with the stomach; on visual inspection due to discrepancy in transition point between stomach/duodenum (**figure 1**).

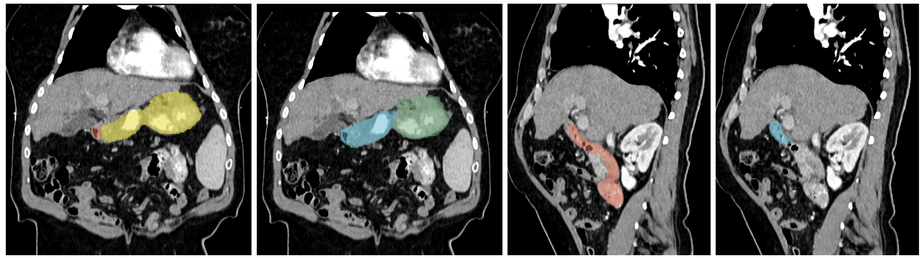


Figure 1 a-d (left to right*): Example of the splitting of stomach and duodenum differing between AI algorithm and human contour. Figures 1a) (Totalsegmentator: yellow = stomach, red = duodenum) and 1b) (human contour: green = stomach, blue = duodenum) illustrate the different transition points between the stomach and duodenum along an identical coronal plane. Figures 1c) (Totalsegmentator: duodenum) and 1d) (human contour: duodenum) shows differing inferior points of duodenal contour between Totalsegmentator and human contour along an identical sagittal plane (before truncation).*

KEY FINDINGS

- TotalSegmentator had good performance against the human ground truth. As an open-source tool this may be beneficial for treatment settings where a commercial algorithm is not economically viable.
- A key clinical finding is the potential disagreement between human and auto-contouring tool in the position of the duodenal/stomach junction. We highlight this as important to consider when using autocontouring tools, particularly if the duodenum or stomach is a dose critical structure (e.g. for upper abdominal SBRT). This is illustrated in the UK 2022 consensus for SBRT normal tissue dose-volume constraints, with differing mandatory constraints between stomach and duodenum for 1 and 3 fractions.[2]

LIMITATIONS

- Only one open-source algorithm was evaluated.
- Limited generalisation due to small sample size and restriction to cross-sectional images of one specific pathology from a single centre.
- Dosimetric impact or clinical acceptability were not evaluated.
- Future comparisons of open-source and commercial algorithms should encompass larger datasets with heterogeneity in images and pathologies, as well as compare more algorithms.

REFERENCES

1. Wasserthal J, Breit HC, Meyer MT, et al. TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images. *Radiology: Artificial Intelligence*. 2023;5(5):e230024. doi:10.1148/ryai.230024
2. Diez P, Hanna GG, Aitken KL, et al. UK 2022 Consensus on Normal Tissue Dose-Volume Constraints for Oligometastatic, Primary Lung and Hepatocellular Carcinoma Stereotactic Ablative Radiotherapy. *Clinical Oncology*. 2022;34(5):288-300. doi:10.1016/j.clon.2022.02.010