

# Auto-segmentation of pelvic structures using MRI Planner; A quantitative evaluation

Laura G. Bentvelzen<sup>1</sup>, Rickard O. Cronholm<sup>1</sup>, Anders Karlsson<sup>1</sup>, Carl Siversson<sup>1,2\*</sup>

## Abstract

Accurate delineation of organs at risk (OAR) is a crucial step in radiotherapy to ensure efficient treatment delivery. However, the delineation process is labor-intensive and prone to inter-observer variation. As a solution, MRI Planner (Spectronic Medical AB, Helsingborg, Sweden), distributed as commercially available software, provides automated organ delineations using only a single standard MRI acquisition. The aim of this study is to quantitatively evaluate the quality of these automatically generated delineations for OAR in the male pelvic anatomy.

Forty-one prostate cancer patients from multiple centers were included in the study. The bladder, femoral heads and rectum were considered as OAR for the treatment. For every patient, ground-truth segmentations were manually created and automated segmentations were generated by the MRI Planner software. Testing of the segmentation performance was done by comparing multiple quantitative similarity metrics between automated and corresponding ground-truth manual OAR segmentations for each case. The dice similarity coefficient (DSC), mean surface distance (MSD), volume similarity ( $Vol_{sim}$ ) and surface DSC (SDSC) were computed for the femoral heads, rectum and bladder structures.

Evaluation results showed accurate segmentation performance. Average DSC results were  $0.96 \pm 0.01$  (femoral heads),  $0.95 \pm 0.03$  (bladder) and  $0.90 \pm 0.02$  (rectum). Mean SDSC values at distance tolerance of 3mm were  $0.99 \pm 0.01$ ,  $0.98 \pm 0.03$  and  $0.95 \pm 0.03$  for femoral heads, bladder and rectum respectively. The average MSD obtained were  $0.53 \pm 0.29$  mm (bladder),  $0.83 \pm 0.24$  mm (rectum) and  $0.43 \pm 0.11$  mm (femoral heads).

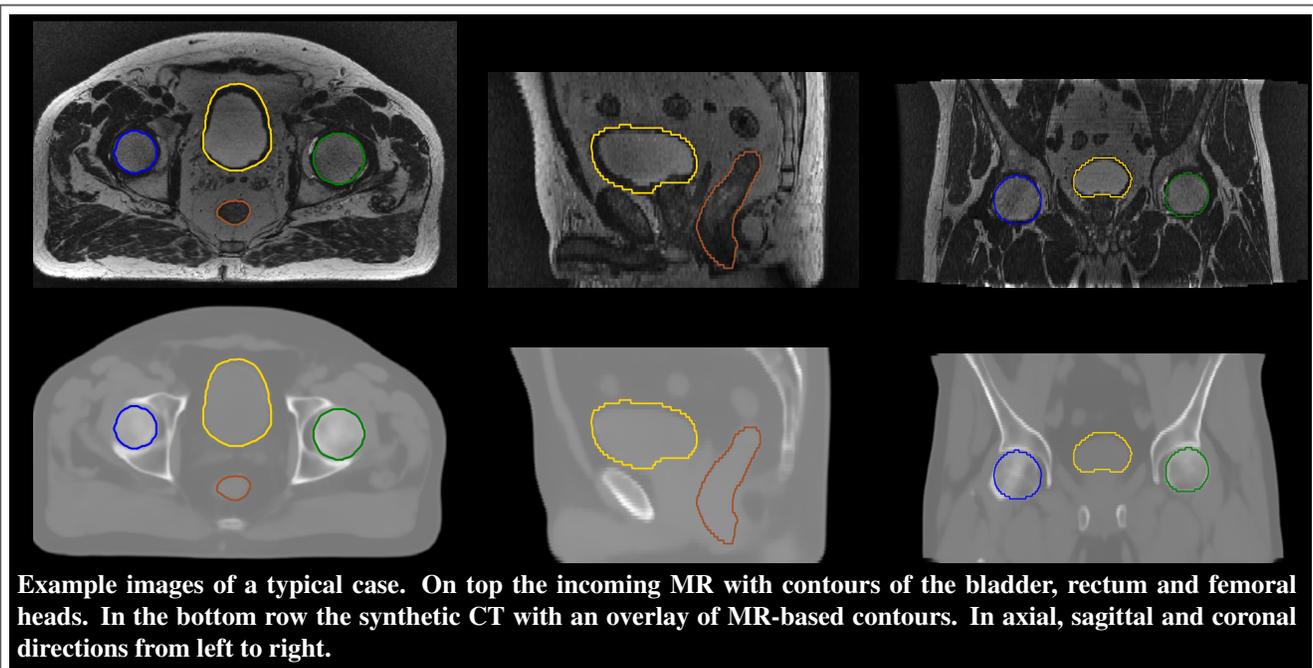
Volume differences were on average not larger than 5% for bladder and rectum and not larger than 3% for the femoral heads. The quantitative metrics show that our software provides highly accurate segmentation performance and indicate that the use of MRI Planner is an effective way to improve the prostate radiotherapy workflow.

