Can organs at risk segmented by AI-based auto segmentation algorithms be used in inverse planning

2026 JOINT AAPM | COMP MEETING
JULY 19–22 | VANCOUVER, BC

Chenyu Yan
Cleveland Clinic Foundation
yanc2@ccf.org

Cleveland Clinic
Taussig Cancer Institute

Abstract

Purpose: this study aims at investigating wehther unedited organs at risk (OARs) automatically segmented by Artificial Intelligence-based algorithms can be used by inverse optimizer to generate plans for lung SBRT cases such that the plans are still clinically acceptable after the contours has been edited by radiation oncologists.

Introduction

Segmentation of medical images aiming at locating anatomic structures and delineating their boundaries on a digital source such as computed tomography (CT) is a crucial step in radiation treatment planning. Traditionally, segmentation is done by an experienced clinical expert, and this process is often time-consuming and subject to inter-observer variations. To improve efficiency and reduce workload of clinicians, auto-segmentation has been introduced to radiation therapy and has been an active area of research for many years. Most recently, convolutional neural networks – a concept from the field of deep learning – has been applied to auto segment CT images. In this paper, we call this type of algorithm: *Artificial Intelligence (AI)-based auto segmentation algorithms*. Lustberg etc. explored the use of AI-based auto segmentation algorithm for lung cancer and concluded that user adjustment of software generated contours is a viable strategy to reduce contouring time of OARs for lung radiotherapy while conforming to local clinical standards. AI-based segmentation algorithms have been routinely used to segment CT images to generate OARs in our clinics and these OARs need to be adjusted by experienced radiation oncologists before their use in treatment planning. Even though it significantly reduced clinical workload and improved efficiency, planners still have to wait for radiation oncologists to finalize the contours before planning starts. This prompted us to ask whether unedited contours could be used in inverse planning to generate clinically acceptable plans after physician edited the contours that is whether the inverse planning can start without waiting for physician finalizing the contours. Formally the auto planning problem is defined as follows: *With the contours generated by AI-based segmentation algorithms, can the inverse optimizer generate a set of plans $\{P_1, \ldots, P_n\}$ such that after the AI-segmented contours has been adjusted by radiation oncologists, there are at least one plan which can meet all the dosimetric criteria based on the physician-edited contours*. The optimization process is aimed at minimizing the number of plans generated and simultaneously maximizing the probability of at least one plan can still meet all clinically used dosimetric criteria after physician finalizes the contours. Answering this question could potentially benefit auto planning and adaptive radiation therapy. Plans' qualities were often evaluated based on dose metrics recommended in and each dose metric may be adjusted by a physician based on his/her past experience and patient condition.

When evaluating auto segmentation performance, Dice Coefficient (DC) defined as $2 * V_1 \cap V_2/(V_1 + V_2)$, which describes the intersection of contour volume V_1 and V_2 divided by the average volume, is one of the most popular criteria. DC has a value of 1 when two contours completely overlap with each other and a value of 0 when two contours are entirely disjointed (no overlap). The use of AI segmentation algorithms may introduce new challenges of evaluating the performance of segmentation algorithms. Since AI-based segmentation algorithms is trying to simulate the human brain and it is well known that there are inter- and intra-observer variations, therefore, an AI-based segmentation algorithm might very likely have an inherent uncertainty when delineating certain structures especially in low contrast regions. These contour variations add new challenges to auto planning as well since dosimetric criteria used to evaluate plan qualities are contour dependent and the dependence of OARs and target volumes are likely different. This study mainly focused on the impact of OARs' inaccuracy (relative to physician-edited contours) on inverse planning.

Compared with forward planning, inverse planning is like a black box and non-intuitive, the impact of contour inaccuracy on inverse planning process is not straight forward and well understood. Therefore, the main aim of this study is to investigate whether unedited OARs generated by AI-based auto segmentation algorithms for lung SBRT cases could be used by inverse optimizers to generate plans such that the plan is still clinically acceptable after physician adjust/approve the OARs. Since traditional radiotherapy workflow often starts with segmentation of the planning CT image set, followed by positioning and optimization of radiotherapy beams. We are hoping that our study could be helpful for medical physics community to design a fully automatic end-to-end radiotherapy process by combining the two tasks.

