Radiotherapy and Oncology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Deep learning-based auto-segmentation of organs at risk in high-dose rate brachytherapy of cervical cancer

Reza Mohammadi [a,*], Iman Shokatian [a], Mohammad Salehi [a], Hossein Arabi [b], Isaac Shiri [b], Habib Zaidi [b,c,d,e]

[a] Department of Medical Physics, Iran University of Medical Sciences, Tehran, Iran; [b] Division of Nuclear Medicine and Molecular Imaging; [c] Geneva University Neurocenter, Geneva University, Switzerland; [d] Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Netherlands; [e] Department of Nuclear Medicine, University of Southern Denmark, Odense, Denmark

ARTICLE INFO

Article history:
Received 25 January 2021
Received in revised form 20 March 2021
Accepted 24 March 2021
Available online 6 April 2021

Keywords:
Deep learning
High-dose rate brachytherapy
Segmentation
Locally-advanced cervical cancer

ABSTRACT

Background and purpose: Delineation of organs at risk (OARs), such as the bladder, rectum and sigmoid, plays an important role in the delivery of optimal absorbed dose to the target owing to the steep gradient in high-dose rate brachytherapy (HDR-BT). In this work, we propose a deep convolutional neural network-based approach for fast and reproducible auto-contouring of OARs in HDR-BT.

Materials and methods: Images of 113 patients with locally-advanced cervical cancer were utilized in this study. We used ResU-Net deep convolutional neural network architecture, which uses long and short skip connections to improve the feature extraction procedure and the accuracy of segmentation. Seventy-three patients chosen randomly were used for training, 10 patients for validation, and 30 patients for testing. Well established quantitative metrics, such as Dice similarity coefficient (DSC), Hausdorff distance (HD), and average symmetric surface distance (ASSD), were used for evaluation.

Results: The DSC values for the test dataset were 95.7 ± 3.7%, 96.6 ± 1.5% and 92.2 ± 3.3% for the bladder, rectum, and sigmoid, respectively. The HD values (mm) were 4.05 ± 5.17, 1.96 ± 2.19 and 3.15 ± 2.03 for the bladder, rectum, and sigmoid, respectively. The ASSDs were 1.04 ± 0.97, 0.45 ± 0.09 and 0.79 ± 0.25 for the bladder, rectum, and sigmoid, respectively.

Conclusion: The proposed deep convolutional neural network model achieved a good agreement between the predicted and manually defined contours of OARs, thus improving the reproducibility of contouring in brachytherapy workflow.

High-dose rate brachytherapy (HDR-BT) is one of the most important treatment modalities for cervical cancers commonly employed as a primary or boost treatment option. The delineation of organs at risk (OARs), such as the bladder, rectum and sigmoid affects the delivered dose to the target volume, which is an important factor in high-dose rate brachytherapy owing to the high steep dose gradient [1]. Moreover, there is limited time to perform the whole planning process for both the radiation oncologist and physicist from patient scanning to patient treatment. Recommendations and guidelines have been published by Groupe Européen de Curiethérapie (GEC) and the European Society for Radiotherapy & Oncology (ESTRO) based on three-dimensional (3D) image-guided treatment planning for organs delineation, applicator reconstruction, prescribing and reporting [2–6]. Various studies investigated the effect of uncertainties in organ delineation, which consist of inter- and intra-observer variability in dosimetric parameters, such as the maximum dose to 2 cm^3 (D_{2cc}) of OARs [1,7,8]. Hellebust et al. reported 5–8% inter-observer variability based on MR images [7], while Saarnak et al. observed 10–11% based-on CT images for the bladder and rectum [1].

Recently, a number of approaches have been proposed to reduce inter- and intra-observer variability and processing time using auto-segmentation techniques. This included atlas-based auto-segmentation (ABAS) algorithms, which are widely used for contouring of OARs [9–14]. ABAS algorithms use deformable image registration to transform OAR contours from the atlas into the new image. There is a lack of certainty in deformable registration algorithms [15–17], particularly when intensity-based algorithms are used. In addition, there are a number of issues that must be considered when using ABAS since there is a large variability in the pelvic region owing to internal organs motion, bladder filling, and presence of gas in the rectum. These factors result in sub-optimal performance of ABAS, in particular when a small number of atlases (typically 10–20) are used [18–21]. Moreover, the presence of applicators and CT markers for brachytherapy implantation tends to decrease contrast to noise ratio and to increase high-density

* Corresponding author at: Department of Medical Physics, Iran University of Medical Sciences, Shahid Hemmat Highway, Tehran 1449614535, Iran.
E-mail address: Reza021mohammadi@gmail.com (R. Mohammadi).

https://doi.org/10.1016/j.radonc.2021.03.030
0167-8140/

artifacts of CT images. ABAS algorithms usually exhibit suboptimal performance for small and thin OARs, such as swallowing muscles [22].

Deep learning approaches have been recently applied to various medical imaging modalities to increase the accuracy, reproducibility, and consistency of organ segmentation [23–28]. Various deep learning algorithms, including recurrent neural networks (RNNs), restricted residual networks (ResNets), encoder decoder, and convolutional neural networks (CNNs) were devised to address typical challenges encountered in medical imaging research [29–32]. The CNN model is one of the most common algorithms used in medical imaging owing to its ability to extract the representative features that enable to solve most challenging tasks. For specific applications, CNN models perform better than restricted Boltzmann machines (RBMs) and stacked auto-encoders (SAEs) since the spatial and structural information in the input images is directly explored by CNN models while the inputs for RBMs and SAEs models should be in the form of vectors, wherein structural information is missed or become difficult to detect. CNN models have been used for image registration, auto-segmentation, and classification [33–36]. Some studies used CNN in auto-contouring for regions, such as the head and neck and lung cancers [37–39]. Van Dijk et al. [34] compared deep learning-guided segmentation with ABAS for delineation of OARs in head and neck cancers using an independent validation dataset. ABAS failed to delineate the esophagus, glottic area, and left and right regions of arytenoid. They concluded that the deep learning model was superior to ABAS for twenty-two OARs.

