

# Comparative clinical evaluation of auto segmentation methods in contouring of prostate cancer

A. Lastrucci<sup>1</sup>, F. Meucci<sup>2</sup>, M. Baldazzi<sup>3</sup>, L. Marciello<sup>4</sup>, N.L.V. Cernusco<sup>4</sup>, E. Serventi<sup>1</sup>, S. Marzano<sup>4</sup>, R. Ricci<sup>1</sup>

<sup>1</sup> Radiation Oncology Unit, Santo Stefano Hospital, Department of Health Professions, Azienda USL Toscana Centro, Prato, 59100, Italy

<sup>2</sup> Medical Physics Unit, Santo Stefano Hospital, Azienda USL Toscana Centro, Prato- Pistoia, 59100, Italy

<sup>3</sup> Radiation Oncology Unit, University of Florence, Florence, 50100 Italy

<sup>4</sup> Radiation Oncology Unit, Santo Stefano Hospital, Department of Oncology, Azienda USL Toscana Centro, Prato, 59100, Italy

SUMMARY

**Purpose:** The aim of this study is to compare two different methods for auto-contouring for prostate radiation therapy. On the one hand Atlas-Based segmentation, on the other a deep-learning Artificial Intelligence (AI) based method by means of the recently developed software module Contour ProtégéAI of MIM Maestro (MIM Software Inc., Cleveland, OH, USA).

**Methods:** Ten patients with prostate cancer treated at the Radiotherapy Unit of S.Stefano Hospital of Prato (IT) were selected retrospectively by specific inclusion criteria. To make a comparison between the auto-segmentation methods in prostate radiation therapy the manual contouring was used as a reference, called Gold Standard. Similarity indices, like as Dice Similarity Coefficient (DSC) and Mean Distance to Agreement (MDA), are used to compare AI and Atlas-based contours with Gold Standard contours.

**Results:** Data analysis show a significant difference in results obtained by Atlas based segmentation and AI. Significant differences in DSC and MDA (in terms of mean and SD) values between the two automatic methods of segmentation are present in the prostate (Mean DSC AI  $0.78 \pm 0.07$  vs Atlas-based  $0.64 \pm 0.10$ ; Mean MDA AI  $1.42 \pm 0.34$  vs Atlas-based  $6.56 \pm 2.85$ ) and rectum (Mean DSC AI  $0.86 \pm 0.05$  vs Atlas-based  $0.58 \pm 0.13$ ; Mean MDA AI  $1.89 \pm 0.53$  vs Atlas-based  $3.75 \pm 1.51$ ). There is not a significant difference in segmentation of both femurs. Even for empty bladder both methods give good results.

**Conclusion:** In summary in the case of prostate treatment the use of software Contour ProtégéAI is extremely valid and the superiority in terms of accuracy of this method in comparison with the Atlas-based one has been shown.

**Key words:** prostate cancer, positron emission tomography, choline, recurrence

## INTRODUCTION

The contouring of target volumes and surrounding Organs at Risk (OARs) is a critical process in Radiotherapy. In general, contours are delineated manually [1]. However the manual segmentation is time-consuming and is a large contributor to radiotherapy treatment planning loss of time [2,3]. Another problem related to manual segmentation is the Inter-Oberserver Variability (IOV). IOV exists with manual contours due to differences in Radiation Oncologist training or due to the inherent quality of imaging studies; IOV may cause inaccuracy in radiotherapy treatment [4].

To standardize the contours related to target volumes and OARs, improve contouring efficiency and reduce the time of contouring many automatic segmentation methods have been developed [5].

One of the available methods for auto-contouring is Atlas-based segmentation. This automatic algorithm uses a library of images, referred to as atlases, contoured by experienced healthcare professionals, with different organ shapes and sizes that represent a wide range of anatomic variations. Atlas-based auto-contouring uses deformable image registration between an atlas image and the image to be contoured (e.g. the planning CT) to calculate a transformation which will be used to transfer atlas contours onto the planning image to be contoured [6].

