

Basic Original Report

Clinical Evaluation of Deep Learning and Atlas-Based Auto-Contouring of Bladder and Rectum for Prostate Radiation Therapy



W. Jeffrey Zabel, BSc,^{a,b} Jessica L. Conway, MD,^{b,c}
Adam Gladwish, MSc, MD,^{b,c} Julia Skliarenko, MD,^{b,c}
Giulio Didiodato, MD, PhD,^b Leah Goorts-Matthews, MRT(T),^b
Adam Michalak, MRT(T),^b Sarah Reistetter, MRT(T),^b
Jenna King, MSc,^b Keith Nakonechny, MSc,^b Kyle Malkoske, MSc,^b
Muoi N. Tran, PhD,^b and Nevin McVicar, PhD^{b,*}

^aDepartment of Physics and Astronomy, McMaster University, Hamilton, Ontario, Canada; ^bRoyal Victoria Regional Health Centre, Barrie, Ontario, Canada; and ^cDepartment of Radiation Oncology, University of Toronto, Toronto, Ontario, Canada

Received 3 March 2020; revised 24 May 2020; accepted 28 May 2020

Abstract

Purpose: Auto-contouring may reduce workload, interobserver variation, and time associated with manual contouring of organs at risk. Manual contouring remains the standard due in part to uncertainty around the time and workload savings after accounting for the review and editing of auto-contours. This preliminary study compares a standard manual contouring workflow with 2 auto-contouring workflows (atlas and deep learning) for contouring the bladder and rectum in patients with prostate cancer.

Methods and Materials: Three contouring workflows were defined based on the initial contour-generation method including manual (MAN), atlas-based auto-contour (ATLAS), and deep-learning auto-contour (DEEP). For each workflow, initial contour generation was retrospectively performed on 15 patients with prostate cancer. Then, radiation oncologists (ROs) edited each contour while blinded to the manner in which the initial contour was generated. Workflows were compared by time (both in initial contour generation and in RO editing), contour similarity, and dosimetric evaluation.

Results: Mean durations for initial contour generation were 10.9 min, 1.4 min, and 1.2 min for MAN, DEEP, and ATLAS, respectively. Initial DEEP contours were more geometrically similar to initial MAN contours. Mean durations of the RO editing steps for MAN, DEEP, and ATLAS contours were 4.1 min, 4.7 min, and 10.2 min, respectively. The geometric extent of RO edits was consistently larger for ATLAS contours compared with MAN and DEEP. No differences in clinically relevant dose-volume metrics were observed between workflows.

Conclusion: Auto-contouring software affords time savings for initial contour generation; however, it is important to also quantify workload changes at the RO editing step. Using deep-learning auto-contouring for bladder and rectum contour generation reduced contouring time without negatively affecting RO editing times, contour geometry, or clinically relevant dose–volume metrics.

Sources of support: none.

Disclosures: The authors have no conflicts of interest to declare.

Research data are stored in an institutional repository and will be shared upon request to the corresponding author.

* Corresponding author: Nevin McVicar, PhD; E-mail: mcvicarn@rvh.on.ca

<https://doi.org/10.1016/j.prro.2020.05.013>

1879-8500/© 2020 The Author(s). Published by Elsevier Inc. on behalf of American Society for Radiation Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

This work contributes to growing evidence that deep-learning methods are a clinically viable solution for organ-at-risk contouring in radiation therapy.

© 2020 The Author(s). Published by Elsevier Inc. on behalf of American Society for Radiation Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Prostate cancer is the most common cancer among men in Canada, and incidence is increasing rapidly due to the aging population.¹ Radiation therapy is an important treatment option for management of this disease. Accurate organ-at-risk (OAR) contours are integral to prostate radiation therapy planning. At most centers, an initial observer manually draws the bladder and rectum contours using computed tomography (CT) scans. Then, an independent radiation oncologist (RO) reviews and, if necessary, edits the contours before treatment planning. Initial contour generation is a time-consuming step that requires expertise and long periods of attention. The RO editing step is also resource-intensive and represents an important component of the quality assurance program.² Auto-contouring methods have been proposed to improve efficiency of the contouring workflow.³

Low spatial accuracy of auto-contours relative to conventional manual contours has been a barrier to widespread adoption of auto-contouring tools. This is especially true for low-contrast organs, such as the bladder or rectum, which can have a variable appearance as a result of physiological filling or gas. More recently, auto-contouring algorithms that use organ-specific spatial and radiologic features have been developed to delineate low-contrast organs on CT.³ For example, atlas-based methods have become available in commercial radiation therapy software,³⁻⁶ and machine-learning methods, such as deep learning, have been shown to accurately contour OARs for various anatomic sites.⁷⁻¹⁰ Atlas-based contouring algorithms use a library of expert contoured images, referred to as atlases, with different organ sizes and shapes that represent the range of anatomic variations. Atlas-based auto-contouring uses image registration between an atlas image and the new image to calculate the transformation, which is used to transfer atlas contours onto the new image. Atlas-based methods vary in size and quality of atlas library,¹¹ image-registration techniques (eg, rigid or elastic), contour postprocessing techniques (eg, smoothing), and the number of atlases used for auto-contouring (eg, single or multiple).¹² Deep-learning methods for auto-contouring often use convolution neural networks (CNNs). CNNs are “trained” by passing a large set of contoured images, called the training set, through a backpropagation algorithm that optimizes CNN parameters for identifying complex spatial representations related to specific objects in an image.¹³ Deep-learning

auto-contouring methods vary in the CNN architecture used (eg, U-Net¹⁴ and ResNet¹⁵), the quantity and quality of the training data set,¹⁶ and the contour postprocessing techniques.

