Physics and Imaging in Radiation Oncology 28 (2023) 100501

Contents lists available at ScienceDirect

Physics and Imaging in Radiation Oncology

journal homepage: www.sciencedirect.com/journal/physics-and-imaging-in-radiation-oncology

Original Research Article

Real-world validation of Artificial Intelligence-based Computed Tomography auto-contouring for prostate cancer radiotherapy planning

Gabriele Palazzo [a], Paola Mangili [a], Chiara Deantoni [b], Andrei Fodor [b], Sara Broggi [a], Roberta Castriconi [a], Maria Giulia Ubeira Gabellini [a], Antonella del Vecchio [a], Nadia G. Di Muzio [b,c], Claudio Fiorino [a,*]

[a] Medical Physics, IRCCS San Raffaele Scientific Institute, Milano, Italy
[b] Radiotherapy, IRCCS San Raffaele Scientific Institute, Milano, Italy
[c] Vita-Salute San Raffaele University, Italy

A R T I C L E I N F O

Keywords:
Auto-contouring
Artificial Intelligence in radiotherapy
Prostate cancer
Radiotherapy
Planning

A B S T R A C T

Background and purpose: Artificial Intelligence (AI)-based auto-contouring for treatment planning in radiotherapy needs extensive clinical validation, including the impact of editing after automatic segmentation. The aims of this study were to assess the performance of a commercial system for Clinical Target Volumes (CTVs) (prostate/seminal vesicles) and selected Organs at Risk (OARs) (rectum/bladder/femoral heads + femurs), evaluating also inter-observer variability (manual vs automatic + editing) and the reduction of contouring time.

Materials and methods: Two expert observers contoured CTVs/OARs of 20 patients in our Treatment Planning System (TPS). Computed Tomography (CT) images were sent to the automatic contouring workstation: automatic contours were generated and sent back to TPS, where observers could edit them if necessary. Inter- and intra-observer consistency was estimated using Dice Similarity Coefficients (DSC). Radiation oncologists were also asked to score the quality of automatic contours, ranging from 1 (complete re-contouring) to 5 (no editing). Contouring times (manual vs automatic + edit) were compared.

Results: DSCs (manual vs automatic only) were consistent with inter-observer variability (between 0.65 for seminal vesicles and 0.94 for bladder); editing further improved performances (range: 0.76–0.94). The median clinical score was 4 (little editing) and it was <4 in 3/2 patients for the two observers respectively. Inter-observer variability of automatic + editing contours improved significantly, being lower than manual contouring (e.g.: seminal vesicles: 0.83vs0.73; prostate: 0.86vs0.83; rectum: 0.96vs0.81). Oncologist contouring time reduced from 17 to 24 min of manual contouring time to 3–7 min of editing time for the two observers ($p < 0.01$).

Conclusion: Automatic contouring with a commercial AI-based system followed by editing can replace manual contouring, resulting in significantly reduced time for segmentation and better consistency between operators.

1. Introduction

The manual delineation of target volumes (CTV) and organs at risk (OARs) remains a crucial phase of the radiotherapy chain for prostate cancer patients, and represents a substantial amount of working time for radiation oncologists. Manual contouring on Computed Tomography (CT) images is not only cumbersome, but is also affected by uncertainties that translate into clinically significant inter-observer variability, both in delineating CTV (i.e.: prostate and seminal vesicles) as well as OARs [1–5]. Automatic tools aiming to both reduce contouring time and improve contouring consistency have been developed over the past 10–15 years [6–15].

As is the case in other branches of Radiation Oncology, AI-based solutions seem to have great potential in supporting and partially replacing repetitive human activities through proper model training using large datasets, and may represent a major clinical challenge to professionals in terms of safety, accuracy and confidence [16–21]. In the last few years, the application of advanced AI-based methods for automatic segmentation based on deep learning (DL) in the prostate cancer scenario have been developed for both CT and Magnetic Resonance Imaging (MRI) [7]: it was also suggested that AI-based automatic contouring may have significant potential applications in the Quality

* Corresponding author.
E-mail address: fiorino.claudio@hsr.it (C. Fiorino).

https://doi.org/10.1016/j.phro.2023.100501
Received 6 June 2023; Received in revised form 9 October 2023; Accepted 10 October 2023
Available online 13 October 2023

Assurance (QA) of clinical trials [22], and in facilitating adaptive radiotherapy through fast (re)-contouring on CT/Cone Beam CT (CBCT), as well as on MRI [23].

Quite recently, several AI-based commercial systems for auto-contouring have begun to appear. Commercial software for AI-based automatic contouring should be validated in terms of its ability in contouring prostate cancer patients on CT images, including both CTVs (prostate and seminal vesicles) and OARs in a real-life context. On the one hand, some studies have shown promising results in terms of both substantial reduction of contouring time and improvement in quality of delineation in real-life clinical contexts [13,14,24,25]. On the other hand, the reported experiences are still very limited, especially when considering true clinical impact. Very few studies have clearly compared performances of AI-based auto-contouring against manual delineation [25], or against inter-institution variations [24]. Reasonable editing is expected especially for CTV, at least in a fraction of patients, as already shown in other scenarios [26].