Methods and Materials

Planning CT images from twenty-four patients treated with SBRT for lung cancer were selected and automatically segmented by a commercially available AI-based software. The OARs were then edited by an experienced radiation oncologist and were also revieded by a second radiation oncologist before treatment planning starts. The unedited OARs automatically generated by AI-based algorithm were used by an inverse optimizer to manually generated VMAT plans denoted as *AIContours_P*. The edited OARs were used to inversely generate clinically approved plans denoted as *Clinical_P*. The plan quality between *AIContours_P* and *Clinical_P* were compared. The unedited OARs were fed into an auto planning algorithm to automatically generate VMAT plans denote as *AIContoursAuto_P*. The plan quality between *AIContoursAuto_P* and *Clinical_P* were compared. Even though unedted OARs were used in inverse planning, only edited contours' clinically used dosimetric criteria were used in evaluating the plans' qualities.

Figure 1: Dosimetric criteria or scorecards of plans: (a) clinical plans optimized with physician edited and approved contours; (b) plans optimized with unedited OARs generated by AI-based algorithms. Both scorecards are based on physician-edited contours.

Results

All plans manually optimized with AI-segmented OARs met the clinically used scorecards while four out of the twenty-four cases optimized by the auto inverse planner failed at least one dosimetric criteria in the clinically used scorecards. For one case, the *Dice Coefficient* (DC) between unedited and physician-adjusted heart is 0.92, however, all three plans *Clinical_P, AIContours_P* and *AIContoursAuto_P* met the clinical scorecards. For another case, DC between unedited and physician-adjusted great vessel is 0.52, *Clinical_P* and *AIContours_P* could meet all dose metrics while *AIContoursAuto_P* failed at least one dose metric. For all twenty-four cases, dose differences for heart $D_{0.03cc}$ varies from -3.78 Gy (-5.61) to 1.81 Gy (4.21) with a mean value of -0.78 Gy (-0.61) between *Clinical_P* and *AIContours_P (AIContoursAuto_P)*. Dose differences for proximal bronchial tree (PBT) $D_{0.03cc}$ varies from -5.72 Gy (-5.22) to 1.47 Gy (8.41) with a mean value of -0.74 Gy (-0.38) between *Clinical_P* and *AIContours_P* (*AIContoursAuto_P*). For spinal cord, $D_{0.03cc}$ differs from -3.92 Gy (-17.12) to 1.38 Gy (5.96) with a mean value of -0.38 Gy (-0.81) between *Clinical_P* and *AIContours_P* (*AIContoursAuto_P*). Differences of $D_{0.03cc}$ for great vessel and V_{20Gy} of whole lung excluding gross tumor volume (GTV) were also evaluated.

