Technical Note

# Comparison of atlas-based auto-segmentation accuracy for radiotherapy in prostate cancer

Takahiro Aoyama [a,b,*], Hidetoshi Shimizu [a], Tomoki Kitagawa [a], Kazushi Yokoi [a], Yutaro Koide [a], Hiroyuki Tachibana [a], Kojiro Suzuki [c], Takeshi Kodaira [a]

[a] *Department of Radiation Oncology, Aichi Cancer Centre, 1-1 Kanokoden, Chikusa-Ku, Nagoya, Aichi 464-8681, Japan*
[b] *Graduate School of Medicine, Aichi Medical University, 1-1 Yazako-karimata, Nagakute, Aichi 480-1195, Japan*
[c] *Department of Radiology, Aichi Medical University, 1-1 Yazako-karimata, Nagakute, Aichi 480-1195, Japan*

## ARTICLE INFO

## ABSTRACT

Atlas-based auto-segmentation (ABS) procedure used in radiotherapy can be classified into two groups, one using one atlas per patient (sSM) and the other using multiple atlases (sMM). This study evaluated auto-contouring accuracy and contouring time in patients with prostate cancer using the two procedures. The Dice similarity coefficient of sMM was significantly better than that of sSM (prostate [median, 0.81 (range, 0.66–0.91) vs. 0.64 (0.27–0.71), $p < 0.01$], seminal vesicles [0.49 (0.31–0.80) vs. 0.18 (0.01–0.60), $p < 0.05$], and rectum [0.81 (0.37–0.91) vs. 0.57 (0.31–0.77), $p < 0.01$]). The median contouring times were 2.6 (sMM) and 1.3 min (sSM).

## 1. Introduction

Intensity-modulated radiation therapy (IMRT) is a technique that specifically targets tumors while sparing the normal tissue, thereby leading to improved treatment outcomes for patients with prostate cancer [1,2]. Consequently, several clinics have widely used IMRT for such patients. However, considerable time is required to contour the relevant structures to establish an IMRT plan [3].

The atlas-based auto-segmentation (ABS) software is used to automatically contour target tumors and normal tissues on the computed tomography (CT) images of a new patient using predefined atlases and a non-rigid registration technique [4]. Several studies have used this software to reduce interobserver contouring variation and time [5–11]. Other studies have focused on improving the accuracy of non-rigid registration to enhance the accuracy of auto-contouring [12–15]. Some reports in recent years have focused on improving ABS procedures based on neuroimaging research [16,17], and these improved procedures have been applied to commercial treatment planning systems (TPSs) [8]. However, because structures created using the ABS software for treatment planning are far from satisfactory, they cannot be used for treatment planning without manual editing performed by a radiation

oncologist or a planner [8–10,18].

The ABS procedure can be classified into two groups, one using one atlas per patient and the other using multiple atlases per patient. Studies have compared the accuracy of ABS using these two procedures [19] and reported that the latter has a better auto-contouring accuracy [4]. Recently, the hybrid non-rigid registration algorithm, which combines image information by intensity with the structure provided by the contours, has been widely used [12] and reported to perform well compared with conventional algorithms [8].

Because the accuracy of ABS varies with a change in the procedure and algorithm used, verifying the accuracy for each procedure or algorithm is essential. Delpon et al. [10] reported the accuracy of ABS for prostate cancer using the hybrid non-rigid registration algorithm; however, the effect of ABS procedures was not clarified. To the best of our knowledge, the superiority of the procedure of multiple atlases over one atlas has not been observed in the hybrid non-rigid registration algorithm. Therefore, this study aimed to compare the accuracy of ABS for prostate cancer between the two procedures using the hybrid non-rigid registration algorithm.

## 2. Materials and methods

### 2.1. ABS procedure

RayStation version 6.2.0.7 (RaySearch Laboratories, Stockholm, Sweden), as a default procedure, implements the selected single atlas from multiple atlases (sSM) procedure (i.e., one atlas per patient). The sSM procedure can be divided into four steps: (1) searching for an atlas similar to the patient's CT images by similarity measurement, (2) selecting the atlas most similar to the patient's CT images from pre-defined atlases using the correlation coefficient, (3) performing non-rigid registration using ANAtomically CONstrained Deformation Algorithm (ANACONDA) as a transformation model, and (4) contouring the patient's CT images [12]. The present study investigated another procedure (i.e., multiple atlases per patient) by changing from predefined atlases to structures. The process of the procedure can be divided into four steps: (1) making multiple structures per structure, (2) searching and selecting each similar structure using the correlation coefficient in the RayStation functionality, (3) performing non-rigid registration using ANACONDA, and (4) contouring with each structure. Hence, this procedure was named as selected multiple structures from the multiple structures (sMM) procedure.