Recently software tools based on Deep Learning (DL) or Artificial Intelligence (AI) were developed to assist or even completely replace the current contouring methods. Recent studies have shown excellent results for contours obtained by Artificial Intelligence software, with a greater accuracy compared to atlas-based methods [7]. In particular DL methods for segmentation use Convolution Neural Networks (CNNs). CNNs are “trained” by analysing a very large set of contoured images, i.e. the training set, through a backpropagation algorithm that optimizes CNN parameters for identifying complex spatial representations related to specific objects in an image [8].

In this study we assess the impact of auto segmentation methods to contour the bladder, rectum, prostate and femoral heads in prostate radiation therapy. Specifically the aim of this study is to compare manual contours, which are deemed as the “gold standard” contour [9], made by experts Radiation Oncologists and two different methods for auto-contouring, one atlas based and one deep learning-based.

### Address for correspondence:

Andrea Lastrucci, Radiation Oncology Unit, Santo Stefano Hospital, Department of Health Professions, Azienda USL Toscana Centro, Prato, 59100, Italy, email: andrea.lastrucci@uslcentro.toscana.it

**Word count:** 2324 **Table:** 00 **Figures:** 02 **References:** 17

Received:- 11 January, 2022, Manuscript No. M-OAR-22-51521

Editor assigned:- 13 January, 2022, PreQC No. OAR-22-51521(PQ)

Reviewed:- 20 February, 2022, QC No. OAR-22-51521(Q)

Revised:- 21 February, 2022, Manuscript No. OAR-22-51521(R)

Published:- 28 February, 2022, Invoice No. J- OAR-22-51521

## MATERIAL AND METHODS

### Patient selection

Ten patients with prostate cancer treated at the Radiotherapy Unit of S. Stefano Hospital of Prato (IT) from October 2020 and June 2021 were selected retrospectively. For each patient the contouring was done on a 3D planning CT with a slice thickness of 3 mm. All patients were prepared with an empty rectum and empty bladder. The empty bladder preparation protocol may provide better patient comfort and reproducibility during the radiation treatment and has non-inferior acute and intermediate post treatment gastrointestinal and genitourinary toxicities, compared with full bladder preparation [10]. Contrast medium was not used for any patients in this study. In this study no patient had hip implant, to avoid image artifacts. There wasn't age limit for patients. All patients with prostate cancer were treated with radical radiotherapy.

### Gold-standard

To make a comparison between the auto-segmentation methods in prostate radiation therapy the manual contouring was used as a reference [11]. To reduce IOV the reference contours, named Gold-Standard, were delineated by three senior Radiation Oncologists. Target volume and OARs in ten CT-Simulation selected images are contoured manually by each Radiation Oncologist. Starting from the contours made by all Radiation Oncologists, by means of the majority vote algorithm, Gold-Standard contours were created.

### Atlas based segmentation

Atlas-based segmentation is an automatic method used to automatically contour target volumes and OARs on the planning CT using predefined atlases and a non-rigid registration technique [12, 13].

In this study we used a software module of MIM Maestro (MIM Software Inc., Cleveland, OH, USA) that offers the possibility to work with a customizable workflow which allows users to choose some options, such as the registration algorithm or some post-processing operations.

A clinical workflow, named "Prostate Empty", was created to contour automatically target volume and OARs. In this workflow the atlas "MIM Air Masked Rectum" (version 1.0.4.G518) to contour rectum and "MIM High Risk Prostate" (version 1.0.1.G518) to contour prostate and femoral heads are added. Both atlases are proprietary and are not user editable. So we created a new customized atlas, added to workflow "Prostate Empty", to contour the empty bladder. This atlas contains 14 subjects and it is based on patient contours delineated by the expert Radiation Oncologists.

### Artificial intelligence segmentation

A newer software module developed by MIM Maestro, Contour ProtégéAI, was used for this study. Contour ProtégéAI is based on a neural network framework for automated segmentation of structures on CT images. At the moment no detailed informations on the method are available.

Contour ProtégéAI's neural network is based on the U-Net architecture, which has been used for segmentation in numerous different applications [14]. The AI database on the MIM cloud contains approximately 500-1000 registered training data for each treatment site [2].

### Quantitative evaluations of auto-segmentation

Similarity indices are used to compare AI and Atlas-based contours with Gold Standard contours and so quantify the algorithm accuracy [15].