The aims of our study were to assess the quality of performances of a commercial system, by comparing automatic against manual segmentation. We also quantified the need for additional manual editing, and its impact on the agreement between automatic (+edited) and manual contours, evaluating any improvement in consistency between different radiation oncologists due to automatic contouring. Moreover, we quantified the contouring time saved by automatic contouring, including the editing after automatic contour generation.

2. Materials and methods

2.1. Study design

The study was approved by the Ethical Committee of San Raffaele Institute, Milano (187/INT/2021 on Feb 3, 2022). Two radiation oncologists of our Institution with >10 years' experience in contouring prostate cancer patients for planning purposes were asked to manually contour planning CT images of 20 intermediate/high risk prostate cancer patients. All patients were randomly selected from among those treated with radical intent within the previous year, according to an Institutional protocol delivering 74.2 Gy in 28 fractions to prostate Planning Target Volume (PTV), using a Knowledge Based (KB) automatic plan optimization approach [27–29]. Planning CT images were acquired with a 0.3–0.5 cm step from below the external genitals up to the L2 vertebral body, and sent to our treatment planning station (Eclipse v 13.5 Varian) for contouring. The two observers, blinded from each other's delineations, contoured bladder, femoral heads + femurs, rectum, prostate and seminal vesicles, according to our Institutional policy: in particular, the rectum was contoured from the anus up to the point at which it turns into the sigmoid [2], and the femoral heads + femurs were contoured up to approximately the expected caudal limit of PTV. CTs were sent to MIM (v. 7.2.8), on which we set up a workflow for the generation of AI-based automated segmentation of OARs and CTVs using the MIM Protegè system, v. 1.1.2 (Prostate CT 2.0.0). As reported by internal documentation of the vendor, the system runs a U-Net model trained on a multi-institution dataset. Once generated, contours were sent back to Eclipse in order to allow radiation oncologists to visualize them in the more familiar environment. An independent copy of the structures was also generated to allow editing to be performed (if necessary).

Regarding femoral heads + femurs, the automatically generated contours extended up to the most caudal CT slice. Since only the upper part of the contour is relevant for treatment planning, and is usually contoured by the radiation oncologists (as previously described), automatic contours were cut *a posteriori* at the same level as the manual structures for each observer in order to render the contours comparable. The two observers were also asked to report the time spent for the manual contouring for each patient and that spent editing.

2.2. Assessing performances of automatic (with/without editing) contouring

The clinical validity of automatic segmentation was quantified for each observer by first comparing volume values, DSCs and Hausdorff Distances (HDs) [30] between automatically generated vs manual contours. Contour volumes and all other metrics were computed using SimpleITK Python library [31]. DICOM CT images were converted into nifti files (of size 512 × 512 × number of DICOM slices), maintaining the same CT voxel spacing (0.98 mm × 0.98 mm × 5 mm). DICOM RT Structure sets were converted into nifti files using the same voxel grid size and the same voxel spacing. RT Structure nifti masks were expanded into 0.5 mm × 0.5 mm × 5 mm voxels using linear resampling along the xy plane, in order to make segmentation masks smoother.

In addition, a clinical quality score of automatic contouring ranging from 1 to 5 was reported by each radiation oncologist according to the following scale: 1-Very poor quality (need for major correction, re-contouring); 2-Major editing necessary; 3-Moderate editing necessary; 4-Minor editing necessary; 5-Acceptable editing (no or negligible editing). DSC and HD between contours obtained by automatic generation followed by editing were then compared also against manual ones to better assess performance in the more realistic scenario of operator intervention after automatic segmentation. In all cases, average values were considered across all patients, using the standard deviation as error. The signed-rank Wilcoxon test was used to assess whether differences between groups were statistically significant. Intra-observer DSC and HD were compared by averaging values across all observers.

2.3. Inter-observer variability and time spared in a clinical setting

Inter-observer consistency was evaluated using DSC and HD for both manual contours and contours resulting after editing of the automatic ones, aiming to estimate possible improvements in inter-observer agreement due to the introduction of automatic segmentation followed by final editing by the radiation oncologist. This procedure is expected to be operatively implemented in the clinical activity.

In order to assess the contouring time, the average time spent for manual and automatic + edit contouring and their difference were considered across all patients. Mann-Whitney test was used to assess possible improvements of inter-observer variability and time spent for contouring between manual vs automatic + edit contouring.

Additionally, an automated workflow for automatic segmentation takes about 8 min per patient with our current hardware configuration (CPU: Intel® Xeon® Silver 4208 with 8 cores and 16 threads, RAM: 32 GB); this time was not computed as it does not involve any operator time as the workflow is completely automated.