To the best of our knowledge, there is a limited number of studies reporting on the auto-segmentation of OARs in brachytherapy treatment planning for cervical cancers [40]. The aim of this study is to assess the feasibility of auto-segmentation using a deep learning approach for OARs delineation considering manual contouring as standard of reference.

Materials and methods

Clinical dataset

Images of 113 patients with locally-advanced cervical cancer collected between 2017 and 2019 were included in this study protocol. All patients were initially treated with external beam radiation therapy (25 fractions, 2 Gy/fraction) and HDR-BT (3 fractions, 8.6 Gy/fraction) as a boost treatment. For each patient, Foley catheter was inserted into the bladder before each brachytherapy implantation session. A combination of 1 cm^3 contrast agent (Meglumine compound, Amp VISIPAQUETM 320) and 6 cm^3 normal saline was injected into the balloon of Foley catheter. The bladder was filled with 120 cm^3 of normal saline in the CT scanning room. All patients were scanned on a volumetric HiSpeed Dual Slice CT scanner (GE Healthcare, USA). Images were reconstructed with 512 $\times$ 512 matrix size and 3 mm slice thickness.

A radiation oncologist contoured the clinical target volume (CTV), including the cervix, entire uterus, bilateral parametria, upper half of vagina, and lymph nodes for external beam radiation therapy and high risk (HR)-CTV and intermediate risk (IR)-CTV for brachytherapy according to international guidelines and recommendations. Relevant OARs included for both treatments were the bladder, rectum, and sigmoid.

Data preprocessing

Binary masks of the OARs were created for each patient using the 3D-Slicer software and RT structure data. All images were cropped to body contour for computational cost reduction and effective training process. All binary masks (supplemental Fig. 1) were converted into one-hot vector with values ranging from 0 to 4 as indicated in supplemental Table 1.

Convolutional neural network model

A combination of ResNet [41] and Unet [42] model (called ResU-Net) was developed to create a deeper convolutional neural network for delineating OARs from CT images as depicted in supplemental Fig. 2b. We also used a loss function formulated in Eq. (1) based on a combination of binary cross-entropy and Dice similarity coefficient (DSC) regulated with constant weighting factors.

$$Loss = [-\frac{1}{N}\sum_{i=1}^{N} y_i \times log(p(y_i)) + (1 - yi) \times log(1 - p(y_i))] \times w + (1 - w) \times [-\frac{2\sum_i p(y_i) y_i}{\sum_i p(y_i) + \sum_i y_i}] \quad (1)$$

N stands for the number of classes, $p(y_i)$ is a probability distribution of each channel, y_i is ground truth and w is a weighting factor. Details are provided in Supplemental material.

Training details

Seventy-three patients were used for training, 10 patients for validation and 30 patients for testing (hold-out, unseen during training). The training set was generated from randomly selected 2D slices with their corresponding binary masks. For the training, 2D image slices were used as input and all OARs corresponding to slices used as ground truth. In the test phase, 2D slices of each patient were used as input where the output was a 2D mask with similar resolution that was converted to OARs contour using the marching squares algorithm [43]. The same training procedure was repeated for the standard Unet architecture considered as a standard baseline model in medical image segmentation [44,45]. Then, the performance of the proposed model (ResU-Net) was compared to the Unet model providing a baseline for an insightful assessment of the results. Implementation details is provided in Supplement material.

Performance measurement

The 30 patients were used to assess the performance of our proposed model wherein 2D slices of CT images underwent segmentation slice-by-slice. Segmentation results from the CNN model and manual contouring were quantitatively compared using DSC, hausdorff distance (HD), hausdorff distance 95th percentile ($HD_{95\%}$), volume correlation (VC), average symmetric surface distance (ASSD) and precision metrics. Details are provided in Supplement material.

Dosimetric assessment

To assess the effect of predicted segmentation inaccuracies on dose-volume parameters (such as D5cc, D2cc, D1cc, and D0.1 cc for OARs, where Dxcc represents the minimum dose received by $\times$ cm^3 of an OAR), treatment plans based on manual segmentations were generated. Subsequently, the contours predicted from ResU-net and Unet models were imported to the treatment planning system to calculate the abovementioned dose-volume parameters and to compare them with the corresponding parameters derived from manual contouring. Plans were generated using SagiPlan v2.0.2 (Eckert & Ziegler BEBIG Co., Germany) platform according to the OARs dose constraint and prescribed dose for the target (HR-CTV) recommended by the GEC-ESTRO, and the American Brachytherapy Society (ABS). A prescription dose of 80–90 Gy_{EQD2} (45–50 Gy as EBRT) was considered for the target lesion, equivalent

to an absorbed dose of 2 Gy per fraction (EQD2) assuming α/β = 10, and maximum dose (D2cc) of 90 Gy_{EQD2}, 75 GyEQD2, and 75 GyEQD2 to the bladder, rectum, and sigmoid OARs, respectively, assuming α/β = 3.

Statistical analysis

Numbers v10 (Apple, Inc.) was used for statistical analyses, where mean ± standard deviation (SD) or the median (minimum, maximum, first-quartile (Q1), third-quartile (Q3)) were used for presenting and summarizing the results. Interquartile range (IQR) was also employed to find outliers in the HD, HD95% and ASSD results for all organs which were equal to differences between the third and first quartiles. 1.5 IQR was considered as the outlier criteria. Independent two samples Student's *t*-test with equal variance was used to compare the parameters considering p-value < 0.05 as threshold for statistical significance.