### 2.2. Patients and manual contouring

We retrospectively reviewed the volumetric CT datasets of 30 patients with prostate cancer who received IMRT at our center between April 2015 and April 2017. The following seven anatomical structures of the prostate region were contoured on the patient's CT images: the prostate, seminal vesicles, rectum, bladder, pubis, ischium, and femoral head. The Aquilion LB CT system (Canon Medical Systems, Tochigi, Japan) was utilized under the following conditions: tube voltage, 120 kV; tube current as determined by auto-exposure control (standard deviation = 10); slice width, 2.0 mm; matrix size, $512 \times 512$; and field of view, 550 cm. All structures of the 30 patients were manually contoured using the delineation consensus guideline for patients with prostate cancer as a reference [20]. The rectum was defined only in the 10-mm area of the craniocaudal direction of the planning target volume with the policy of the center of the present study [21]. The study was approved by the local ethics committee (no. 2018–1-415).

### 2.3. Creating predefined atlases and structures

The structure sets of 20 patients that were randomly selected from the 30 patients were used for predefined atlases [22] and structures using the ABS software of RayStation to create two different predefined atlases or structures. The two different ABS procedures (sSM and sMM) were then performed for the remaining 10 patients.

### 2.4. Evaluation of auto-contouring accuracy and contouring time

To evaluate auto-contouring accuracy, similarity between ABS and manual contouring was assessed using the Dice similarity coefficient (DSC) calculated as follows (Eq. (1)):

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \tag{1}$$

where $A$ and $B$ are the two structures evaluated. DSC may take any values between 0 (no overlap) and 1 (complete overlap). $A$ is the manually contoured structure and $B$ is the structure that was auto-contoured using the sSM or sMM procedure. Following previous studies [23,24], DSC values $> 0.7$, 0.7–0.4, and $<0.4$ were defined as substantial agreement, moderate agreement, and large variation, respectively. The Hausdorff distance computation utilizes the maximum–minimum function as defined by Eq. (2):

$$h(a,b) = \max_{a \in A} \{ \min_{b \in B} \{ d(a,b) \} \} \tag{2}$$

where $a$ and $b$ are the points of structures $A$ and $B$ and $d(a,b)$ is the Euclidian distance between $a$ and $b$.

The time required to perform ABS was measured in each procedure. Auto-contouring on RayStation was performed using a desktop computer with Windows 7 Professional with an Intel Xeon CPU at 3.2 GHz and 64 GB RAM.

### 2.5. Statistical analysis

The Wilcoxon signed-rank test was used to compare differences in DSC values and Hausdorff distance between the sSM and sMM procedures. Two-sided $p$ values $< 0.05$ were considered significantly different. All statistical analyses were performed using EZR version 1.3.6 (Saitama Medical Centre, Jichi Medical University, Saitama, Japan), a graphical user interface for R (The R Foundation for Statistical Computing, Vienna, Austria) [25].

## 3. Results

### 3.1. Auto-contouring accuracy

In the sSM procedure, DSC values indicated substantial agreement ($>0.7$) for the pubis, ischium, and femoral head; moderate agreement (0.7–0.5) for the prostate, rectum, and bladder; and large variation ($<0.5$) for the seminal vesicles. By contrast, in the sMM procedure, DSC values indicated substantial agreement ($>0.7$) for the prostate, rectum, bladder, pubis, ischium, and femoral head and moderate agreement (0.7–0.5) for the seminal vesicles. Significant differences in DSC values between the two procedures were noted in the prostate (median, 0.64 (range, 0.27–0.71) vs. 0.81 (0.66–0.91), $p < 0.01$), seminal vesicles (0.18 (0.01–0.60) vs. 0.49 (0.31–0.80), $p < 0.05$), and rectum (0.57 (0.31–0.77) vs. 0.81 (0.37–0.91), $p < 0.01$) (Fig. 1a).

The qualitative examples of the segmentation results are shown in Fig. 2. In particular, the contouring accuracy of sMM was improved compared with that of sSM with respect to the prostate, seminal vesicles, and rectum. Likewise, the result of the Hausdorff distances showed a significant difference among the prostate ($p > 0.05$), seminal vesicles ($p > 0.05$), and rectum ($p > 0.05$), but no significant difference was found in other contours (Fig. 1b).