3. Results

3.1. Performances of automatic (with/without editing) contouring

The results regarding the consistency between manual and automatic

Table 1
DSC for edited segmentation compared against manual contours for both observers (O1 and O2). DSC between manual and automatic segmentation is also shown in order to better appreciate the performance. DSC values were compared between different observers with the Wilcoxon test (p value shown in the table).

	Manual vs edited DSC			Manual vs automatic DSC		
	O1	O2	p value	O1	O2	p value
Bladder	0.95	0.94	0.12	0.94	0.94	0.99
Prostate	0.90	0.82	< 0.001	0.83	0.81	0.11
Seminal Vesicles	0.78	0.73	0.16	0.65	0.66	0.76
Rectum	0.85	0.83	0.13	0.82	0.82	0.99
Femoral Heads	0.94	0.73	< 0.001	0.93	0.71	< 0.001

(with/without editing) contours are shown in Table 1 and Fig. 1. The average values between the two observers of DSC regarding manual vs (fully) automatic contours showed values ranging between 0.65 for seminal vesicles and 0.94 for bladder with significant differences between the two observers for prostate (after editing) and for femoral heads + femurs (this last not clinically relevant). After editing, DSC significantly improved ($p \leq 0.002$) for both OARs and CTVs: the larger gain was found for seminal vesicles (0.76 vs 0.65) and for prostate (0.86 vs 0.82). The values near 1 for OARs and lower for CTVs suggested that editing was in general very limited or none for OARs and greater for CTVs. Results regarding HD are shown in the Supplementary Fig. S1, with values below 20 mm for most patients for Bladder, Prostate and Seminal Vesicles. Larger values were found for Rectum and Femoral Heads, due to the variability of the cranio-caudal limit of the structures chosen by different radiation oncologists in manual contours.

Considering the volumes, manual CTVs (Fig. 2) were statistically larger than unedited automatically generated volumes (Wilcoxon p value: Prostate: $p < 0.001$, Seminal Vesicles: $p = 0.001$). In contrast, there was no statistical difference between manual and automatic OAR volumes.

Regarding the clinical scores, as shown in the Supplementary material (Fig S3 and S4), they reflected generally high satisfaction. Median value was 4 (i.e.: little editing) for both operators, with only one patient (only for Obs2) showing a score of 2, and two patients (for both Obs1 and Obs2) with a score of 3. Despite the excellent performances, in a limited fraction of patients the segmentation was not fully satisfactory (2 and 3 out of 20 patients for the two observers, respectively), requiring more extensive editing for a few patients.

3.2. *Improvement of inter-observer variability*

As shown in Table 2, apart from femoral heads + femurs, inter-observer variability of manual contours in terms of DSC ranged between 0.73 (seminal vesicles) and 0.93 (bladder): the poorer value for femoral heads + femurs depended on the variable caudal limit chosen by the observers and has no clinical significance. After automatic contouring and manual editing, inter-observer variability improved significantly, particularly for femoral heads + femurs, rectum and seminal vesicles. As shown in Fig. 3, apart from femoral heads + femurs, inter-observer variability of manual contouring was quite similar to the (intra-observer) consistency between manual vs automatic contour (without editing). The DSC values between the two observers for all patients for prostate, seminal vesicles, bladder and rectum are shown in the Supplementary material. In Supplementary Fig. S2, inter-observer variability of HD are also shown, with values <20 mm for Bladder, Prostate and Seminal Vesicles; larger values were found for Rectum and Femoral Heads, showing a significant reduction from manual contours to automatic edited contours, due to the same cranio-caudal limit of the structures drawn by the software.

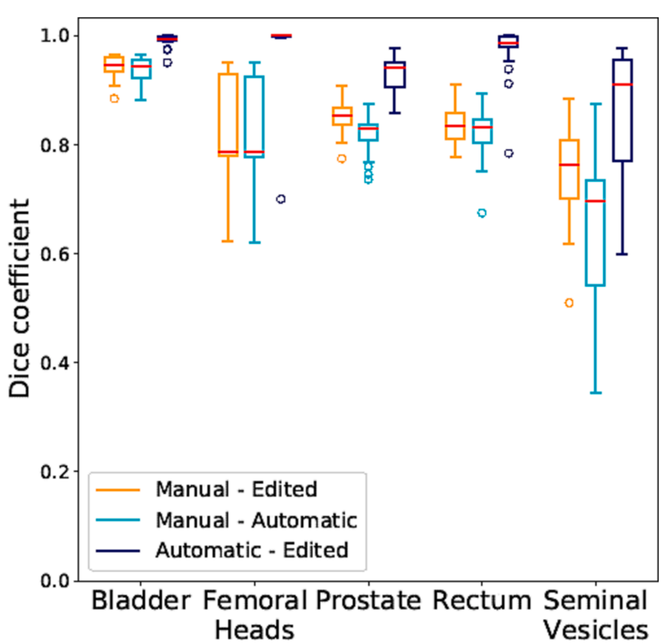

Fig. 1. Comparison of DSCs between manual and edited, between manual and automatic unedited and between edited and automatic unedited contours. Red lines show median values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3. *Time sparing*

As shown in the Supplementary material (Supplementary Fig. S5), the time for contouring (manual vs automatic + editing) was significantly reduced ($p < 0.01$) for both operators: for Obs 1 it was reduced from 24 (±3.5) to 7 min (±3) and for Obs 2 from 17 (±2.5) to 3 min (±1.5). In Fig. 4, the detailed results regarding editing times for the two observers for all patients were shown.