Results

The training time of our ResU-Net CNN model and Unet were approximately 5 and 6 h and the predicted organ contours could be created within 1.5 and 1.6 seconds, respectively. Examples of manual and predicted contours from ResU-Net and Unet models for the bladder, rectum, and sigmoid are shown in Fig. 1. The mean ± SD and the median, minimum, maximum, Q1 and Q3 of DSC, HD, $HD_{95\%}$, ASSD, VC, precision from the ResU-Net and Unet models are summarized in Tables 1 and 2, respectively, whereas

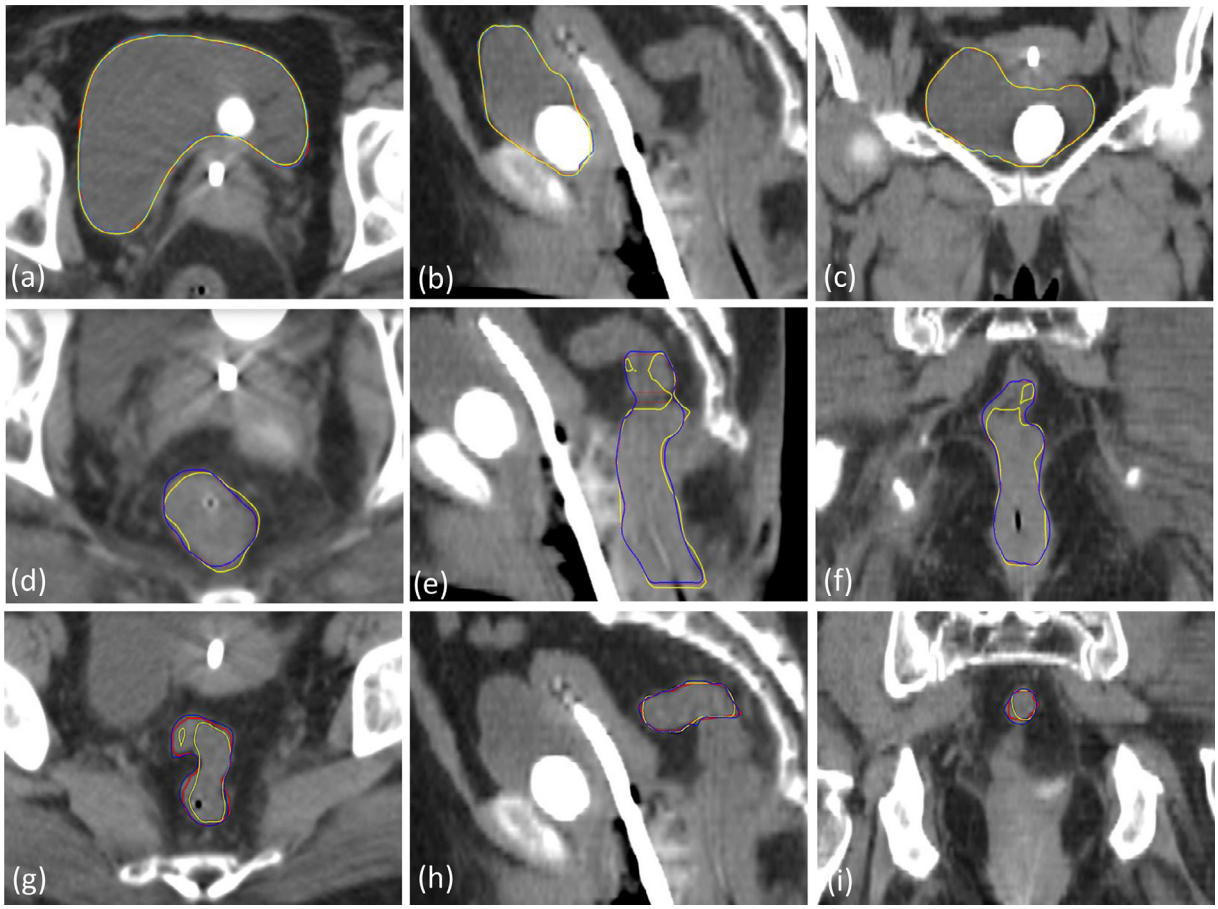


Fig. 1. Representative clinical example showing contours for a patient predicted by ResU-Net and Unet models compared to manual contours for (a-c) the bladder, (d-f) the rectum and (g-i) the sigmoid. Blue color = predicted contour from ResU-Net, Yellow color = predicted contour from Unet, Red color = manual contour.

Table 1
Results of quantitative parameters for the segmentation of the three organs. All values are reported as mean ± SD.

Model	Organ	Quantitative parameters					
		DSC (%)	HD (mm)	$HD_{95\%}$ (mm)	ASSD (mm)	VC	Precision
ResU-Net	Bladder	95.7 ± 3.7	4.05 ± 5.17	2.30 ± 3.37	1.04 ± 0.97	0.976 ± 0.025	0.957 ± 0.065
	Rectum	96.6 ± 1.5	1.96 ± 2.19	1.42 ± 1.41	0.45 ± 0.09	0.975 ± 0.014	0.968 ± 0.036
	Sigmoid	93.0 ± 3.3	3.15 ± 2.03	2.10 ± 1.36	0.79 ± 0.25	0.963 ± 0.018	0.934 ± 0.039
Unet	Bladder	89.9 ± 8.3	5.06 ± 4.86	3.21 ± 3.11	2.20 ± 1.14	0.913 ± 0.033	0.890 ± 0.039
	Rectum	91.7 ± 6.3	2.14 ± 2.45	1.64 ± 1.22	0.62 ± 0.24	0.909 ± 0.032	0.909 ± 0.021
	Sigmoid	88.8 ± 7.9	3.58 ± 2.78	2.57 ± 1.66	1.40 ± 0.31	0.851 ± 0.023	0.847 ± 0.026

DSC = Dice similarity coefficient, HD = Hausdorff distance, $HD_{95\%}$ = Hausdorff distance 95th-percentile, ASSD = average symmetric surface distance, VC = volume correlation. There were no statistically differences between two models for all parameters.