<sup>1</sup>Spectronic Medical AB, Helsingborg, Sweden

<sup>2</sup>Department of Medical Radiation Physics, Lund University, Malmö, Sweden

\*Corresponding author: carl.siversson@spectronic.se

Published: June 2020



Example images of a typical case. On top the incoming MR with contours of the bladder, rectum and femoral heads. In the bottom row the synthetic CT with an overlay of MR-based contours. In axial, sagittal and coronal directions from left to right.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Methods</b>	<b>2</b>
2.A	Data . . . . .	2
2.B	Segmentations . . . . .	2
2.C	Evaluation . . . . .	2
<b>3</b>	<b>Results</b>	<b>3</b>
<b>4</b>	<b>Discussion</b>	<b>4</b>
<b>5</b>	<b>Conclusions</b>	<b>4</b>
	<b>Acknowledgments</b>	<b>4</b>
	<b>References</b>	<b>4</b>

## 1. Introduction

An important step in the current radiotherapy workflow is accurate delineation of organs-at-risk (OAR) to minimize radiation induced toxicities in healthy tissues. Manual delineation is a time-consuming and labor-intensive repetitive task. Moreover, the process is sensitive to inter-observer variability [1].

Auto-segmentation algorithms using deep learning frameworks are quickly gaining ground in medical image segmentation [2]. Automatic OAR segmentation methods in the radiotherapy field reduce variability and decrease the labor-intensive nature of OAR delineation, while achieving high agreement between automated and manual delineations.

The MRI Planner (Spectronic Medical, Helsingborg, Sweden) software generates synthetic CT images from MRI using the deep-learning driven transfer function estimation (TFE) algorithm [3], for use in MRI only radiotherapy treatment planning. In addition to synthetic CT images, it also provides automated delineations of organs at risk from a single standard MRI acquisition. MRI Planner is a commercially available CE marked software that can be installed on a computer at the hospital. MRI Planner uses standard DICOM files to allow for seamless integration with both MRI scanners and treatment planning systems.

The aim of this study is to evaluate the quality of the automatically generated structures by comparing several quantitative similarity metrics, between manual segmentations (serving as ground-truth) and automatically generated segmentations for prostate radiotherapy planning.

## 2. Methods

### 2.A Data

Forty-one patients with prostate cancer undergoing radiotherapy were retrospectively included in this multi-center evaluation study. T2-weighted MR images were acquired at three different radiotherapy centers in Sweden, using 3T GE Discovery, 3T GE Signa and 1.5T Siemens Aera MRI scanners. Imaging was performed as part of a previous non-interventional multicenter study[4]. Patients with one or multiple hernias of the bladder were excluded for this study.

### 2.B Segmentations

The femoral heads, the bladder and the rectum were defined as OARs. The delineations used for the ground-truth rectum segmentations were drawn according to the ESTRO guidelines formed by Salembier et al [5]. The bladder was manually segmented including the bladder wall and femurs were segmented by manual delineation of the femoral heads.

The automated segmentations were generated by the deep learning driven MRI Planner (v2.3) software (Spectronic Medical AB, Helsingborg, Sweden)[3].