Various studies have evaluated different auto-contouring algorithms for contouring OARs and targets in the pelvic region on CT.^{4,5,7,10,12,17-21} These studies measured geometric similarity either quantitatively or qualitatively between manual and auto-contours to validate performance of auto-contouring algorithms; however, geometric similarity does not predict changes in clinical workflow. Although auto-contouring methods can reduce workload associated with initial contour generation, questions remain about how much extra time is needed for ROs to edit auto-contours compared with manual contours. The RO editing step must be carefully evaluated when comparing processes because RO time is so valuable to health care systems. To be accepted as a clinically viable solution, auto-contours must not result in significantly more time during the RO editing step.

In this preliminary study, we assess the impact of auto-contouring on the clinical workflow for contouring the bladder and rectum in prostate radiation therapy. Specifically, we compare manual contouring versus auto-contouring using atlas-based or deep-learning-based methods. The initial contour generation step and the RO editing step are compared between workflows by both time and geometric evaluation of contours. Dosimetric differences between workflows are also assessed using RO-edited contours.

Methods and Materials

Patient selection

This study was approved by our institution’s research ethics board. Fifteen consecutive patients with prostate cancer treated at our institution from January to May 2019 were selected retrospectively for this study. For each patient, a 3-dimensional planning CT with 2-mm slice thickness was used for contouring. Patients were prepared with a full bladder and empty rectum for their scans. Contrast was not used for any patients in this study. One patient had a unilateral hip implant, and the CT scan was reconstructed with the orthopedic metal artefact reduction algorithm (v3.5.6, Philips Healthcare, Andover, MA) per institutional protocol.

Contouring workflows

Three contouring workflows were defined based on the initial contour generation method including manual (MAN), atlas-based auto-contour (ATLAS), and deep-learning auto-contour (DEEP), as described below.

- MAN contours were drawn by radiation therapists (RTs) according to Radiation Therapy Oncology Group atlas definitions²² per institutional guidelines. Semiautomated tools including interpolation and flood fill were available for use as needed. Three RTs manually contoured 5 patients each for this study. Each RT has at least 2 years of experience in OAR contouring.
- ATLAS auto-contours were generated using Smart Segmentation Knowledge-Based Contouring software (SSKC, version 13.6, Varian Medical Systems, Palo Alto, CA).¹⁸ SKCC used a library containing 27 atlases (ie, contoured CT scans), including 8 data sets provided with the software and 19 data sets randomly selected from our local database. For each patient CT, a user selects a suitable atlas based on the tumor site, contours needed, and similarity of patient anatomy. SKCC algorithm first segments the CT into tissue classes (body, bone, and air) and locates specific body regions on the CT. Finally, atlas contours are transformed onto the patient CT based on a deformable image registration between the atlas and patient CT.²¹
- DEEP auto-contours were generated using a research version of Limbus Contour software (LC, version 1.0.18, Limbus AI Inc, Regina, SK, Canada). LC uses organ-specific deep convolutional

neural network models based on a U-net architecture.¹⁰ LC models were trained using publicly available CT scans that were reviewed by one RO.²³⁻²⁵ After contour generation, LC employs a series of postprocessing steps including outlier removal, slice interpolation, z-plane cutoffs, and contour smoothing.¹⁰

SKCC was selected for this study because it is a widely accessible commercial software that functions with the Eclipse treatment-planning system (TPS, Varian Medical Systems, Palo Alto, CA). The LC software was selected because a recent study reported promising results for auto-contouring of bladder and rectum.¹⁰ Both auto-contouring softwares are available at our institution; however, they are not used clinically. ATLAS auto-contour and DEEP auto-contour generation were performed on a consumer-grade computer (3.2 GHz Intel Pentium, 8 GB memory).

The design of this workflow comparison study is illustrated in Figure 1. For each patient, 3 initial contour sets were generated (MAN_{preRO}, ATLAS_{preRO}, and DEEP_{preRO}) for geometric analysis. The duration needed to generate each contour set was recorded. All initial contours were then independently edited by 3 ROs who treat prostate cancer. Contour sets were edited one at a time and in a randomized order with ROs blinded to the method of initial contour generation. ROs were instructed to correct any contour errors that they deemed clinically meaningful. The durations for each RO to edit each contour set were recorded. RO-edited contours (MAN_{postRO}, ATLAS_{postRO}, and DEEP_{postRO}) were saved for geometric and dosimetric analysis.