Table 2
Results of quantitative parameters for the segmentation of the three organs. All values are reported as median, minimum, maximum, Q1, Q3.

Model	Organ	Quantitative parameters																	
		DSC (%)			HD (mm)			$HD_{95\%}$ (mm)			ASSD (mm)			VC			Precision		
		Median	Min, Max	Q1, Q3	Median	Min, Max	Q1, Q3	Median	Min, Max	Q1, Q3	Median	Min, Max	Q1, Q3	Median	Min, Max	Q1, Q3	Median	Min, Max	Q1, Q3
ResU-Net	Bladder	97.4	81.6, 98.4	94.4, 97.8	3.51	2.20, 10.47	2.51, 4.86	1.64	1.29, 9.34	1.42, 1.94	0.71	0.51, 3.73	0.61, 0.86	0.989	0.897, 0.993	0.968, 0.990	0.966	0.859, 0.978	0.958, 0.971
	Rectum	97.2	90.7, 98.3	95.8, 97.7	1.89	1.46, 2.81	1.72, 2.2	1.37	1.15, 1.88	1.20, 1.54	0.44	0.33, 0.65	0.40, 0.49	0.981	0.941, 0.988	0.965, 0.984	0.968	0.956, 0.978	0.963, 0.973
	Sigmoid	94.1	83.9, 96.9	91.4, 95.2	2.82	2.05, 5.80	2.54, 3.54	1.84	1.20, 4.10	1.67, 2.20	0.73	0.52, 1.54	0.68, 0.85	0.967	0.913, 0.982	0.961, 0.976	0.935	0.897, 0.961	0.923, 0.944
Unet	Bladder	91.0	81.0, 96.0	86.3, 93.0	3.88	2.23, 11.89	3.04, 6.48	2.63	1.44, 8.54	1.74, 4.28	1.78	0.52, 5.74	0.92, 3.37	0.915	0.850, 0.961	0.892, 0.940	0.905	0.822, 0.950	0.853, 0.920
	Rectum	92.4	85.5, 96.4	89.2, 93.5	1.95	1.46, 3.24	1.8, 2.42	1.53	1.11, 2.74	1.37, 1.86	0.48	0.35, 1.48	0.41, 0.75	0.916	0.850, 0.951	0.883, 0.938	0.910	0.873, 0.942	0.890, 0.920
	Sigmoid	88.8	83.6, 96.0	85.5, 91.8	3.49	2.29, 6.38	2.76, 4.25	2.40	1.21, 4.83	1.94, 2.99	1.50	0.68, 2.67	0.87, 1.74	0.852	0.810, 0.894	0.833, 0.863	0.845	0.810, 0.892	0.830, 0.868

DSC = Dice similarity coefficient, HD = Hausdorff distance, $HD_{95\%}$ = Hausdorff distance 95th-percentile, ASSD = average symmetric surface distance, VC = volume correlation, Min = minimum, Max = maximum. There were no statistically differences between two models for all parameters.