The median time required to contour the seven structures in the sSM and sMM procedures was 1.3 (range, 1.2–1.4) and 2.6 (2.5–2.8) min, respectively.

## 4. Discussion

The auto-contouring accuracy of the sMM procedure using the hybrid non-rigid registration algorithm was found to be better than that of the sSM procedure. The median DSC values of the sSM procedure were nearly equivalent to those found in a previous study [10]. However, accuracy with regard to the prostate, seminal vesicles, and rectum is insufficient for clinical use and necessitates improvement. The sSM procedure has the disadvantage of selecting only one atlas that is similar to the patient. Because it is practically not possible to register and choose an atlas that roughly matches the patient with respect to all regions, the auto-contouring accuracy becomes low. In the sMM procedure, the auto-contouring accuracy was improved because the most similar structure could be selected from multiple structures.

In the bladder, by contrast, no significant difference was observed between the sSM and sMM procedures. Raudaschl et al. reported that the auto-contouring accuracy of large organs is better than that of small organs in ABS [11]. Each patient in the present study was instructed not to empty their bladder 3 h prior to treatment. The median bladder volume of all patients was 175 (range, 91–298) mL. Because the bladder
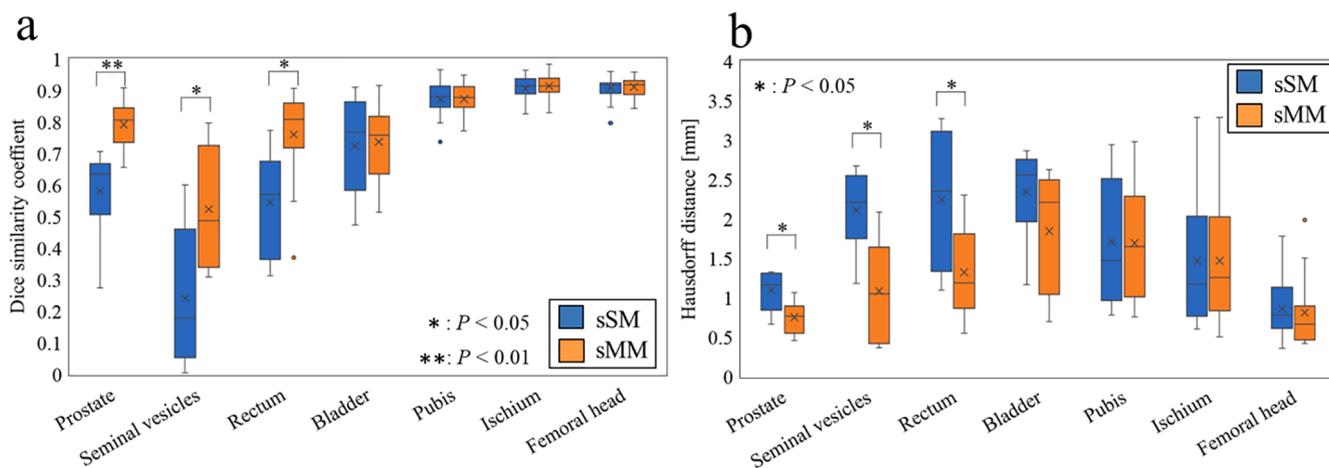
**Fig. 1.** Boxplots of the (a) Dice similarity coefficient and (b) Hausdorff distance between the manual contours and the two procedures. The points represent the maximum, upper quartile, mean, median, lower quartile, and minimum values. sSM, selected single atlas from multiple atlases; sMM, selected multiple structures from the multiple structures.
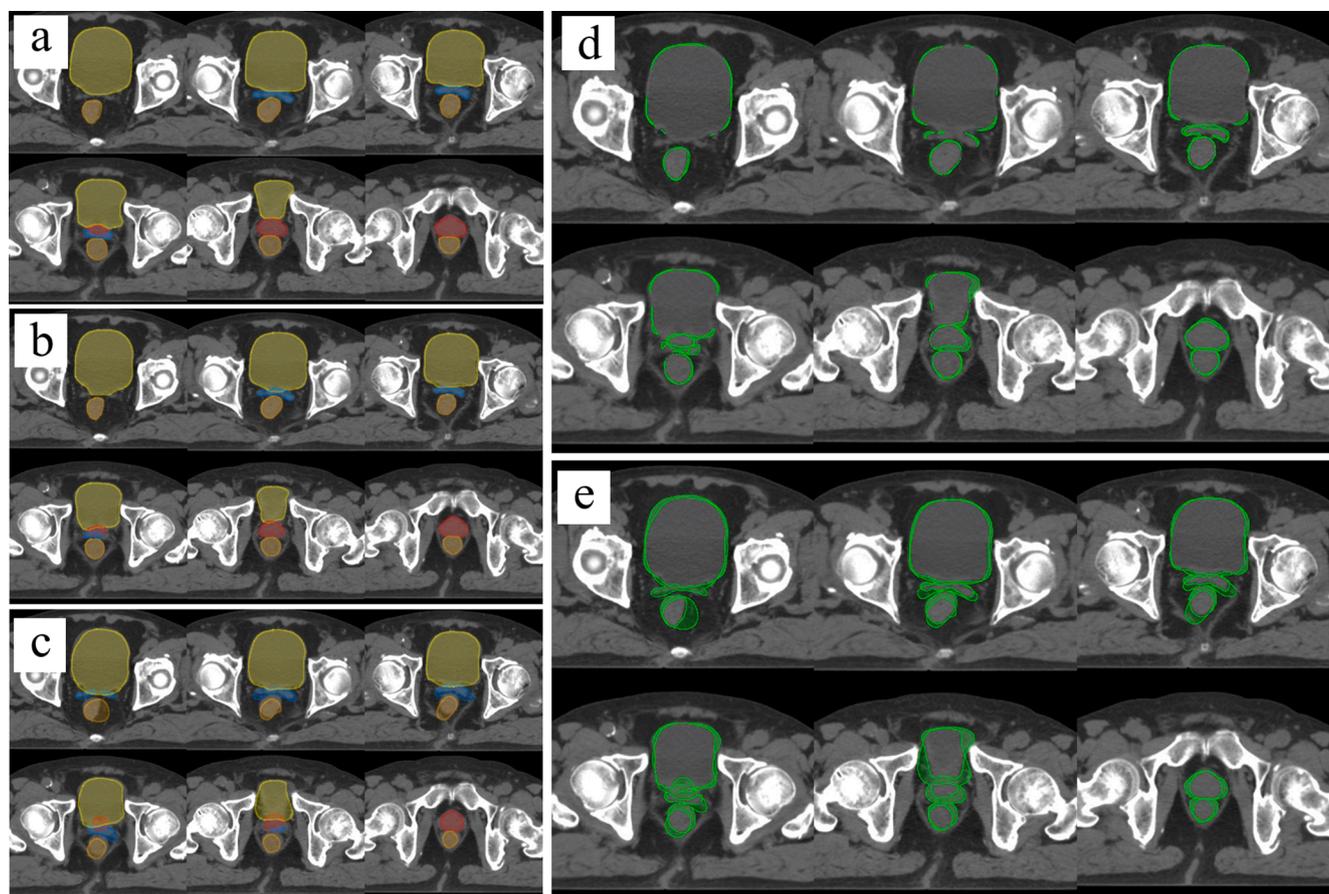