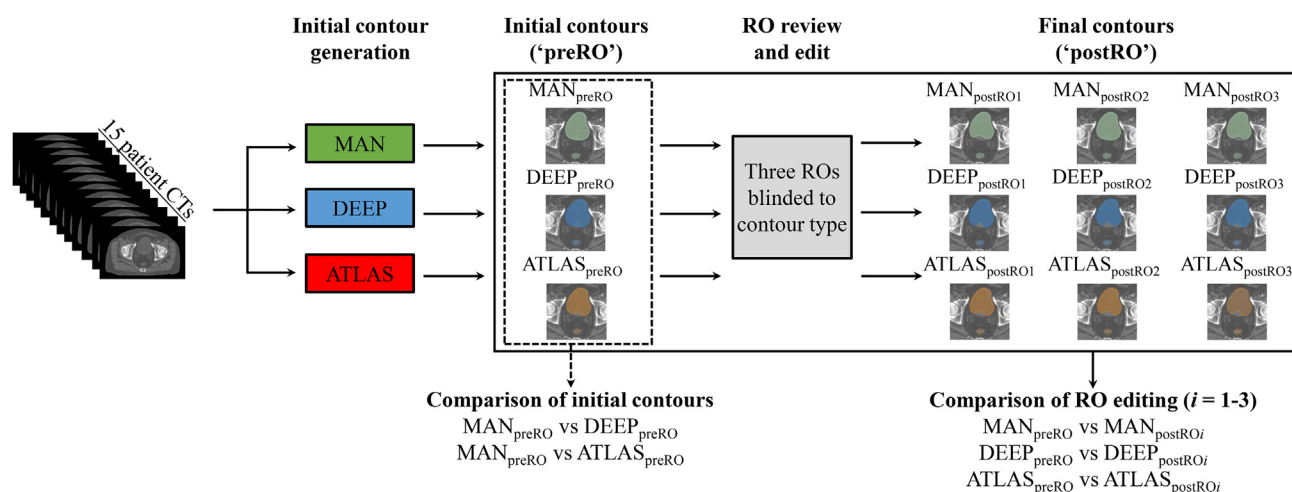


Figure 1 A flowchart of the study design. Computed tomography scans for 15 patients with prostate cancer were used to perform 3 different workflows for contouring the bladder and rectum. Each workflow is characterized by an initial contour-generation method, including manual contouring by a radiation therapist, atlas-based auto-contouring, and deep-learning auto-contouring. After the initial contour-generation step, 3 radiation oncologists independently edited the contours. Workflows were compared using recorded durations of each step, geometric contour comparisons, and dosimetric evaluation of final contours.

Workflow evaluation

The impact of the different contour-generation methods was evaluated at both the initial contour-generation step and the RO editing step. For each step, recorded durations were compared, and a geometric evaluation was performed as described later.

Comparisons of initial contour-generation step

For initial contour comparisons, ATLAS_{preRO} and DEEP_{preRO} contours were each compared with MAN_{preRO} contours. Two metrics were used to measure the similarity between initial contours.

- Dice similarity coefficient (DSC) = $\frac{2|A \cap B|}{|A| + |B|}$, where A and B are sets of voxels defining volumes within 2 contours. DSC provides a unitless measure of the volumetric overlap of 2 contours where DSC = 0 indicates no overlap and DSC = 1 indicates perfect overlap.
- Mean surface separation (MSS) is defined as the mean value of the distance transform (DT), which is the distribution of Euclidean distances between 2 surfaces measured bidirectionally. MSS describes the closeness between 2 contour surfaces in millimeters and has been referred to as the average surface Hausdorff distance.⁷

Comparisons of RO editing step

The extent of RO editing was measured for each workflow by comparing initial contours to RO-edited contours using DSC and another metric described herein.

- Acceptance rate (AR) is the fraction of the DT distribution equal to 0 mm. When DT is calculated between a “preRO” and “postRO” contour, AR is the fraction of contour surface points that the RO accepted (ie, did not edit).

Dosimetric comparison of RO-edited contours

For each patient, a treatment plan was optimized using each set of the RO-edited bladder and rectum contours (MAN_{postRO}, DEEP_{postRO}, ATLAS_{postRO}) along with the clinical target volume from the clinical contour set. A planning target volume was generated by expanding the clinical target volume by a 10-mm margin in all directions except 7 mm posteriorly, as per institutional protocol. For each contour set, a 2-arc volumetric modulated arc therapy plan was optimized to deliver 60 Gy in 20 fractions using our institution’s standard optimization objectives

(Table 1). Optimization objectives were not adjusted throughout optimization, and all plans were normalized such that 99% of the planning target volume received 95% of the prescribed dose. To facilitate plan comparisons for each patient, optimized dose distributions were superimposed onto the clinical bladder, and rectum contours and clinically relevant dose-volume metrics were calculated.²⁶ For rectum, the highest doses received by at least 15% (D_{15%}), 20% (D_{20%}), 35% (D_{35%}), and 50% (D_{50%}) were recorded in Gy. For bladder, the highest doses received by at least 5% (D_{5%}), 30% (D_{30%}), 40% (D_{40%}), and 50% (D_{50%}) were recorded in Gy.

Statistical analysis

Statistical analysis was performed using the Stata/MP software package (version 15.0, StataCorp LLC, College Station, TX). To assess significance of results, 2-tailed, unpaired Student’s *t* tests with Welch’s correction for unequal variance were performed. Statistical analysis of MSS used entire DT distribution data sets. A *P* value ≤ .05 was considered statistically significant.

Results

Comparison of initial contour-generation step

For each patient, initial contour sets were generated using each workflow (n = 45 initial contour sets). The mean ± standard deviation duration required for RTs to draw MAN contours of the rectum and bladder was 10.9 ± 4.6 min, whereas DEEP and ATLAS software required 1.4 ± 0.1 min and 1.2 ± 0.2 min, respectively.

Table 1 Optimization objectives used to generate VMAT plans for dosimetric analysis. All objectives had equal priority of 100 throughout optimization

Contour	Objective, Gy
CTV	D _{min} > 62
PTVopt*	D _{min} > 58
	D _{max} < 62
Bladder	V _{40%} < 40
Rectum	V _{15%} < 52
	V _{20%} < 48
	V _{25%} < 44
	V _{35%} < 40
	V _{50%} < 32

Abbreviations: CTV = clinical target volume; D_{min} = minimum dose; D_{max} = maximum dose; PTV = planning target volume; V_{D%} = fractional volume of a contour that received D% of prescribed dose; VMAT = volumetric modulated arc therapy.