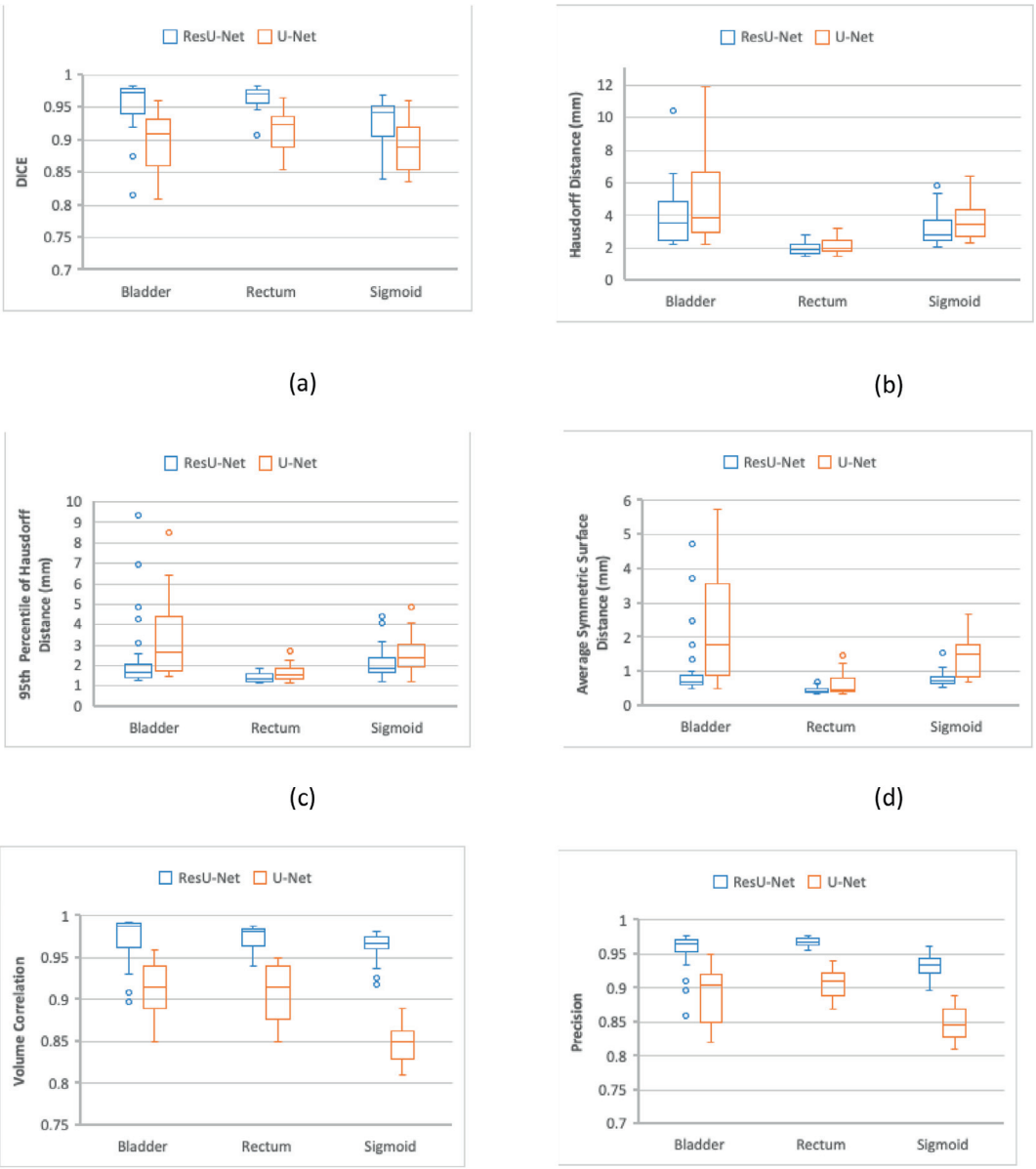


Fig. 2. Box plots of quantitative metrics for both ResU-Net and Unet models for the bladder, rectum and sigmoid. (a) DSC, (b) Hausdorff distance, (c) 95th percentile of Hausdorff distance, (d) Average symmetry surface distance, (e) Volume correlation and (f) precision.

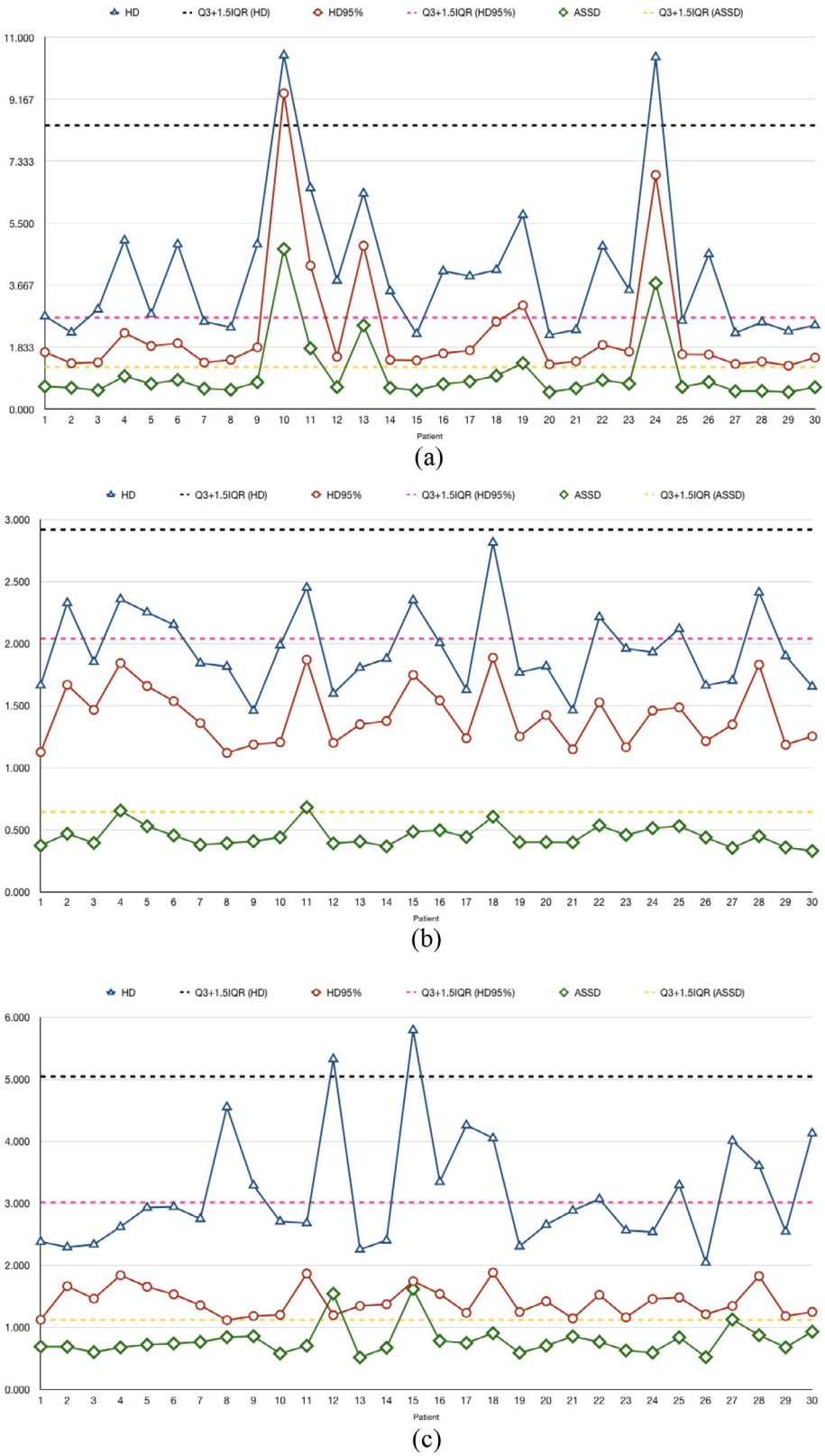
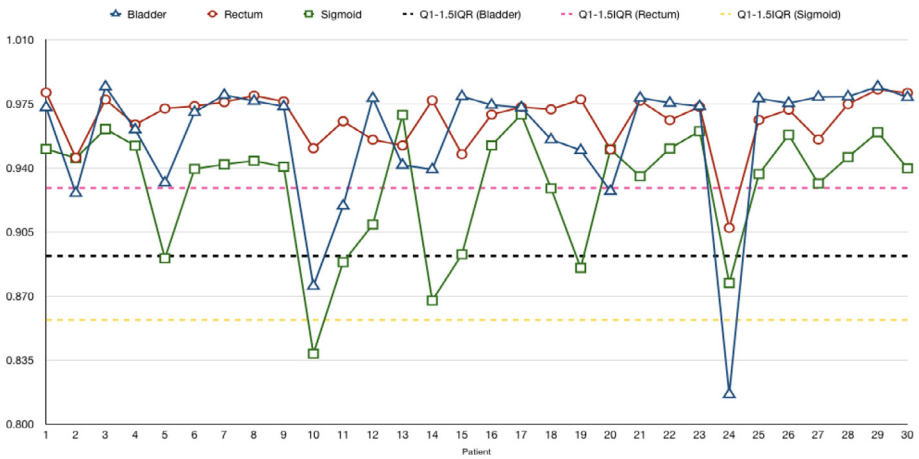
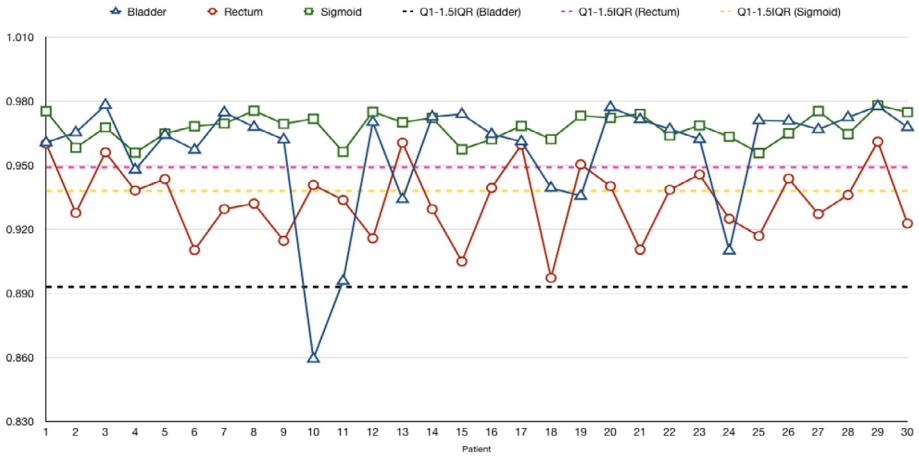