4. Discussion

In this work, we assessed the quality of a currently available commercial system for automated contouring, considering its impact on contour quality and on the spared time by radiation oncologists. Even though contours generated by the neural network were generally not fully accepted, little editing was required by radiation oncologists, thus causing a significant reduction in contouring time. Additionally, fast editing of the automatic contours translated into a significant reduction of inter-observer variability, both for OARs and prostate/seminal vesicles.

The current study was planned by medical physicists together with clinicians, with the above described aims and the final objective of adding an automatic workflow for contouring [29] to the currently used automatic plan workflow [29], rendering the entire treatment chain preparation for prostate cancer patients nearly fully automated.

For this reason, great attention was devoted to estimating the performance of our system in an operatively clinical scenario, including the possibility of editing the automatic contours before contouring approval, keeping the final decision to the doctors. Linked to this, the estimate of the extent of editing also in terms of additional time spent by the radiation oncologist, was quantified. Importantly, we verified a much lower time spent relative to manual contouring: the time saved for an expert radiation oncologist was typically about 15 min per patient.

Time-saving was not the only gain identified: the doctors reported that the clinical scores (based on the subjective degree of editing needed) was very good, with very few cases (3/20 and 2/20 for the two observers) of moderated/limited satisfactoriness. Agreement between automatically generated vs manual contours were also satisfactory, with DSC values comparable to inter-observer variability, apart from femoral heads + femurs, due to the specific (and clinically negligible) variability in assessing the caudal limit of this structure.

More importantly, after fast (typically of a few minutes) editing, performances showed significant further improvement, including a clear reduction of inter-observer variability when considering the final edited contours. In particular, the improvement in inter-observer variability was also significant for prostate and for seminal vesicles, with final DSC values for the two CTVs of 0.88 and 0.83 respectively, against values referred to manual contouring of, respectively, 0.83 and 0.73. The explanation of such results likely depend on the “supervised” guide of automatically generated contours toward more similar contouring between observers compared to manual delineation *ab initio*. In the case of OARs, where consistency is dependent also on the definition of the cranial/caudal limits, such as femoral heads + femurs and, in a minor