**Fig. 2.** Segmentation results for (a) manual contouring, (b) sMM procedure, and (c) sSM procedure; differences between (a–b) and (a–c) are shown in (d) and (e). Red line, the prostate; blue line, the seminal vesicles; orange line, the rectum; and yellow line, the bladder. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

volume was large, auto-contouring accuracy was good in both the sSM and sMM procedures but without significant differences. Moreover, in the pubis, ischium, and femoral head, no significant differences were noted. Because of the high contrast of the bony structure with the neighboring areas and low degree of shape variation, DSC values were good in both procedures but without significant differences [10]. DSC is sensitive to both over- and undercontouring and has the major

advantage of having the ability to evaluate both false positives and false negatives between specific voxels. However, its disadvantage is that the DSC value of a large structure volume is higher than that of a small structure volume [4]. In particular, bigger volumes that only have a small disagreement will still have a larger DSC value. Smaller volumes with a slight disagreement, on the other hand, will have a larger decrease in the DSC value. Hence, the Hausdorff distance was used as a

non-DSC indicator. However, no significant difference was noted in the Hausdorff distance (Fig. 1b).

Although the number of non-rigid registrations in the sMM procedure increased more than that in the sSM procedure, the calculation load was not extremely significant. In fact, the contouring time of the sMM procedure was approximately 1 min longer than that of the sSM procedure, but this increase is clinically acceptable.