Fig. 3. Plots of HD, $HD_{95\%}$, ASSD, DSC, VC, and precision for individual patients to depict the outlier subjects from ResU-Net model. (a), HD, $HD_{95\%}$, ASSD for the bladder, (b) HD, $HD_{95\%}$, ASSD for the rectum, (c) HD, $HD_{95\%}$, ASSD for the sigmoid, (d) DSC for all OARs, (e) VC for all OARs and (f) precision for all OARs. DSC = Dice similarity coefficient, HD = Hausdorff distance, HD95% = Hausdorff distance 95th-percentile, ASSD = average symmetric surface distance, VC = volume correlation.

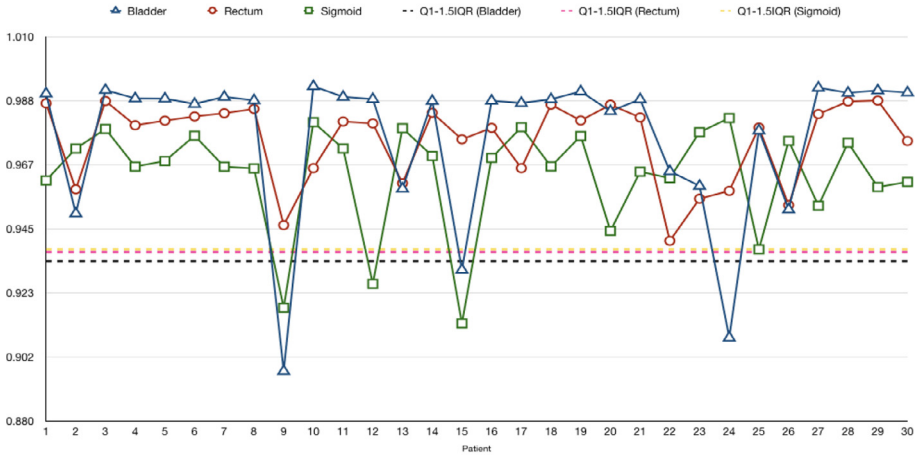
the boxplots are shown in Fig. 2. Overall, the predicted contours of the rectum showed relatively higher accuracy for both models. Outliers detected based on DSC, HD, $HD_{95\%}$, and ASSD parameters were summarized in supplemental Table 2, Fig. 3 and supplemental Fig. 3. An example of erroneously predicted contours of one patient for the bladder, rectum, and sigmoid are shown in Fig. 4.



(d)



(e)



(f)

Fig. 3 (*continued*)

The dosimetric accuracy of the ResU-Net and Unet models were compared to results obtained from manually defined contours used as reference (Table 3). The ResU-Net model achieved better results compared to Unet model for all dosimetric parameters. The differences between all dosimetric parameters for the sigmoid were statistically significant between the two models (p-value < 0.05). There is no proof of statistical difference for the bladder and rectum considering all dosimetric parameters.