### 2.C Evaluation

The dice similarity coefficient (DSC), mean surface distance (MSD), volume similarity ( $Vol_{sim}$ ) and surface DSC (SDSC) are used as quantitative measures to evaluate the automatically generated delineations. For each patient, the metrics were compared between automatically generated and ground truth delineations. The DSC is defined in equation 1 where  $GT$  (ground truth) and  $INF$  (inferred) refer to the volumes of the manual and automatically generated delineations respectively.  $|GT \cap INF|$  represents the intersection and  $|GT| + |INF|$  the union of  $GT$  and  $INF$ .

$$DSC = \frac{2|GT \cap INF|}{|GT| + |INF|} \quad (1)$$

While DSC indicates geometrical similarity between two structures, it does not necessarily relate to clinical importance and/or time needed to do contour adaptations. The surface dice coefficient (SDSC) proposed by Nikolov *et al.* indicates the overlap of two surfaces instead of two volumes at a specified distance tolerance  $\tau$  [6]. Resulting values range between 0 and 1, where a higher value corresponds to reduced likelihood that the contour needs manual adjustments/corrections. Calculations were done using the open source DeepMind implementation of the surface DSC [7].

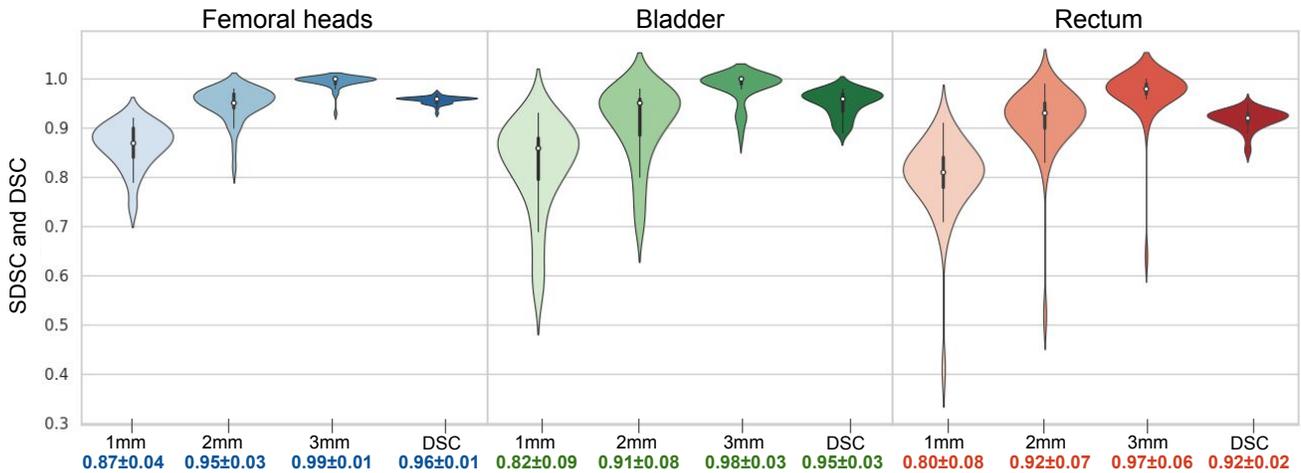
The volumetric similarity is defined in equation 2 with  $INF$  and  $GT$  referring to the volume of the automated and manual delineations respectively. The  $Vol_{sim}$  indicates the volume difference of the predicted segmentation with respect to the ground truth segmentation.

$$Vol_{sim} = \frac{2(INF - GT)}{INF + GT} \quad (2)$$

The MSD (equation 3) is used as a symmetric measure, using the mean of the euclidean distances in mm from the contour points of the  $GT$  segment to the contour of the  $INF$  segment, and vice versa.

$$MSD = mean\left(\underset{x \sim X}{mean}(d(x, Y)), \underset{y \sim Y}{mean}(d(y, X))\right) \quad (3)$$

with  $d$  = euclidean distance from a point on contour A to contour B.



**Figure 1.** Violinplots showing the distribution of results across cases for femoral heads, bladder and rectal structures. The plots extend to minimum and maximum values along with the median (white dot) and the third to first interquartile range (black bar). Mean values  $\pm$  SD are indicated at the bottom in colored text.

The results are obtained by evaluating on the union of ground-truth and automatically generated delineations in 3D for the bladder and femoral heads. The cranial boundary of the rectum is known to be challenging to determine [8], resulting in a possible variation between manual and automatically generated delineations. To reflect clinically relevant differences between delineations, rectum delineations are evaluated on the intersection in caudal-cranial direction and on the union in anterior-posterior and left-right directions.

### 3. Results

The mean MSD between manual and automatically generated delineations ranged between 0.43 mm and 0.63 mm for all structures and the mean volumetric similarity between -0.04 and 0.05 (Table 1).