* PTVopt is PTV plus 1 mm margin and is used for optimization only.

Figure 2 summarizes geometric comparisons of $ATLAS_{preRO}$ and $DEEP_{preRO}$ versus MAN_{preRO} contours. $DEEP_{preRO}$ contours showed better geometric similarity to MAN_{preRO} based on significantly higher DSC and lower MSS measurements for both bladder and rectum (Table 2).

Comparison of RO editing step

Three ROs edited 45 initial contour sets, resulting in 135 RO-edited contour sets in total. Mean \pm standard deviation for RO editing durations for MAN, DEEP, and ATLAS were 4.1 ± 2.5 , 4.7 ± 2.6 , and 10.2 ± 5.7 min, respectively. ATLAS contours demanded significantly more RO editing time compared with MAN contours ($P < .001$), whereas the durations for DEEP and MAN contours were not significantly different ($P = .45$). Figure 3 illustrates the required steps in each workflow based on our institution's standard procedures. The DEEP workflow required additional "file import" and "file export" steps compared with MAN and ATLAS because LC is a separate software from our TPS, whereas SKCC is a module available in our TPS. Durations for file importing, file exporting, and task scheduling were not recorded in this study; however, they were each estimated to take 10 sec. Therefore, the mean total durations for each workflow are estimated to be 15.3 min, 6.8 min, and 11.7 min for MAN, DEEP, and ATLAS, respectively.

Figure 4 summarizes the DSC and AR values calculated between initial and RO-edited contours. ROs consistently had higher AR for MAN contours compared

with ATLAS and DEEP contours. Notably, ROs edited the majority of ATLAS rectum contour points, and consequently, the geometric extent of RO editing for ATLAS rectum contours was large as seen by significantly lower DSC values (Table 3). DEEP rectum contours required more edits compared with MAN; however, ROs appeared to make a similar number of edits for DEEP and MAN bladder contours. For all workflows, ROs made more edits to rectum contours compared with bladder contours, based on consistently lower AR values.

Comparison of dosimetry for RO-edited contours

Table 4 summarizes dosimetric comparisons of plans optimized with $ATLAS_{postRO}$ and $DEEP_{postRO}$ versus plans optimized with MAN_{postRO} contours. No significant dose differences were measured between manual and auto-contour workflows.

The contouring workflows were not remarkably affected by the reconstructed CT of the patient with the unilateral hip implant. Each geometric, time, and dosimetric parameter value measured in this patient was within 2 standard deviations of the overall mean value.

Discussion

In this preliminary study, atlas-based and deep-learning auto-contouring workflows were compared with a manual contouring workflow. Previous auto-contour assessment studies have primarily focused on comparing initial auto-contours with manual contours. By analyzing

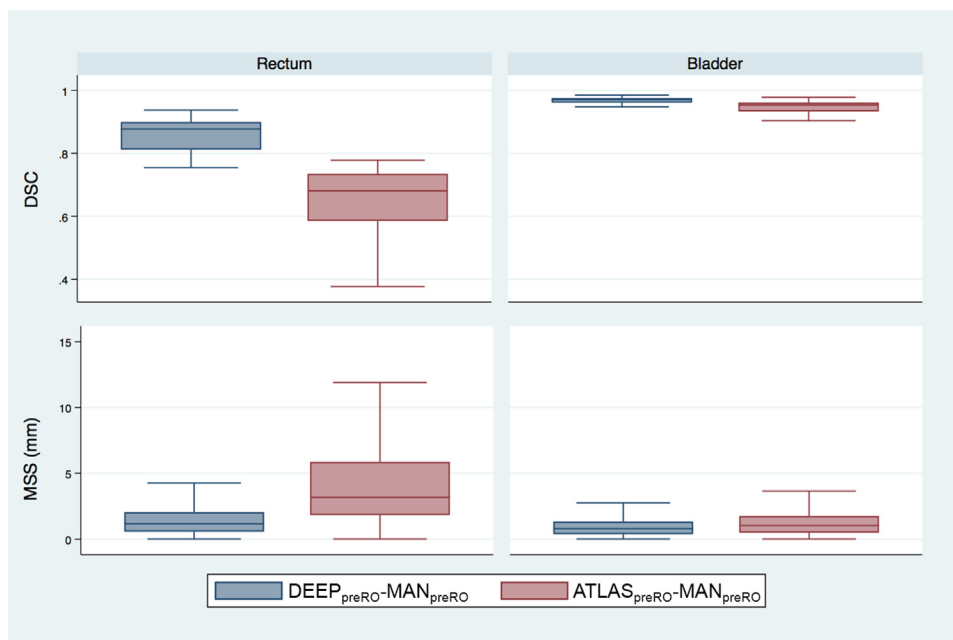


Figure 2 Geometric comparison of initial contours. Dice similarity coefficient and mean surface separation initial contour comparison data for bladder and rectum. Abbreviations: $ATLAS_{preRO}$ = initial ATLAS contour; $DEEP_{preRO}$ = initial DEEP contour; MAN_{preRO} = initial MAN contour.

