Qualitative Evaluation of AI Contours

Radim Barta 1, Ian Nygren 1, Charles Kirkby 2, Conor Shaw 3, Samuel See 4


Corresponding author: Radim Barta, radim.barta@albertahealthservices.ca

Categories: Medical Physics, Radiation Oncology, Healthcare Technology

Keywords: radiation therapy contouring, ct based delineation, ct (computed tomography) imaging, image quality, organ at risk, artificial intelligence (ai), radiation therapy treatment planning, treatment planning automation, auto-segmentation

How to cite this poster

Abstract

Purpose:

The differences between prostate CT contours generated by artificial intelligence (Limbus Contour, Limbus AI, Regina, SK, Canada) and those expertly drawn are characterized qualitatively.

Materials and Methods:

Patient CT scans imaged at three Alberta sites were contoured with Limbus Contour and independently by dosimetrists. For a minority of patients, the Limbus Contours were available to dosimetrists during the contouring step, allowing editing or re-contouring as deemed appropriate. The clinical contours were compared to the Limbus contour using Dice-coefficient (DC) and distance-to-agreement (DT) metrics. Further the contours were qualitatively reviewed to identify volumes of disagreement, and the potential for downstream clinical impact.

Results:

Quantitative measurements are in line with past studies of U-Net algorithm based auto-segmentation. Qualitatively, auto-segmentation contours differ from the clinical contours for different reasons. The agreement for bladder (DC=97%, DT=4.3 mm) speaks to its relatively consistent shape and position in the pelvic region, as well as soft tissue contrast. Auto-segmentation generates overlapping bladder and prostate, and this may contribute to disagreement in prostate contours (90%, 7.7 mm) when clinically mutual exclusivity is enforced.

Disagreement for the high-contrast femurs (83%, 12.1 mm) results from variation in the number of slices contoured. While auto-segmentation tends to contour the full extent of the femoral heads, expert-generated contours may truncate to only the relevant slices, i.e. those likely to receive doses relevant for plan optimization or assessment. This was also seen in the rectum (80%, 14 mm) and bowel bag (56%, 33 mm).

Penile bulb (59%, 6 mm) highlights a common evaluation challenge with small structures. Even single slice differences in contours lead to significant DC differences, the 6 mm DT better captures the variation seen.

Conclusions:

This study suggests auto segmentation can be used effectively to begin the contouring process. It allows trained experts to focus more time on difficult contours and defining problem areas.
Qualitative Evaluation of AI Contours

Background: Emerging Contours, Locoma AI, Negm, et al. (Canada) as a tool for the granting validation of those community imaging studies in Alberta (SSG, SPRC, and ARC) have identified the influence of automated contouring. Early contouring was performed by the community centres.

Methods: Determination of false positive and false negatives during the triennial audits of all the major imaging centres.

Results: Contours are commonly required without which planning cannot and proceed evidence applied. But no major significant changes of contours to require internal contouring protocol.

Conclusion: Multi-segmentation is a valuable tool to provide to your hands. It is important to gain an understanding of how your tool performs in clinical practice. We provide all the major community imaging centres with information on automated contouring. This tool is important to develop a working knowledge of common and uncommon errors.

Keywords: Multi-segmentation, Liver, Liver Cancer, Metabolic, Pathology, and Liver Imaging. For the publication title during AI, multi-segmentation, preparation of data for presentation, and impact of new datasets.

Contact: Emre Barta - emre.barta@albertahealthservices.ca