The SDSC at a tolerance level of 3mm was 0.97, 0.98 and 0.99 on average for rectum, bladder and femoral heads respectively (Table 2). Indicating for example that for the rectum 97% of the automatically generated contour does not deviate more than 3mm from the manual contour.

Structures	MSD [mm] [ $\mu \pm \sigma$ ]	$Vol_{sim}$ [ $\mu \pm \sigma$ ]
<b>Bladder</b>	0.53 $\pm$ 0.29	-0.04 $\pm$ 0.05
<b>Rectum</b>	0.63 $\pm$ 0.33	0.05 $\pm$ 0.06
<b>Femoral Heads</b>	0.43 $\pm$ 0.11	-0.02 $\pm$ 0.03

**Table 1.** The average MSD (mean surface distance) and average  $Vol_{sim}$  (volumetric similarity) of the predicted segmentations compared to the ground-truth segmentations over all cases.

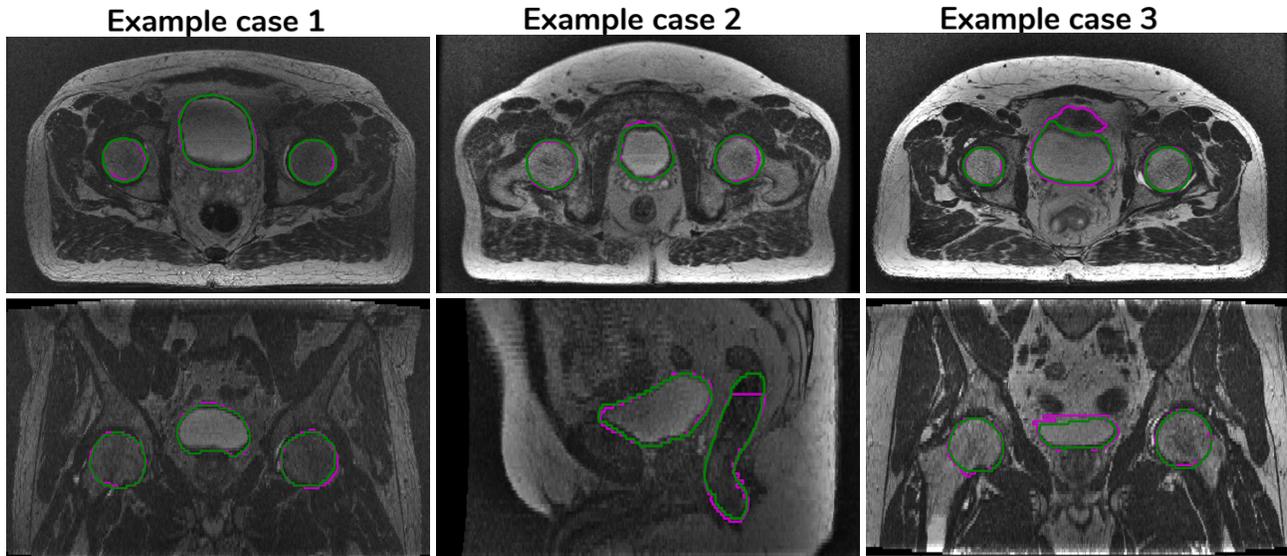
In Figure 1 the SDSC results for tolerances set at 1mm, 2mm and 3mm are visualized, as well as the DSC. By its nature, SDSC results improve with higher physical tolerances.

DSC values of 0.92, 0.95 and 0.96 were obtained for rectum, bladder and femoral heads respectively (Table 2).

In Figure 2 automatically generated delineations for three example cases are shown together with the manual delineations. Case 1 shows a typical segmentation result, with negligible differences between automatically generated and manual contours. Case 2 (sagittal slice) shows a longer rectum contour compared to the manual contour. This example case also shows the largest femoral head discrepancy seen across all cases. Case 3 is interesting in the sense that the automatically generated bladder contour is clearly superior to the manual contour, since it does not extend into the colon wall.

Structures	DSC [ $\mu \pm \sigma$ ]	SDSC* [ $\mu \pm \sigma$ ]
<b>Bladder</b>	0.95 $\pm$ 0.03	0.98 $\pm$ 0.03
<b>Rectum</b>	0.92 $\pm$ 0.02	0.97 $\pm$ 0.06
<b>Femoral Heads</b>	0.96 $\pm$ 0.01	0.99 $\pm$ 0.01

**Table 2.** The average DSC (Dice Similarity Coefficient) and average SDSCs (Surface DSC) at [ $\tau$ ]=3mm of the predicted segmentations compared to the ground-truth segmentations over all cases.