Table 2 Summary of geometric comparison between initial auto-contours and manually generated contours. DSC and MSS measurements are summarized using overall mean values for rectum and bladder

	Rectum		Bladder	
	DSC	MSS, mm	DSC	MSS, mm
ATLAS _{preRO} – MAN _{preRO}	0.65	5.2	0.94	1.6
DEEP _{preRO} – MAN _{preRO}	0.86	2.7	0.97	1.1
Δ^*	–0.21	3.48	–0.02	0.51
95% CI of Δ^*	[–0.27 to –0.16]	[3.43-3.53]	[–0.04 to –0.004]	[0.50-0.52]
P value	<.001	<.001	.02	<.001

Abbreviations: ATLAS_{preRO} = initial ATLAS contour; CI = confidence interval; DEEP_{preRO} = initial deep learning contour; DSC = Dice similarity coefficient; MAN_{preRO} = initial manual contour; MSS = mean surface separation.

* Difference calculated as ATLAS_{preRO} – MAN_{preRO} minus DEEP_{preRO} – MAN_{preRO} for each metric.

the initial contour-generation step and the RO editing step, we evaluated the entire clinical contouring workflow that routinely occurs between CT acquisition and RO approval of contours for planning.

The current study design is based on several assumptions. First, contours associated with the MAN workflow are assumed to be the gold standard because this represents our department's current standard of care. Our results support this because MAN contours required the least amount of RO time for editing. It is also assumed

that RO editing ensures acceptable contour accuracy before treatment planning, and therefore the RO editing step offers a clinically relevant endpoint for contour accuracy. This was considered reasonable because ROs have extensive training and expertise in normal tissue and tumor anatomy as well as contouring,² and it was ultimately supported by the dosimetric analysis that showed no differences between RO-edited contours, regardless of initial contour-generation method. Finally, the observed time differences between workflows are assumed to estimate the effects of auto-contouring software on staff workload, which is only accurate if the time and work associated with file transfers and task scheduling are not significantly different between workflows.

Several groups have reported similar geometric parameters for comparing auto-contours to manual contours for the bladder and rectum. Overall, in our study both ATLAS and DEEP achieved initial bladder DSC values that compared favorably to the literature, and rectum DSC values were within the relatively large range (0.49-0.92) reported by similar studies.^{4,5,7,10,20,21} In one study, Gambacorta et al reported a mean DSC of 0.585 and 0.695 when comparing manual and auto-contours for the bladder and mesorectum, respectively, in patients with locally advanced rectal cancer.²¹ Interestingly, Gambacorta et al also used SSKC yet found considerably lower bladder DSC and higher mesorectum DSC compared with our results. These differences may be due to variability between SSKC atlases or versions, contour definitions, or differences in patient populations. In another study, Wong et al¹⁰ showed that Limbus Contour software generates bladder and rectum contours with accuracy comparable to expert interobserver variation for patients with prostate cancer. The geometric similarity between initial DEEP and MAN contours reported in our study agree well with that study's findings (mean bladder DSC of 0.97, mean rectum DSC of 0.78).

ROs edited significantly more ATLAS contour points compared with MAN and DEEP, resulting in over a twofold increase in RO editing time for ATLAS. Initial ATLAS rectum contours were particularly inaccurate,

CT acquisition		
MAN	DEEP	ATLAS
1) Import CT to TPS	1) Import CT to DEEP software	1) Import CT to TPS
2) Initial contour generation (10.9 ± 4.6 min)	2) Initial contour generation (1.4 ± 0.1 min)	2) Initial contour generation (1.2 ± 0.2 min)
3) Task RO to review contours	3) Export contours to TPS	3) Task RO to review contours
4) RO contour review and editing (4.1 ± 2.5 min)	4) Import CT and contours to TPS	4) RO contour review and editing (10.2 ± 5.7 min)
	5) Task RO to review contours	
	6) RO Contour review and editing (4.7 ± 2.6 min)	
Total: 15.3 min	Total: 6.8 min	Total: 11.7 min

Figure 3 Summary of all required steps for each workflow. Measured mean ± standard deviation durations are shown for “initial contour generation” and “radiation oncology contour review and editing” steps. The time required for file importing, file exporting, and task scheduling steps were each estimated to be 10 sec. Mean total durations for manual contouring by a radiation therapist, deep-learning auto-contouring, and atlas-based auto-contouring workflows were estimated to be 15.3 min, 6.8 min, and 11.7 min, respectively.