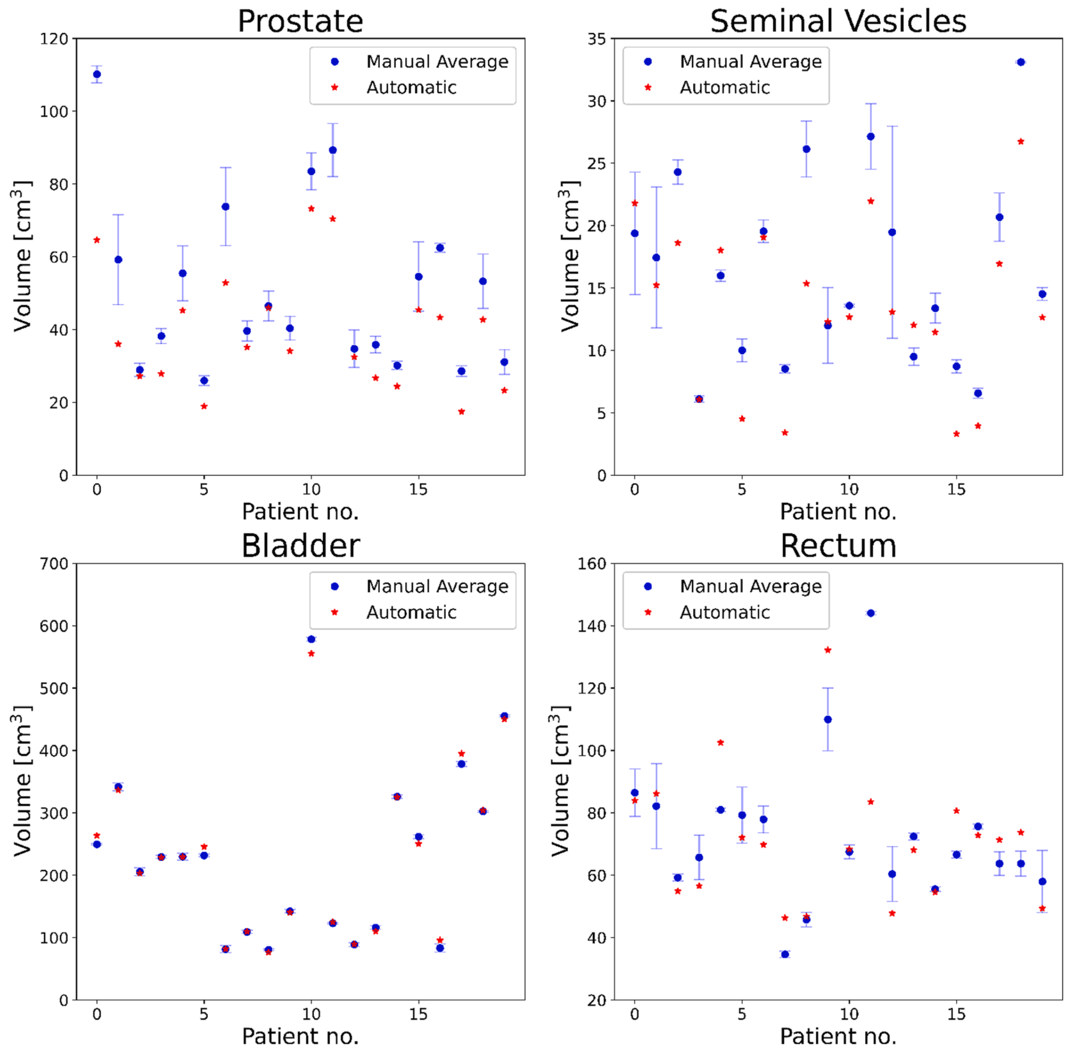

Fig. 2. Segmentation volumes comparison (manual vs unedited automatic). Average manual contour volumes are compared to unedited automatic volumes for all patients. Error bars show the standard deviation between the two observers.

Table 2
DSC between the manual segmentation and the edited automatic segmentation for different operators. *this result was due to the shift of the caudal limit to the same level as for the observer (i.e.: not due to manual editing).

Organ	Edited inter-observer DSC	Manual inter-observer DSC	p value
Bladder	0.98	0.93	< 0.001
Prostate	0.88	0.83	0.009
Seminal Vesicles	0.83	0.73	0.001
Rectum	0.96	0.81	< 0.001
Femoral Heads	0.97*	0.65	< 0.001

extent, rectum, this phenomenon is still stronger, as shown in Table 2.

Even though DSC does not correlate well with editing time [32], it has been always used in similar studies, therefore allowing for a broader comparison. However, we also considered another metric (HD) which is more sensitive to slight changes in contours: apart from femoral heads + femurs, HD results were found to be reasonably low, and consistent with few other recent studies [33,34].

Of note, HD values for manual contours of femoral heads were rather large due to the different (arbitrary) caudal limit of the structures, which has not clinical relevance.

Similarly, HD values for inter-observer manual rectum contours were also rather large due to the different cranial limit chosen by different radiation oncologists.

More generally, when trying to compare our results with similar recent studies dealing with the validation of AI-based auto-contouring commercial systems in the prostate cancer scenario (as summarized in the Supplementary material (Table S1)), our results are generally in quite good agreement with the literature. The findings that can be better compared are those referring to the DSC values between automatic (unedited) vs manual contours: apart from femoral heads + femurs, the results are comparable with or slightly better than the findings reported by using both MIM [13,34] and other systems [14,24,25,34,35]. In particular, results in segmenting rectum and bladder were better than those reported by Walker et al [24]; results referring to prostate and seminal vesicles were better than those reported by Duan et al [34].The estimates of time reduction were also similar [25,33,35] or better [14,24] than in other reported studies. Unfortunately, the results regarding the impact of editing are difficult to compare, as, to our knowledge, only one very recent study has clearly reported results concerning contours edited by doctors after automatic segmentation. In this study, Doolan et al [33] validated five commercial systems (different from the one used in this study) reporting a range of DSC values very similar to those reported by our study: for instance, DSC values for prostate ranged between 0.85 and 0.91 (against our value of 0.86). Interestingly, the same study reported average editing times

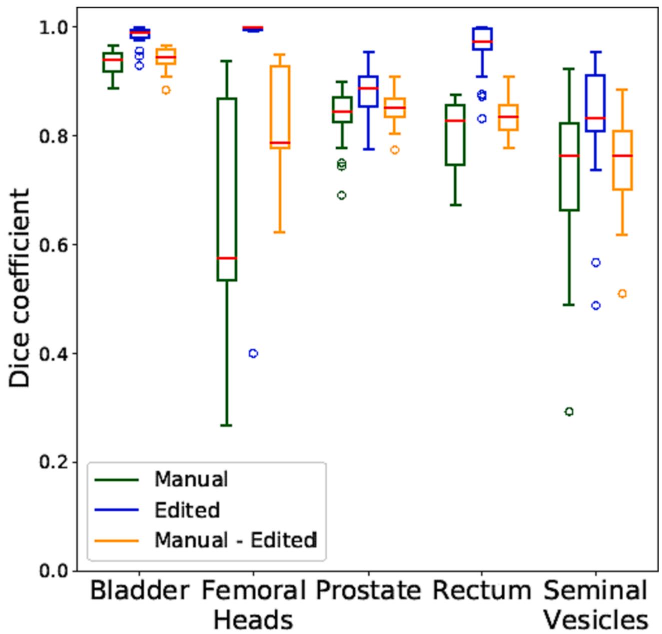

Fig. 3. Inter-observer DSC comparison. Comparison of average inter-observer DSCs on manual contours, inter-observer DSCs on edited contours and intra-observer DSC differences between manual and edited contours. Red lines show median values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