Delays in radiotherapy initiation should be as short as reasonably achievable because delays are associated with an increased local recurrence rate [26]. However, the manual contouring of a large number of structures is time-consuming, burdening radiation oncologists and planners. Hence, auto-contouring technology is considered important in clinical practice. The mainstream of current research on auto-contouring is deep learning, showing an improved accuracy over ABS [27]. However, deep learning requires a large-scale dataset that is suitable for the target site. Moreover, commercial TPS does not generate datasets for all sites. By contrast, good results can be obtained with only a few dozen datasets with ABS, as shown in this study. In addition, the accuracy of deep learning can be improved by continuous learning on a cloud basis. However, cloud-based operation is not possible in certain countries because of legal regulations. Thus, it is believed that ABS may be an option in cases where deep learning datasets are not generated or when cloud management is difficult.

The present study has two major limitations. First, the number of predefined atlases was set to 20 in this study, as per Larrue et al.'s report [22]. However, the optimal number of atlases remains an open problem and depends on certain variables, including the target anatomy and image modality, among others. A detailed data analysis of other sites (e. g., head and neck) is necessary for future studies to acquire more useful results. Second, this study was conducted in a single facility that used a single commercial TPS. Because the similarity measurement and non-rigid registration algorithm adopted for each TPS are different, a multicenter study is needed to compare the robustness and effectiveness of the sMM procedure.

In conclusion, we evaluated auto-contouring accuracy and contouring time for prostate cancer using two different ABS procedures using the hybrid non-rigid registration algorithm. The sMM procedure demonstrated more enhanced auto-contouring accuracy compared with the sSM procedure. Furthermore, although the contouring time of the sMM procedure was slightly increased compared with that of the sSM procedure, this increase is clinically acceptable

## 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.

## References

[1] Zelefsky MJ, Fuks Z, Hunt M, Lee HJ, Lombardi D, Ling CC, et al. High dose radiation delivered by intensity modulated conformal radiotherapy improves the outcome of localized prostate cancer. J Urol 2001;166:876–81. https://doi.org/10.1097/00005392-200109000-00017.

[2] Zelefsky MJ, Fuks Z, Hunt M, Yamada Y, Marion C, Ling CC, et al. High-dose intensity modulated radiation therapy for prostate cancer: early toxicity and biochemical outcome in 772 patients. Int J Radiat Oncol Biol Phys 2002;53: 1111–6. https://doi.org/10.1016/s0360-3016(02)02857-2.

[3] Tomita N, Kodaira T, Teshima T, Ogawa K, Kumazaki Y, Yamauchi C, et al. Japanese structure survey of high-precision radiotherapy in 2012 based on institutional questionnaire about the patterns of care. Jpn J Clin Oncol 2014;44: 579–86. https://doi.org/10.1093/jjco/hyu041.

[4] Rohlfing T, Brandt R, Menzel R, Russakoff DB, Maurer CR. Quo vadis, atlas-based segmentation? In: Suri J, Wilson DL, Laxminarayan S, editors. Handbook of biomedical image analysis. Boston: Springer; 2005. p. 435–86.

[5] Nelms BE, Tomé WA, Robinson G, Wheeler J. Variations in the contouring of organs at risk: test case from a patient with oropharyngeal cancer. Int J Radiat Oncol Biol Phys 2012;82:368–78. https://doi.org/10.1016/j.ijrobp.2010.10.019.

[6] Tao CJ, Yi JL, Chen NY, Ren W, Cheng J, Tung S, et al. Multi-subject atlas-based auto-segmentation reduces interobserver variation and improves dosimetric parameter consistency for organs at risk in nasopharyngeal carcinoma: a multi-institution clinical study. Radiother Oncol 2015;115:407–11. https://doi.org/10.1016/j.radonc.2015.05.012.

[7] Chao KS, Bhide S, Chen H, Asper J, Bush S, Franklin G, et al. Reduce in variation and improve efficiency of target volume delineation by a computer-assisted system using a deformable image registration approach. Int J Radiat Oncol Biol Phys 2007; 68:1512–21. https://doi.org/10.1016/j.ijrobp.2007.04.037.

[8] Teguh DN, Levendag PC, Voet PW, Al-Mamgani A, Han X, Wolf TK, et al. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. Int J Radiat Oncol Biol Phys 2011;81:950–7. https://doi.org/10.1016/j.ijrobp.2010.07.009.

[9] Fortunati V, Verhaart RF, van der Lijn F, Niessen WJ, Veenland JF, Paulides MM, et al. Tissue segmentation of head and neck CT images for treatment planning: a multiatlas approach combined with intensity modeling. Med Phys 2013;40: 071905. https://doi.org/10.1118/1.4810971.