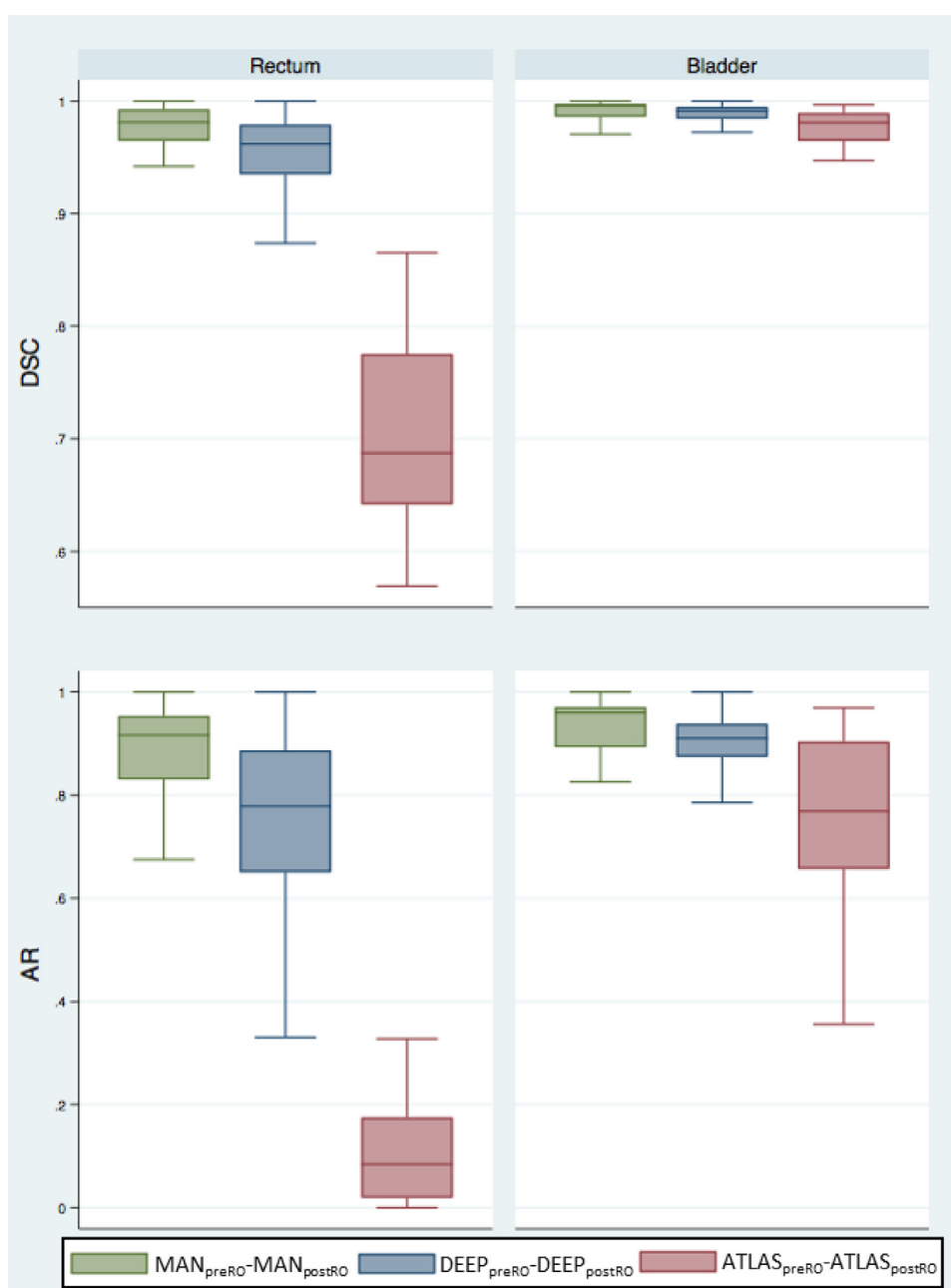


Figure 4 Geometric comparison of radiation oncologist editing for each workflow. Dice similarity coefficient and acceptance rate contour comparisons data measured between initial and radiation oncologist–edited contours for the bladder and rectum. *Abbreviations:* ATLAS_{postRO} = RO-edited ATLAS contour; ATLAS_{preRO} = initial ATLAS contour; DEEP_{postRO} = RO-edited DEEP contour; DEEP_{preRO} = initial DEEP contour; MAN_{postRO} = RO-edited MAN contour; MAN_{preRO} = initial MAN contour.

which agrees with some previous reports by other groups using SSKC.^{18,27} Mean RO editing time was slightly longer for DEEP than MAN (<1 min) but no statistical difference was measured. In a similar study, Gambacorta et al reported that atlas-based auto-contours required ~1 min more for RO editing compared with manual contours of rectal cancer target volumes.²¹

Commercial auto-contouring software predominantly uses atlas-based methods.^{5,12,17} As deep-learning methods

continue to make rapid progress, several groups have also compared deep-learning models with state-of-the-art atlas-based software. For example, both Zhu et al²⁸ and Lustberg et al⁶ compared atlas-based and deep-learning auto-contours of lung OARs by measuring geometric similarity to expert manual contours. In agreement with our findings, both of these studies reported that deep-learning auto-contours were more similar to expert manual contours. Furthermore, Lustberg et al also

Table 3 Summary of mean DSC and AR values measured between initial and RO-edited contours for comparing the extent of RO editing for the rectum and bladder

	Rectum		Bladder	
	DSC	AR	DSC	AR
MAN _{preRO} – MAN _{postRO}	0.97	0.87	0.99	0.92
DEEP _{preRO} – DEEP _{postRO}	0.95	0.76	0.99	0.89
Δ^*	0.01	0.11	0.00	0.03
95% CI of Δ^*	[−0.01 to 0.03]	[0.05-0.17]	[−0.01 to 0.00]	[0.00-0.06]
<i>P</i> value	.19	<.001	.56	.09
ATLAS _{preRO} – ATLAS _{postRO}	0.69	0.12	0.97	0.75
Δ^\dagger	0.27	0.75	0.02	0.18
95% CI of Δ^\dagger	[0.23-0.31]	[0.69-0.81]	[0.01-0.04]	[0.13-0.23]
<i>P</i> value	<.001	<.001	<.001	<.001

Abbreviations: AR = acceptance rate; ATLAS_{postRO} = RO-edited ATLAS contour; ATLAS_{preRO} = initial ATLAS contour; CI = confidence interval; DEEP_{postRO} = RO-edited DEEP contour; DEEP_{preRO} = initial DEEP contour; DSC = Dice similarity coefficient; MAN_{postRO} = RO-edited MAN contour; MAN_{preRO} = initial MAN contour; RO = radiation oncologist.

* Difference calculated as MAN_{preRO} – MAN_{postRO} minus DEEP_{preRO} – DEEP_{postRO} for each metric.