Discussion

The delineation of OARs is an important task in brachytherapy of gynecological cancers. However, the process suffers from inter- and intra-observer variability in organ delineations when performed manually. The toxicity of OARs in brachytherapy of gynecological cancer can be accurately predicted through the precise delineation of OARs. Auto-contouring of OARs is highly desired

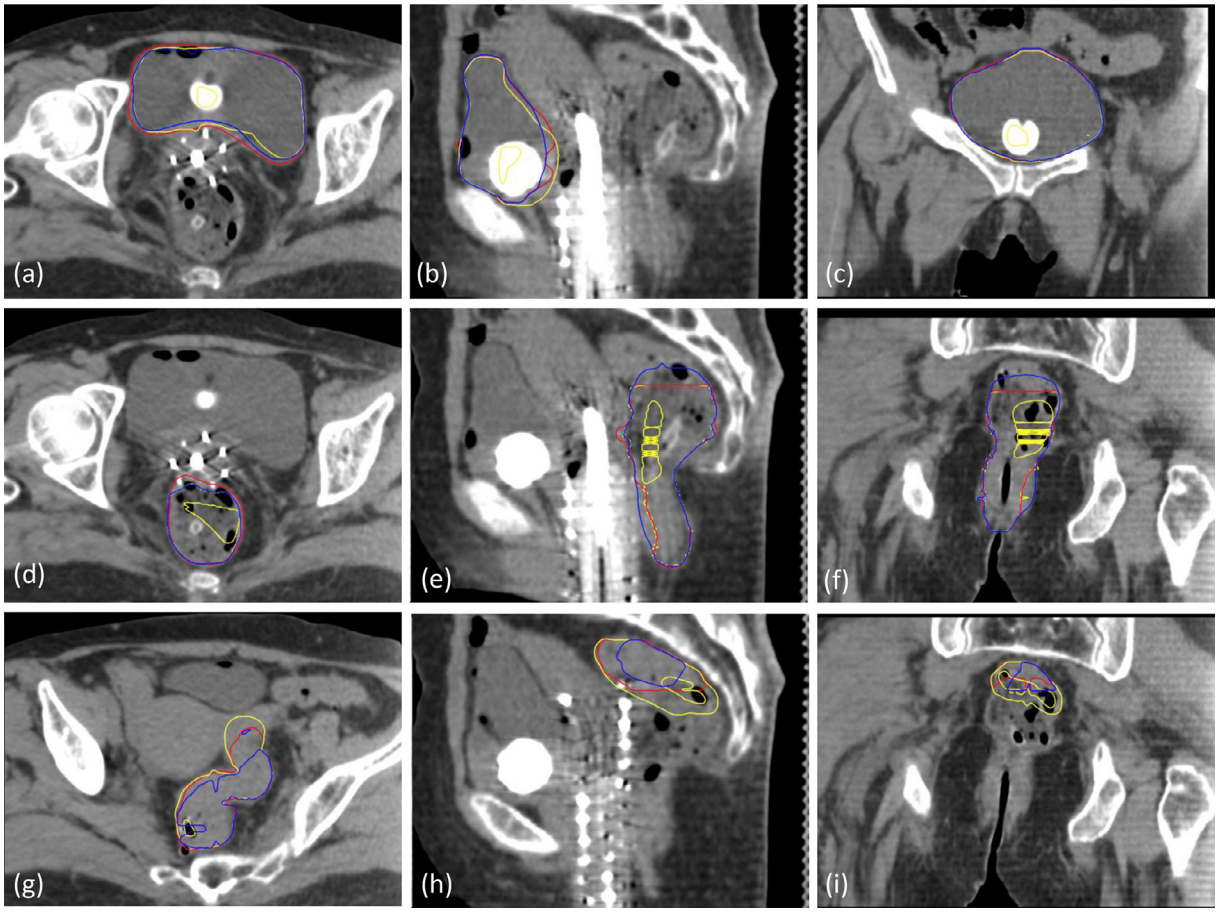


Fig. 4. Representative example of erroneously predicted contours by ResU-Net and Unet models for (a-c) the bladder, (d-f) the rectum and (g-i) the sigmoid. Blue color = predicted contour from ResU-Net, Yellow color = predicted contour from Unet, Red color = manual contour.

Table 3
Results of dosimetric parameters and volume differences for the segmentation of the three organs. All values are reported as mean ± SD.

Model	Quantitative parameters				
	Organ	ΔD_{5cc} (Gy)	ΔD_{2cc} (Gy)	ΔD_{1cc} (Gy)	ΔV (cm^3)
ResU-Net	Bladder	−0.383 ± 0.230	−0.499 ± 0.444	−0.757 ± 0.367	34.2 ± 7.5
	Rectum	0.194 ± 0.111	0.321 ± 0.255	0.467 ± 0.268	16.7 ± 8.8
	Sigmoid	−0.322 ± 0.311	−0.467 ± 0.266	−0.521 ± 0.264	5.2 ± 3.7
Unet	Bladder	0.498 ± 0.365	0.601 ± 0.211	0.888 ± 0.156	48.8 ± 10.1
	Rectum	0.226 ± 0.204	0.383 ± 0.295	0.528 ± 0.244	20.5 ± 5.9
	Sigmoid	−0.489 ± 0.289	−0.622 ± 0.198	−0.789 ± 0.236	8.4 ± 3.8

There were statistically differences between the two models only for the sigmoid (p-value < 0.03).

in brachytherapy owing to the limited time between implantation and treatment of patients and the need to minimize the variability in organ delineations to avoid unwanted toxicity of OARs. Accurate delineation of OARs allows for dose escalation in the target volumes. This study set out to employ a deep convolutional neural network for simultaneous multi-organ contouring in gynecological brachytherapy. We realized that the contours predicted using ResU-Net model had high overlap/agreement with manual contours delineated by radiation oncologists and had better performance and dosimetric accuracy compared to the standard Unet model.