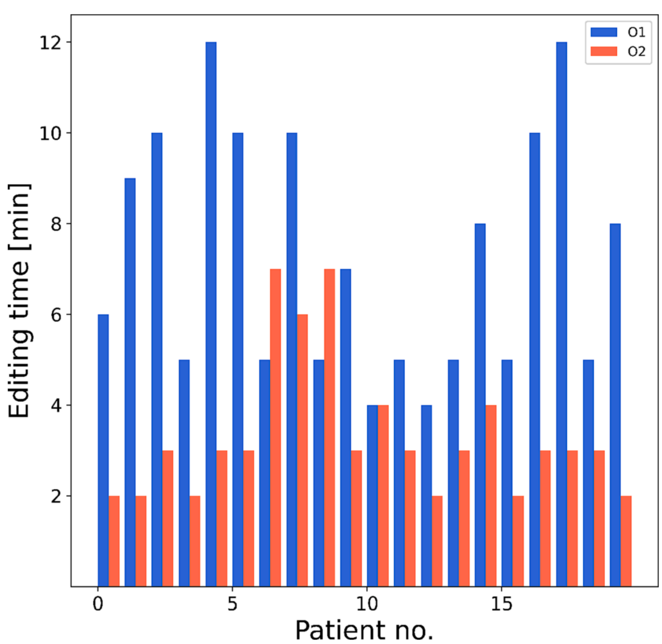

Fig. 4. Editing time across all patients. The plot shows editing times of the two observers for all patients.

(0.3–7.4 min) consistent with ours (3–7 min).

Another important result of our study is the quantification of the reduction of inter-observer variability due to the use of AI-based auto-contouring followed by editing, compared to manual segmentation. The reduction of inter-observer variability after editing of automatic contours was not reported by previous studies, although it may be reasonably expected and has been reported for other districts (see for instance [24]).

Our study has some limitations: first, the number of observers was limited to two, and current findings could benefit from further confirmation by the addition of new observers. Another limitation concerns the monocentric characteristic of the study: inter-institute contouring variability may be an issue, and the quantitative results reported here cannot therefore be automatically translatable to other Institutes. Our results would be clearly reinforced by replications in other Institutions.

At the moment, a fully automated workflow has been implemented and clinically activated: once the patient CT is imported into the TPS station, it may be sent to the auto-contouring workstation where a proper scripting automatically starts the segmentation and sends the automatic generated contours, which are easily imported, back to the TPS. The entire procedure requires less than 8 min, while operator time is less than 2 min. In the future, further efforts should be made to extend this approach to the case of pelvic node irradiation, which will require proper training for pelvic node CTVs, and for bowel loops.

In conclusion, the current study demonstrated the benefit of implementing AI-based automatic segmentation for prostate cancer radiotherapy in terms of both time-saving and improvement in contouring quality, including the reduction of inter-observer variability.

CRediT authorship contribution statement

Gabriele Palazzo: Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft. **Paola Mangili:** Conceptualization, Methodology. **Chiara Deantoni:** Methodology, Investigation, Writing – review & editing. **Andrei Fodor:** Methodology, Investigation, Writing – review & editing. **Sara Broggi:** Writing – review & editing. **Roberta Castriconi:** . **Maria Giulia Ubeira-Gabellini:** Software, Methodology, Writing – review & editing. **Antonella del Vecchio:** Writing – review & editing. **Nadia G. Di Muzio:** Resources, Writing – review & editing. **Claudio Fiorino:** Conceptualization, Funding acquisition, Writing – original draft, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The study was supported by AIRC (Associazione Italiana per la Ricerca sul Cancro), IG25951 and IG23150 grants. D Raspanti of Tema Sinergie is gratefully acknowledged for support in the use of MIM and MIMassistant softwares. Frank Bagg is acknowledged for the linguistic revision.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.phro.2023.100501.

References

[1] Fiorino C, Reni M, Bolognesi A, Cattaneo GM, Calandrino R. Intra- and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. Radiother Oncol 1998;47:285–92. https://doi.org/10.1016/s0167-8140(98)00021-8.

[2] Foppiano F, Fiorino C, Frezza G, Greco C, Valdagni R. The impact of contouring uncertainty on rectal 3D dose-volume data: results of a dummy run in a multicenter trial (AIROPROS01-02). Int J Radiat Oncol Biol Phys 2003;57:573–9. https://doi.org/10.1016/s0360-3016(03)00659-x.