**Figure 2.** Magenta: Manual, Green: Automatically generated. Column 1: A typical case. Column 2: Axial slice showing the largest difference in femoral heads across cohort (worst case) and sagittal slice demonstrating a difference in the length of rectum contours. Column 3: Axial slice shows the manual contour falsely including part of colon, whereas the automatically generated contour was able to distinguish between the organs.

## 4. Discussion

Automatically generated OAR contours provided by the MRI Planner software are accurate and show high similarity with manually created contours based on multiple quantitative similarity metrics (DSC, SDSC, MSD and  $Vol_{sim}$ ).

Among the OARs, the femoral heads achieved on average the highest SDSC and DSC and lowest MSD. DSC and MSD results for all structures are in the top range as compared to other related work [9, 10, 11, 12, 13, 14, 15]. The femoral heads yield a DSC of  $0.96 \pm 0.01$ , the corresponding range found in other work is [0.92 - 0.97]. For the bladder and rectum, we achieved DSC results of  $0.95 \pm 0.03$  and  $0.90 \pm 0.02$  with corresponding ranges from recent literature [0.93 - 0.95] and [0.84 - 0.92] respectively.

Our MSD results are  $0.53 \pm 0.29$  mm and  $0.83 \pm 0.24$  mm with ranges in literature of [0.44 - 1.37] mm and [0.72 - 1.83] mm for bladder and rectum respectively [10, 12, 13, 14]. The MSD for femoral heads is not often reported in literature, so no comparison could be made. We achieved a MSD of  $0.43 \pm 0.11$  mm which is in the order of the in-plane resolution.

The results of the SDSC at  $\tau = 3$ mm are  $0.99 \pm 0.01$ ,  $0.98 \pm 0.03$  and  $0.95 \pm 0.03$  for femoral heads, bladder and rectum and indicate high similarity between automated and manual contours.

Differences in volume between 3D contours are on average not larger than 5% for bladder and rectum and not larger than 3% for the femoral heads.

Some degree of disagreement could be expected as we consider manual delineations as the ground-truth for evaluation in this work. Taking into account the possible inter-observer vari-

ability the manual contours cannot be considered the perfect ground-truth but rather a gold standard [16, 17].

The achieved results are in good agreement with the manual delineations. The findings of this study suggests that the performance of the commercial MRI Planner software is well on par with state of the art results presented in the scientific literature.

## 5. Conclusions

This study demonstrates that MRI Planner creates highly accurate automatically generated delineations of OARs.

By using a deep-learning based method that requires only a single MR acquisition, the software aids in decreasing the labor required for OAR delineation and provides excellent segmentation performance without the need of manual interventions.

Quantitative metrics show that our software performance is in the top range compared to recent scientific work and indicate that the use of MRI Planner is an effective way to streamline the radiotherapy workflow.

## Acknowledgments

This work was supported the Swedish Agency for Innovation Systems (VINNOVA).

We would like to thank Skåne University Hospital, Umeå University Hospital and Sahlgrenska University Hospital for the image material used in this work[4].

The authors declare that they are employed by Spectronic Medical AB.