MIM Maestro software provides a specific tool, named "Compare Contours", that automatically calculates several statistical parameters between two contours, in this study between the Gold-Standard contours and automatically generated ones. In this study, Dice Similarity Coefficient (DSC) and Mean Distance to Agreement (MDA) were used to make the comparison. In particular, DSC index is a metric expression of contours spatial overlapping. DSC is calculated as follows:

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

Where respectively A and B are the two contours evaluated. DSC is a dimensional and takes value between zero and one. When DSC value approaches unity there's complete overlapping between two contours, while if DSC value approaches zero there is not overlap [16].

The other used statistical parameter, MDA, is a measure of average similarity between two different contours. MDA represents the average distance across all contour points for each couple of selected contours. The measurement unit of MDA is millimeter. So, a high MDA between two contours is evidence of region of dissimilarity in segmentation [17].

## DISCUSSION

Data analysis show a significant difference in results obtained by Atlas based segmentation and AI. For AI segmentation DSC values indicates an optimal agreement (mean>0.85 and SD<0.07) for bladder, right and left femurs and rectum. Figure 1 shows boxplots of the range of DSC for each structure contoured.

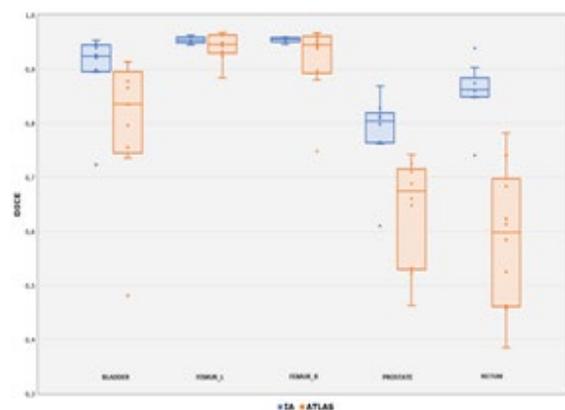


Fig. 1. Box plot of AI and Atlas-based DSC distribution

Significant differences in DSC (in terms of mean and SD) values between the two automatic methods of segmentation are present in the prostate (Mean DSC AI  $0.78 \pm 0.07$  vs Atlas-based  $0.64 \pm 0.10$ ) and rectum (Mean DSC AI  $0.86 \pm 0.05$  vs Atlas-based  $0.58 \pm 0.13$ ). There is not a significant difference in segmentation of both femurs. Even for empty bladder both methods give good results (Mean DSC AI  $0.91 \pm 0.07$  vs Atlas-based  $0.81 \pm 0.13$ ): we must however observe that the good results with Atlas are due to our customized atlas i.e. a user more time-spending customization of atlases may be mandatory. We notice also that AI data show a significant lesser spread (i.e. standard deviation) than Atlas.

In Figure 2 the mean MDA is reported for each contoured structure with the two automatic methods. MDA confirms the results obtained with DSC criterium:

- Significant differences in MDA (in terms of mean and SD) for rectum (Mean MDA AI  $1.42 \pm 0.34$  vs. Atlas-based  $6.56 \pm 2.85$ ) and prostate (Mean MDA AI  $1.89 \pm 0.53$  vs. Atlas-based  $3.75 \pm 1.51$ ).
- Similarity in segmentation of both femurs

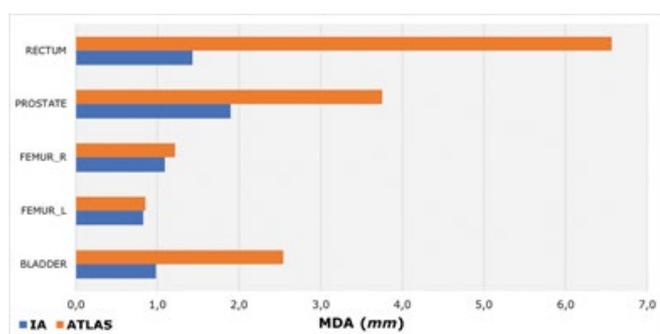


Fig. 2. Histogram of AI and Atlas- based MDA distribution

## CONCLUSION

In this study we have compared, in respect to Gold-Standard contours, two different methods of auto segmentation in ten patients with prostate cancer treated at the Radiotherapy Unit of S. Stefano Hospital of Prato. On the one hand atlas-based segmentation, on the other AI software based on deep-learning, called Contour ProtégéAI, is used to generate automatically contours. The use of the Contour ProtégéAI software proved to be particularly efficient as regards the contouring of the rectum and prostate as well as the bladder, albeit with regard to the latter the results obtained by the appropriately created atlas can be considered valid. On the basis of the obtained data, a minor accuracy of Atlas-based segmentation method has been shown, which do not allow obtaining similar or comparable results to the Gold Standard method.