† Difference calculated as MAN_{preRO} – MAN_{postRO} minus ATLAS_{preRO} – ATLAS_{postRO} for each metric.

computed time required for initial contour generation and RO editing steps and found that both auto-contouring workflows reduced the overall workflow duration compared with manual contouring of lung OARs.⁶ In another study, Ibragimov et al geometrically compared expert manual contours to deep-learning auto-contours of head and neck OARs.⁹ That study reported that deep-learning auto-contour similarity to manual contours compared favorably to atlas-based results within the literature. In the current study, the first direct comparison of atlas-based and deep-learning auto-contouring methods applied to bladder and rectum for prostate radiation therapy is reported. In addition, our study is the first to include a dosimetric evaluation of RO-edited contours

from each workflow. Importantly, RO-edited contours from both ATLAS and DEEP workflows were found to be dosimetrically equivalent to the MAN workflow, which was expected because ROs were instructed to make edits that they deemed clinically meaningful. This finding highlights the value of the RO editing step in each workflow and suggests that clinical implementation of the auto-contour workflows presented in this study would be safe. Dosimetric analysis is important because standard geometric parameters including DSC and MSS are insufficient to determine clinical equivalence.²⁹⁻³¹

Improving the efficiency of radiation therapy treatment provides important benefits to patients and the health care system. Shorter treatment-planning times result in

Table 4 Summary of mean D_{15%}, D_{20%}, D_{35%}, and D_{40%} for rectum and mean D_{5%}, D_{30%}, D_{40%}, and D_{50%} values for bladder

	Rectum				Bladder			
	D _{15%} (Gy)	D _{20%} (Gy)	D _{35%} (Gy)	D _{50%} (Gy)	D _{5%} (Gy)	D _{30%} (Gy)	D _{40%} (Gy)	D _{50%} (Gy)
MAN	52.4	48.4	38.8	31.3	60.1	30.8	23.6	17.4
DEEP	52.4	48.5	39.0	31.4	60.3	31.0	23.8	17.5
Δ^*	−0.1	−0.1	−0.1	−0.1	−0.1	−0.2	−0.2	−0.1
95% CI of Δ^*	[−2.2 to 2.0]	[−2.2 to 2.0]	[−2.1 to 1.8]	[−3.2 to 3.1]	[−1.5 to 1.2]	[−5.2 to 4.9]	[−4.9 to 4.6]	[−4.6 to 4.5]
<i>P</i> value	.96	.93	.88	.97	.84	.95	.95	.97
ATLAS	52.1	48.0	38.1	30.3	60.8	31.1	23.7	17.5
Δ^\dagger	−0.2	−0.4	−0.8	−1.0	−0.7	−0.3	−0.1	0.0
95% CI of Δ^\dagger	[−1.9 to 2.4]	[−1.7 to 2.6]	[−1.2 to 2.7]	[−2.2 to 4.2]	[−2.0 to 0.7]	[−5.4 to 4.7]	[−4.9 to 4.6]	[−4.6 to 4.5]
<i>P</i> value	.82	.69	.42	.55	.33	.90	.95	.99

Abbreviations: ATLAS = atlas-based auto-contouring; CI = confidence interval; DEEP = deep-learning auto-contouring; D_{V%} = minimum dose to the most exposed V% of contour volume; DVH = dose–volume histogram; MAN = manual contouring by a radiation therapist.

Dosimetric parameters were measured by optimizing treatment plans on each contour set and then propagating dose onto clinically approved contour set for DVH calculation.

* Difference calculated as MAN minus DEEP for each metric.

† Difference calculated as MAN minus ATLAS for each metric.

increased patient throughput and reduced wait times. Our results suggest that deep-learning contour generation reduces the total contouring workflow time by approximately 8.5 min, with time savings afforded to the initial delineator. If a department replaced the MAN workflow with the DEEP workflow, staffing resources dedicated to initial delineation could potentially be redirected toward other tasks to improve patient care. For example, faster contouring workflows may facilitate more rigorous dose-monitoring and adaptive planning for patients who experience anatomic changes throughout their treatment.

The small sample size is a limitation of this study. In addition, the study was retrospective, and this likely affected our results. For example, both auto-contouring methods in this study reduced initial contouring time by >85% compared with manual contouring; however, this must be interpreted carefully. During the study, RTs were protected from interruptions during manual contouring, and this does not reflect reality in a busy radiation therapy department. Therefore, manual contouring times could be longer than reported. Furthermore, the actual time savings associated with initial auto-contour generation per patient depends on computing power, presence of steps requiring human input, and the number of contours. Nevertheless, significant time savings were noted, consistent with previous studies of automatic pelvic contour generation methods.^{5,10,12,21} In addition, dosimetric analysis was performed using static plans only, and therefore geometric uncertainties were not taken into account.³²

Conclusions

We have demonstrated that DEEP bladder and rectum contours result in similar RO editing time compared with manual contours, thereby suggesting that this method is an ideal candidate for further auto-contouring workflow studies in radiation therapy.

Acknowledgments

The authors would like to acknowledge the generous support of the RVH Foundation.