[3] Carillo V, Cozzarini C, Perna L, Calandra M, Gianolini S, Rancati T, et al. Contouring variability of the penile bulb on CT images: quantitative assessment using a generalized concordance index. Int J Radiat Oncol Biol Phys 2012;84:841–6. https://doi.org/10.1016/j.ijrobp.2011.12.057.

[4] Roach D, Holloway LC, Jameson MG, Dowling JA, Kennedy A, Greer PB, et al. Multi-observer contouring of male pelvic anatomy: highly variable agreement

across conventional and emerging structures of interest. J Med Imaging Radiat Oncol 2019;63:264–71. https://doi.org/10.1111/1754-9485.12844.
[5] Guzen L, Beddok A, Nioche C, Modzelewski R, Loiseau C, Sailleron J, et al. Assessing interobserver variability in the delineation of structures in Radiation Oncology: a systematic review. Int J Radiat Oncol Biol Phys 2023;115:1047–60. https://doi.org/10.1016/j.ijrobp.2022.11.021.
[6] Huyskens DP, Maingon P, Vanuytsel L, Remouchamps V, Roques T, Dubray B, et al. A qualitative and a quantitative analysis of an auto-segmentation module for prostate cancer. Radiother Oncol 2009;90:337–45. https://doi.org/10.1016/j.radonc.2008.08.007.
[7] Kalantar R, Lin G, Winfield JM, Messiou C, Lalondrelle S, Blackledge MD, et al. Automatic segmentation of pelvic cancers using deep learning: State-of-the-art approaches and challenges. Diagnostics 2021;11:1964. https://doi.org/10.3390/diagnostics11111964.
[8] La Macchia M, Fellin F, Amichetti M, Cianchetti M, Gianolini S, Paola V, 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. https://doi.org/10.1186/1748-717x-7-160.
[9] Kearney VP, Chan JW, Wang T, Perry A, Yom SS, Solberg. Attention-enabled 3D boosted convolutional neural networks for semantic CT segmentation using deep supervision. Phys Med Biol 2019;64:135001. https://doi.org/10.1088/1361-6560/ab2818.
[10] Balagopal A, Kazemifar S, Nguyen D, Lin M-H, Hannan R, Owrangi AM, et al. Fully automated organ segmentation in male pelvic CT images. Phys Med Biol 2018;63:245015. https://doi.org/10.1088/1361-6560/aaf11c.
[11] Greenham S, Dean J, Fu CKK, Goman J, Milligan J, Deanna T, et al. Evaluation of atlas-based auto-segmentation software in prostate cancer patients. J Med Radiat Sci 2014;61:151–8. https://doi.org/10.1002/jmrs.64.
[12] Olsson CE, Suresh R, Niemela J, Akram SU, Valdman A. Autosegmentation based on different sized training datasets of consistently curated volumes and impact on rectal contours in prostate cancer radiation therapy. Phys Imaging Radiat Oncol 2022;22:67–72. https://doi.org/10.1016/j.phro.2023.100416.
[13] Urago Y, Okamoto H, Kaneda T, Murakami N, Kashihara T, Takemori M, et al. Evaluation of auto-segmentation accuracy of cloud-based artificial intelligence and atlas-based models. Radiat Oncol 2021;16:175. https://doi.org/10.1186/s13014-021-01896-1.
[14] Zabel WJ, Conway JL, Gladwish A, Skliarenko J, Didiodato G, Goorts-Matthews L, et al. Clinical evaluation of deep learning and atlas-based auto-contouring of bladder and rectum for prostate radiation therapy. Pract Radiat Oncol 2021;11:e80–9. https://doi.org/10.1016/j.prro.2020.05.013.
[15] Aoyama T, Shimizu H, Kitagawa T, Yokoi K, Koide Y, Tachibana H, et al. Comparison of atlas-based auto-segmentation accuracy for radiotherapy in prostate cancer. Phys Imaging Radiat Oncol 2021;19:126–30. https://doi.org/10.1016/j.phro.2021.08.002.
[16] Boon IS, Au Yong TPT, Boon CS. Assessing the role of artificial intelligence (AI) in clinical oncology: utility of machine learning in radiotherapy target volume delineation. Medicines (Basel) 2018;5:E131. https://doi.org/10.3390/medicines5040131.
[17] Unkelbach J, Bortfeld T, Cardenas CE, Gregoire V, Hager W, Heijmen B, et al. The role of computational methods for automating and improving clinical target volume definition. Radiother Oncol 2020;153:15–25. https://doi.org/10.1016/j.radonc.2020.10.002.
[18] Fiorino C, Guckenberger M, Schwarz M, van der Heide UA, Heijmen B. Technology-driven research for radiotherapy innovation. Mol Oncol 2020;14(1500–1513). https://doi.org/10.1002/1878-0261.12659.