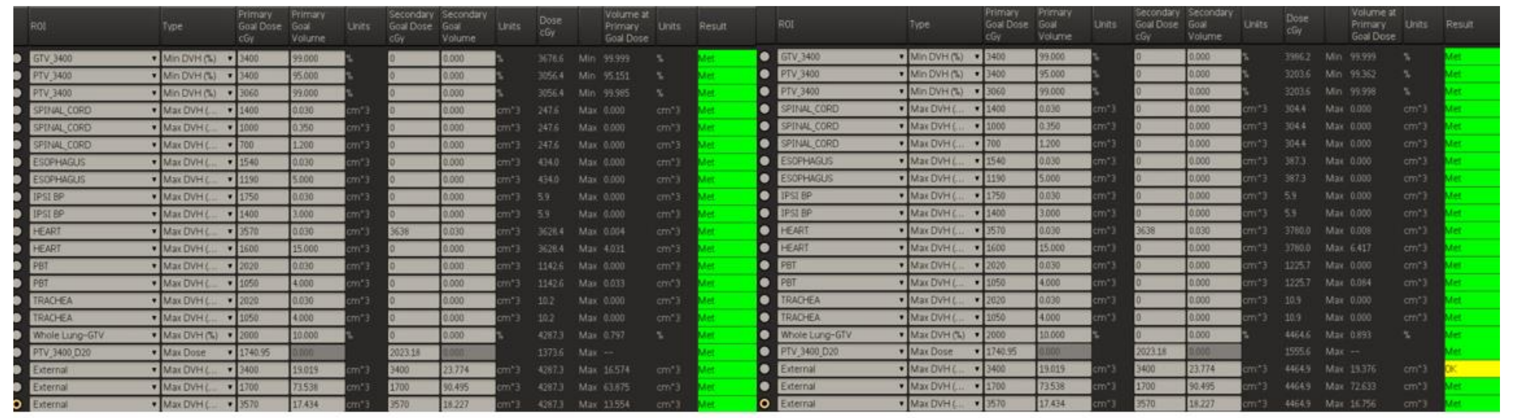

Figure 2:Scorecards for clinically approved plan (a) and automatically generated plan with unedited OAR (b).

Conclusions

It is possible to use unedited OARs segmented by AI-based algorithms to do inverse planning for lung SBRT cases. The plan can still meet all clinically used dose metrics after physician adjusted the contours. The unedited OARs can be used by an auto inverse optimizer to automatically generate clinically acceptable plans with high probability.

Discussion

The promise of tremendously reducing planning time and increasing planning consistency promoted active research in automated treatment planning process. The qualitative and quantitative evaluation of radiotherapy treatment plan varies between treatment sites, patient condition and physician experience. In our institution, physicians usually expect lung SBRT to be a hot plan and their prescribed isodose line (IDL) is normally around 70% of maximum point dose. This implies the maximum point dose could be close to 30% higher than the prescribed dose. However, for bone lesion, most radiation oncologists prefer a less hot plan, and they normally prescribe 85% IDL of maximum point dose to increase complication-free survival rate. Therefore, the evaluation of plan quality is often site dependent. Even for the same site, radiation oncologists may adjust recommended dose metrics depending on tumor location and patient condition.

In addition to the uncertainties in evaluating treatment plan quality, there are reported intra- and inter-observer contour variations. Since AI-based auto-segmentation algorithms are based on mimicking a human brain, these algorithms should inherently have contour variability as well. To objectively evaluate the performance of AI-based segmentation algorithm, a single DC between AI segmented contour and physician-drawn contour may not be sufficient. Rather, in the era of AI-based segmentation algorithm, medical physicist community might want to consider comparing the set of physician-drawn contours $\{C_{p1}, C_{p2}, \ldots, C_{pn}\}$ with the set of automatically segmented contours $\{C_{AI1}, \ldots, C_{AIn}\}$. Better understanding of the contour distributions could be very helpful in designing fully automatic inverse planning algorithms. One strategy of dealing with the challenge of contour uncertainties is to automatically generate a set of plans $\{P_{1, \ldots,} P_m\}$ such that after an experienced radiation oncologist finalize the contours, there is at least one plan that will be clinically acceptable. Ideally, the plan needs no or minimum re-optimization to meet the radiation oncologists' expectations.

Our study shows that unedited OARs for lung SBRT cases can be used to generate clinically equivalent plans even after the radiation oncologists edited the contours. This study only evaluated OARs since our clinically available AI-based algorithms only contour OARs, and treatment planning techniques used to handle tumor volume variations will be more complicated than OARs. Hence, detailed techniques of handling tumor volume variations will be presented in a separate paper. The fact that all manually optimized plans with unedited OARs can meet all clinical scorecards while four out of the twenty-four automatically generated plans failed implies that auto planning algorithm needs further improvement. It is worth mentioning that auto planning algorithms used in this study was developed more than ten years ago and there has been recent progress on this topic. A helpful message for automatic treatment planning algorithm developer may be that imperfect unedited OARs (DC less than one when compared with physician adjusted OARs) might not necessarily imply the unedited OARs cannot be used by inverse optimizers to generate clinically acceptable plans.

Contact

Chenyu Yan
Cleveland Clinic Foundation
yanc2@ccf.org
(216) 308 8761

References

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.