In summary, In the case of prostate treatment the use of software Contour ProtégéAI is extremely valid. To apply Contour ProtégéAI to clinical practice, however, it will be necessary on the one hand to increase the number of cases in the population examined, to further confirm the obtained data so far, and on the other hand making the population more heterogeneous to evaluate the contouring of OAR and target volume even in those patients in which there may be hip implants.

- |            |  |
|------------|--|
| REFERENCES | <ol style="list-style-type: none"> <li>Kim N, Chang JS, Kim YB. Atlas-based auto-segmentation for postoperative radiotherapy planning in endometrial and cervical cancers. <i>Radiat Oncol J.</i> 2020;15:1-9.</li> <li>Urago Y, Okamoto H, Kaneda T. Evaluation of auto-segmentation accuracy of cloud-based artificial intelligence and atlas-based models. <i>Radiat Oncol J.</i> 2021;16:1-3.</li> <li>Wong J, Huang V, Wells D. Implementation of deep learning-based auto-segmentation for radiotherapy planning structures: a workflow study at two cancer centers. <i>Radiat Oncol J.</i> 2021 ;16 :1-10.</li> <li>Wong J, Fong A, Mc Vicar N. Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning. <i>Radiat Oncol J.</i> 2020 ;144:152-158.</li> <li>Chen M, Wu S, Zhao W. Application of deep learning to auto-delineation of target volumes and organs at risk in radiotherapy. <i>Cancer Radiother</i> 2021;26.</li> <li>Zabel WJ, Conway JL, Gladwish A. Clinical evaluation of deep learning and atlas-based auto-contouring of bladder and rectum for prostate radiation therapy. <i>Pract Radiat Oncol</i> 2021;11:e80-e89.</li> <li>Lustberg T, Van Soest J, Gooding M. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. <i>Radiat Oncol</i> 2018;126:312-317.</li> <li>Bengio Y, LeCun Y. Scaling learning algorithms towards AI. Large-scale kernel machines. 2007;34:1-41.</li> <li>Yuzhen NW, Barrett SA. Review of automatic lung tumour segmentation in the era of 4DCT. <i>Reports Rep Pract Oncol Radiother.</i> 2019;24:208-220.</li> <li>Tsang YM, Hoskin P. The impact of bladder preparation protocols on post treatment toxicity in radiotherapy for localised prostate cancer patients. <i>Tech Innov Patient Support Radiat Oncol</i> 2017;3:37-40.</li> <li>Da Silva BA, Fazenda AL, Paixão FC. Femoral Head Autosegmentation for 3D Radiotherapy Planning: Preliminary Results. <i>arXiv preprint arXiv:1812.04682.</i> 2018.</li> <li>Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. <i>BMC Med.</i> 2015;15:1-28.</li> <li>Militello C, Rundo L, Toia P. A semi-automatic approach for epicardial adipose tissue segmentation and quantification on cardiac CT scans. <i>Comput Biol Med.</i> 2019;114:103424.</li> <li>Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. <i>MICCAI</i> 2015;234-241.</li> <li>Casati M, Piffer S, Calusi S. Methodological approach to create an atlas using a commercial auto-contouring software. <i>J. Appl Clin Med Phys.</i> 2020;21:219-230.</li> <li>Zou KH, Warfield SK, Bharatha A. Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. <i>Acad Radiol.</i> 2004;11:178-189.</li> <li>Hammers JE, Pirozzi S, Lindsay D. Evaluation of a commercial DIR platform for contour propagation in prostate cancer patients treated with IMRT/VMAT. <i>J Appl Clin Med Phys.</i> 2020; 2:14-25.</li> </ol> |
|------------|--|