[19] Thompson RF, Valdes G, Fuller CD, Carpenter CM, Morin O, Aneja S, et al. Artificial intelligence in radiation oncology: a specialty-wide disruptive transformation? Radiother Oncol 2018;129:421–6. https://doi.org/10.1016/j.radonc.2018.05.030.
[20] Fiorino C, Jeraj R, Clark CH, Garibaldi C, Georg D, Muren L, et al. Grand challenges for medical physics in radiation oncology. Radiother Oncol 2020;153:7–14. https://doi.org/10.1016/j.radonc.2020.10.001.
[21] Xing L, Krupinski EA, Cai J. Artificial intelligence will soon change the landscape of medical physics research and practice. Med Phys 2018;45:1791–3. https://doi.org/10.1002/mp.12831.
[22] Thor M, Apte A, Haq R, Iyer A, LoCastro E, Deasy JO. Using auto-segmentation to reduce contouring and dose inconsistency in clinical trials: the simulated impact on RTOG 0617. Int J Radiat Oncol Biol Phys 2021;109:1619–26. https://doi.org/10.1016/j.ijrobp.2020.11.011.
[23] Byrne M, Archibald-Heeren B, Hu Y, Teh A, Beserminji R, Cai E, et al. Varian ethos online adaptive radiotherapy for prostate cancer: early results of contouring accuracy, treatment plan quality, and treatment time. J Appl Clin Med Phys 2022;23. https://doi.org/10.1002/acm2.1347.
[24] Walker Z, Bartley G, Hague C, Kelly D, Navarro C, Rogers J, et al. Evaluating the effectiveness of deep learning contouring across multiple radiotherapy centres. Phys Imaging Radiat Oncol 2022;24(121–128). https://doi.org/10.1016/j.phro.2022.11.003.
[25] Wong J, Huang V, Wells D, Giambattista J, Giambattista J, Kolbeck C, et al. Implementation of deep learning-based auto-segmentation for radiotherapy planning structures: a workflow study at two cancer centers. Radiat Oncol 2021;16:101. https://doi.org/10.1186/s13014-021-01831-4.
[26] Vaassen F, Boukerroui D, Looney P, Canters R, Verhoeven K, Peeters S, et al. Real-world analysis of manual editing of deep learning contouring in the thorax region. Phys Imaging Radiat Oncol 2022;22:104–10. https://doi.org/10.1016/j.phro.2022.04.008.
[27] Di Muzio N, Fiorino C, Cozzarini C, Alongi F, Broggi S, Mangili P, et al. Phase I-II study of hypofractionated simultaneous integrated boost with tomotherapy for prostate cancer. Int J Radiat Oncol Biol Phys 2009;74:392–8. https://doi.org/10.1016/j.ijrobp.2008.08.038.
[28] Di Muzio N, Deantoni CL, Brombin C, Fiorino C, Cozzarini C, Zerbetto F, et al. Ten year results of extensive nodal radiotherapy and moderately hypofractionated simultaneous integrated boost in unfavorable intermediate-, high-, and very high-risk prostate cancer. Cancers (Basel) 2021;13(19):4970. https://doi.org/10.3390/cancers13194970.
[29] Castriconi R, Cattaneo GM, Mangili P, Esposito P, Broggi S, Cozzarini C, et al. Clinical implementation of knowledge-based automatic plan optimization for helical tomotherapy. Pract Radiat Oncol 2021;11:e236–44. https://doi.org/10.1016/j.prro.2020.09.012.
[30] Dice LR. Measures of the amount of ecologic association between species. Ecology 1945;26:297e302. https://doi.org/10.2307/1932409.
[31] Lowekamp BC, Chen DT, Ibanez L, Blezek D. The design of simpleitk. Front Neuroinf 2013;7:45. https://doi.org/10.3389/fninf.2013.00045.
[32] Gooding MJ, Smith AJ, Tariq M, Aljabar P, Peressutti D, van der Stoep J, et al. Comparative evaluation of autocontouring in clinical practice: a practical method using the Turing test. Med Phys 2018;45:5105–15. https://doi.org/10.1002/mp.13200.
[33] Doolan PJ, Charalambous S, Roussakis Y, Leczynski A, Peratikou M, Benjamin M, et al. A clinical evaluation of the performance of five commercial artificial intelligence contouring systems for radiotherapy. Front Oncol. 2023;13:1213068. https://doi.org/10.3389/fonc.2023.1213068.
[34] Duan J, Vargas CE, Yu NY, Laughlin BS, Toesca DS, Keole S, et al. Incremental retraining, clinical implementation, and acceptance rate of deep learning auto-segmentation for male pelvis in a multiuser environment. Med Phys 2023;50:4079–91. https://doi.org/10.1002/mp.16537.
[35] Radici L, Ferrario S, Borca VC, Cante D, Paolini M, Piva C, et al. Implementation of a commercial deep learning-based auto segmentation software in radiotherapy: evaluation of effectiveness and impact on workflow. Life (Basel) 2022;12:2088. https://doi.org/10.3390/life12122088.