## References

- [1] Shalini K. Vinod, Michael G. Jameson, Myo Min, and Lois C. Holloway. Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies. *Radiotherapy and Oncology*, 121(2):169–179, 2016.
- [2] Carlos E. Cardenas, Jinzhong Yang, Brian M. Anderson, Laurence E. Court, and Kristy B. Brock. Advances in Auto-Segmentation. *Seminars in Radiation Oncology*, 29(3):185–197, 2019.
- [3] Spectronic Medical AB. MRI only radiotherapy planning using the transfer function estimation algorithm. <http://www.spectronic.se/reports/>. Accessed: 2020-02.
- [4] E. Persson, C. Gustafsson, F. Nordström, M. Sohlin, A. Gunnlaugsson, K. Petruson, M. Rintelä, K. Hed, L. Blomqvist, B. Zackrisson, T. Nyholm, L.E. Olsson, C. Siversson, and J. Jonsson. Mr-opera : A multicenter/multivendor validation of magnetic resonance imaging-only prostate treatment planning using synthetic computed tomography images. 99(3):692–700, 2017.
- [5] Carl Salembier, Geert Villeirs, Berardino De Bari, Peter Hoskin, Bradley R. Pieters, Marco Van Vulpen, Vincent Khoo, Ann Henry, Alberto Bossi, Gert De Meerleer, and Valérie Fonteyne. ESTRO ACROP consensus guideline on CT- and MRI-based target volume delineation for primary radiation therapy of localized prostate cancer. *Radiotherapy and Oncology*, 127(1):49–61, 2018.
- [6] Stanislav Nikolov, Sam Blackwell, Ruheena Mendes, Jeffrey De Fauw, Clemens Meyer, Cían Hughes, Harry Askham, Bernardino Romera-Paredes, Alan Karthikesalingam, Carlton Chu, Dawn Carnell, Cheng Boon, Derek D’Souza, Syed Ali Moinuddin, Kevin Sullivan, DeepMind Radiographer Consortium, Hugh Montgomery, Geraint Rees, Ricky Sharma, Mustafa Suleyman, Trevor Back, Joseph R. Ledsam, and Olaf Ronneberger. Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy. pages 1–31, 2018.
- [7] DeepMind. surface-distance. <https://github.com/deepmind/surface-distance>, 2018. Accessed: 2020-03.
- [8] B. Haas, T. Coradi, M. Scholz, P. Kunz, M. Huber, U. Opitz, L. André, V. Lengkeek, D. Huyskens, A. Van Esch, and R. Reddick. Automatic segmentation of thoracic and pelvic CT images for radiotherapy planning using implicit anatomic knowledge and organ-specific segmentation strategies. *Physics in Medicine and Biology*, 53(6):1751–1771, 2008.
- [9] Anjali Balagopal, Samaneh Kazemifar, Dan Nguyen, Mu Han Lin, Raquibul Hannan, Amir Owrangi, and Steve Jiang. Fully automated organ segmentation in male pelvic CT images. *Physics in Medicine and Biology*, 63(24):245015, dec 2018.
- [10] Xue Dong, Yang Lei, Sibotian, Tonghe Wang, Pretesh Patel, Walter J. Curran, Ashesh B. Jani, Tian Liu, and Xiaofeng Yang. Synthetic MRI-aided multi-organ segmentation on male pelvic CT using cycle consistent deep attention network. *Radiotherapy and Oncology*, 141:192–199, dec 2019.
- [11] Sharif Elguindi, Michael J. Zelefsky, Jue Jiang, Harini Veeraraghavan, Joseph O. Deasy, Margie A. Hunt, and Neelam Tyagi. Deep learning-based auto-segmentation of targets and organs-at-risk for magnetic resonance imaging only planning of prostate radiotherapy. *Physics and Imaging in Radiation Oncology*, 12(October):80–86, 2019.
- [12] Yaozong Gao, Yeqin Shao, Jun Lian, Andrew Z. Wang, Ronald C. Chen, and Dinggang Shen. Accurate Segmentation of CT Male Pelvic Organs via Regression-Based Deformable Models and Multi-Task Random Forests. *IEEE Transactions on Medical Imaging*, 35(6):1532–1543, jun 2016.
- [13] Samaneh Kazemifar, Anjali Balagopal, Dan Nguyen, Sarah McGuire, Raquibul Hannan, Steve Jiang, and Amir Owrangi. Segmentation of the prostate and organs at risk in male pelvic CT images using deep learning. *Biomedical Physics and Engineering Express*, 4(5):055003, jul 2018.
- [14] Yang Lei, Tonghe Wang, Sibotian, Xue Dong, Ashesh B Jani, David Schuster, Walter J Curran, Pretesh Patel, Tian Liu, and Xiaofeng Yang. Male pelvic multi-organ segmentation aided by CBCT-based synthetic MRI. *Physics in Medicine & Biology*, 65(3):035013, dec 2019.
- [15] Kuo Men, Jianrong Dai, and Yexiong Li. Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks:. *Medical Physics*, 44(12):6377–6389, 2017.
- [16] Marcel Van Herk. Errors and Margins in Radiotherapy. *Seminars in Radiation Oncology*, 14(1):52–64, 2004.
- [17] Gregory Sharp, Karl D. Fritscher, Vladimir Pekar, Marta Peroni, Nadya Shusharina, Harini Veeraraghavan, and Jinzhong Yang. Vision 20/20: Perspectives on automated image segmentation for radiotherapy. *Medical Physics*, 41(5):1–13, 2014.