References

1. Canadian Cancer Society's Advisory Committee on Cancer Statistics. *Canadian Cancer Statistics 2015*. Toronto: Canadian Cancer Society; 2015.
2. Wright J, Yom S, Awan M, et al. Standardizing normal tissue contouring for radiation therapy treatment planning: An ASTRO consensus paper. *Pract Radiat Oncol*. 2019;9:65-72.
3. Sharp G, Fritscher K, Pekar V, et al. Vision 20/20: Perspectives on automated image segmentation for radiotherapy. *Med Phys*. 2014;41:050902.
4. Delpont G, Escande A, Ruef T, et al. Comparison of automated atlas-based segmentation software for postoperative prostate cancer radiotherapy. *Front Oncol*. 2016;6:178.
5. Simmat I, Georg P, Georg D, Birkfellner W, Goldner G, Stock M. Assessment of accuracy and efficiency of atlas-based auto-segmentation for prostate radiotherapy in a variety of clinical conditions. *Strahlenther Onkol*. 2012;188:807-815.
6. Lustberg T, van Soest J, Gooding M, et al. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. *Radiation Oncol*. 2018;126:312-317.
7. Kazemifar S, Balagopal A, Nguyen D, et al. Segmentation of the prostate and organs at risk in male pelvic CT images using deep learning. *Biomed Phys Eng Express*. 2018;4:055003.
8. Men K, Zhang T, Chen X, et al. Fully automatic and robust segmentation of the clinical target volume for radiotherapy of breast cancer using big data and deep learning. *Phys Med*. 2018;50:13-19.
9. Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Med Phys*. 2017;44:547-557.
10. Wong J, Fong A, McVicar N, et al. Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning. *Radiat Oncol*. 2020;144:152-158.
11. Wong W, Leung L, Kwong D. Evaluation and optimization of the parameters used in multiple-atlas-based segmentation of prostate cancers in radiation therapy. *Br J Radiol*. 2016;89:1057.
12. La Macchia M, Fellin F, Amichetti M, et al. Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer. *Radiat Oncol*. 2012;7:160-175.
13. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436-444.
14. Ronneberger O, Fischer P, Brox T. "U-net: Convolutional networks for biomedical image segmentation." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference*. Munich: MICCAI; 2015:234-241.
15. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2016:770-778.
16. Al Badawy E, Saha A, Mazurkowski M. Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing. *Med Phys*. 2018;45:1150-1158.
17. Greenham S, Dean J, Fu C, et al. Evaluation of atlas-based auto-segmentation software in prostate cancer patients. *J Med Radiat Sci*. 2014;61:151-158.
18. Huyskens D, Maingon P, Vanuytsel L, et al. A qualitative and a quantitative analysis of an auto-segmentation module for prostate cancer. *Radiation Oncol*. 2009;90:337-345.
19. Ma L, Guo R, Zhang G, et al. Automatic segmentation of the prostate on CT images using deep learning and multi-atlas fusion. *Proc SPIE Int Soc Opt Eng*. 2017;10133:1013320.
20. Geraghty J, Grogan G, Ebert M. Automatic segmentation of male pelvic anatomy on computed tomography images: A comparison with multiple observers in the context of a multicentre clinical trial. *Radiat Oncol*. 2013;8:106.
21. Gambacorta M, Valentini C, Dinapoli N, et al. Clinical validation of atlas-based auto-segmentation of pelvic volumes and normal tissue in rectal tumors using auto-segmentation computed system. *Acta Oncol*. 2013;52:1676-1681.
22. Gay H, Barthold H, O'Meara E, et al. Pelvic normal tissue contouring guidelines for radiation therapy: A Radiation Therapy Oncology Group consensus panel atlas. *Int J Radiat Oncol Biol Phys*. 2012;83:e353-e362.

23. Clark K, Vendt B, Smith KFJ, et al. The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *J Digital Imaging*. 2013;26:1045-1057.
24. Kirk S, Lee Y, Lucchesi F, et al. *Radiology data from the Cancer Genome Atlas Urothelial Bladder Carcinoma [TCGA-BLCA] collection*. Cancer Imaging Archive; 2016.
25. Zuley M, Jarosz R, Drake B. *Radiology data from the Cancer Genome Atlas Prostate Adenocarcinoma [TCGA-PRAD] collection*. Cancer Imaging Archive; 2016.
26. Pryor D, Turner S, Tai K, et al. Moderate hypofractionation for prostate cancer: A user's guide. *J Med Imaging Radiat Oncol*. 2018; 62:232-239.
27. Bowes D, Robar J, Carson J, Wilke D, Lawrence E. Analysis of the accuracy of automated atlas-based segmentation for prostate radiotherapy. *Radiother Oncol*. 2009;92:S38.
28. Zhu J, Zhang J, Qiu B, Liu Y, Liu X, Chen L. Comparison of the automatic segmentation of multiple organs at risk in CT images of lung cancer between deep convolutional neural network-based and atlas-based techniques. *Acta Oncol*. 2019;58:257-264.
29. Frederick A, Roumeliotis M, Grendarova P, et al. A framework for clinical validation of automatic contour propagation: Standardizing geometric and dosimetric evaluation. *Pract Radiat Oncol*. 2019;9: 448-455.
30. Tsuji S, Hwang A, Weinberg V, Yom S, Quivey J, Xia P. Dosimetric evaluation of automatic segmentation for adaptive IMRT for head-and-neck cancer. *Int J Radiat Oncol Biol Phys*. 2010;77: 707-714.
31. Lim T, Gillespie E, Murphy J, Moore K. Clinically oriented contour evaluation using dosimetric indices generated from automated knowledge-based planning. *Int J Radiat Oncol Biol Phys*. 2019;103: 1251-1260.
32. Nourzadeh H, Watkins W, Ahmed M, Hui C, Schlesinger D, Siebers J. Clinical adequacy assessment of autocontours for prostate IMRT with meaningful endpoints. *Med Phys*. 2017;44:1525-1537.