[10] Delpon G, Escande A, Ruef T, Darréon J, Fontaine J, Noblet C, et al. Comparison of automated atlas-based segmentation software for postoperative prostate cancer radiotherapy. Front Oncol 2016;6:178. https://doi.org/10.3389/fonc.2016.00178.

[11] Raudaschl PF, Zaffino P, Sharp GC, Spadea MF, Chen A, Dawant BM, et al. Evaluation of segmentation methods on head and neck CT: auto-segmentation challenge 2015. Med Phys 2017;44:2020–36. https://doi.org/10.1002/mp.12197.

[12] Weistrand O, Svensson S. The ANACONDA algorithm for deformable image registration in radiotherapy. Med Phys 2015;42:40–53. https://doi.org/10.1118/1.4894702.

[13] Kadoya N, Miyasaka Y, Yamamoto T, Kuroda Y, Ito K, Chiba M, et al. Evaluation of rectum and bladder dose accumulation from external beam radiotherapy and brachytherapy for cervical cancer using two different deformable image registration techniques. J Radiat Res 2017;58:720–8. https://doi.org/10.1093/jrr/rrx028.

[14] Jameson MG, Holloway LC, Vial PJ, Vinod SK, Metcalfe PE. A review of methods of analysis in contouring studies for radiation oncology. J Med Imaging Radiat Oncol 2010;54:401–10. https://doi.org/10.1111/j.1754-9485.2010.02192.x.

[15] Vinod SK, Min M, Jameson MG, Holloway LC. A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. J Med Imaging Radiat Oncol 2016;60:393–406. https://doi.org/10.1111/1754-9485.12462.

[16] Kittler J, Hatef M, Duin RP, Matas J. On combining classifiers. IEEE Trans Pattern Anal Mach Intell 1998;20:226–39. https://doi.org/10.1109/34.667881.

[17] Kittler J, Alkoot FM. Sum versus vote fusion in multiple classifier systems. IEEE Trans Pattern Anal Mach Intell 2003;25:110–5. https://doi.org/10.1109/TPAMI.2003.1159950.

[18] Tsuji SY, Hwang A, Weinberg V, Yom SS, Quivey JM, Xia P. Dosimetric evaluation of automatic segmentation for adaptive IMRT for head-and-neck cancer. Int J Radiat Oncol Biol Phys 2010;77:707–14. https://doi.org/10.1016/j.ijrobp.2009.06.012.

[19] Sharp G, Fritscher KD, Pekar V, Peroni M, Shusharina N, Veeraraghavan H, et al. Vision 20/20: perspectives on automated image segmentation for radiotherapy. Med Phys 2014;41:050902. https://doi.org/10.1118/1.4871620.

[20] Salembier C, Villeirs G, De Bari B, Hoskin P, Pieters BR, Van Vulpen M, et al. ESTRO ACROP consensus guideline on CT- and MRI-based target volume delineation for primary radiation therapy of localized prostate cancer. Radiother Oncol 2018;127:49–61. https://doi.org/10.1016/j.radonc.2018.01.014.

[21] Tomita N, Soga N, Ogura Y, Hayashi N, Shimizu H, Kubota T, et al. Preliminary analysis of risk factors for late rectal toxicity after helical tomotherapy for prostate cancer. J Radiat Res 2013;54:919–24. https://doi.org/10.1093/jrr/rrt025.

[22] Larrue A, Gujral D, Nutting C, Gooding M. The impact of the number of atlases on the performance of automatic multi-atlas contouring. Phys Med 2015;31:e30. https://doi.org/10.1186/s13014-015-0579-1.

[23] Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC. Morphometric analysis of white matter lesions in MR images: method and validation. IEEE Trans Med Imaging 1994;13:716–24. https://doi.org/10.1109/42.363096.

[24] 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.

[25] Kanda Y. Investigation of the freely available easy–to–use software "EZR" for medical statistics. Bone Marrow Transplant 2013;48:452–8. https://doi.org/10.1038/bmt.2012.244.

[26] Huang J, Barbera L, Brouwers M, Browman G, Mackillop WJ. Does delay in starting treatment affect the outcomes of radiotherapy? A systematic review. J Clin Oncol 2003;21:555–63. https://doi.org/10.1200/JCO.2003.04.171.

[27] Almeida G, Tavares JMRS. Deep learning in radiation oncology treatment planning for prostate cancer: a systematic review. J Med Syst. 2020;44:179. https://doi.org/10.1007/s10916-020-01641-3.