We observed that our ResU-Net model predicted OARs with a DSC of 95.7% for the bladder, 96.6% for the rectum, and 93.0% for the sigmoid while Unet achieved DSC of 89.9%, 91.7% and 88.8%, respectively. Among all OARs, the rectum presented the highest DSC value and best performance for other quantitative metrics for both ResU-Net and Unet models. Looking at CT images of test patients, we found that the rectum had relatively constant volume and shape among the different patients. The shape and volume of the bladder were different between patients, more than the rectum and sigmoid. However, consistent with the observations made by Jamema et al. [46], the topographical changes in the sigmoid were much more varied than other OARs. The dosimetric parameters were calculated by performing planning on the auto-segmented contours derived from both ResU-Net and Unet models. A good agreement with the results obtained from the planning based on manual contouring was observed. The dosimetric evaluation revealed a similar trend to the one observed for segmentation,

wherein the proposed ResUnet exhibited superior performance over the conventional U-net model, leading to reduced estimated absorbed dose errors within the OARs.

Kazemifar et al. [47] and Balagopal et al. [48] investigated auto-segmentation of the prostate and OARs in male pelvic region using deep learning algorithms. They employed CT images of 85 and 136 subjects and 2D U-Net and 2D + 3D U-Net models, respectively, for automated segmentation of the bladder and rectum to achieve DSC values of 95%, 92% and 95%, 84%, respectively. In comparison to their results, our model showed a better DSC for the rectum while the DSC of the bladder was almost the same. This might be due to the presence of Foley catheter containing contrast agent. The difference in the amount of contrast agent mix with saline changed gray values of Foley catheter inside the bladder among the different patients. Balloon of Foley catheter must be fixed in the trigone area of the bladder after filling. Imperfect fixing causes diversity in balloon location, which might affect the results. Zhang et al. [40] used DSD-UNET and 3D UNET for automated segmentation of HR-CTV and OARs on 91 CT images of patients with cervical cancers. Our results showed superior performance considering the entire quantitative metrics including DSC, Jaccard, and HD OARs segmentation such as bladder, rectum, and sigmoid, though we used a 2D deep learning architecture. This might be due to the size of the training samples (size of dataset) for the training of the model. For 2D models, each 2D slice of CT scan is considered one data or training sample. As such, the number of training samples is dramatically higher in 2D training mode. However, in 3D training mode, each subject (involving the whole organ/structure/target in multiple 2D slices) is considered as a single training sample. In this regard, when the training sample size is limited, the 2D model would be a better option compared to 3D models which may perform suboptimally owing to the limited/insufficient training samples. In this light, the experiment conducted by Rigaud et al. [49] demonstrated that a 2D model trained with DeepLab V3+ to segment cervix-uterus, vagina, parametrium, bladder, rectum, sigmoid, femoral heads, kidneys, spinal cord, and bowel bag from cervical cancer patients performed comparably (and even better in some aspects) with respect to the same model trained in 3D mode. In addition, the 3D models may fail to reach peak performance due to the need for higher computational power and GPU memory to handle large volumetric data. In this light, the 3D model tends to down-sample the input images (either in the input layers or feature space) which may skew the optimal performance of the model. Moreover, 3D models require dramatically larger number of trainable parameters to properly encode the underlying features in volumetric data. A large number of trainable parameters complicates the process of model optimization and might increase the risk of overfitting. Thus, developing models in 3D mode is not necessarily superior to their 2D counterparts. Contraries results were observed in different studies when using 2D or 3D approaches [49–54]. These results might demonstrate that choice of 2D or 3D approaches is task-specific and depend on the size of the training dataset as well as the complexity of the segmentation problem.

Ayadi et al. [55] used ABAS in 26 prostate cancer patients. The DSCs for the bladder and rectum were 80% and 66%, which are substantially lower than those achieved in this work. This might be related to the inherent characteristics of the ABAS algorithm, which uses deformable image registration (DIR) to map structures from atlases to patient's CT images. Hernandez et al. [56] investigated the efficiency of auto-segmentation software in the contouring of OARs for brachytherapy of cervical cancers. They reported DSCs of 83%, 79%, and 52% for the bladder, rectum, and sigmoid, respectively. They also reported that patients with DSC values less than 90% had a median bladder volume about 10% smaller than the atlas dataset. A small DSC (~52%) was observed for the sigmoid due to the large topographic changes that the atlas-based method does not properly cover owing to the limited number of atlases [57]. Some studies reported limitations of performance of DIR for cervical cancer brachytherapy due to applicator artifacts, presence of gas in the rectum, presence of contrast agent in the bladder, large changes in bladder filling and rectum between patients and atlases and imperfection of algorithms, especially intensity-based registration [17,58–60].

We considered values above Q3 + 1.5IQR and below Q1-1.5IQR as outlier data (Fig. 3 and supplemental Table 2). Two patients (n = 10, 24) had the worst HD, $HD_{95\%}$, and ASSD for the bladder. One patient (n = 18) had the worst HD, $HD_{95\%}$, and ASSD for the rectum. Two patients (n = 12, 15) had the worst HD, $HD_{95\%}$, and ASSD for the sigmoid. A potential explanation of these observations is the gross errors in the predicted contour in one or few slices, an example of which has been shown in Fig. 4. supplemental Table 2 shows that choosing outlier data based on each performance metric is different. Besides, the Unet model exhibited fewer gross errors (number of outliers) in the predicted contours compared to the ResU-Net model, while the mean/median of all quantitative metrics were inferior to those of the ResU-Net model. Most studies reported on evaluation using one or few metrics (e.g. DSC, HD, and $HD_{95\%}$), yet the validation of deep learning-guided auto-contouring should include qualitative evaluation by experienced radiation oncologists. We observed that most outliers occurred because one or two slices had large discrepancy between predicted and manual contours, which gave rise to noticeable drops in some quantitative parameters.

This work inherently bears a number of limitations, including uncertainty in OARs contours delineation in the training dataset, the number of training and test datasets, single-center dataset and the ensuing limited variability, variation in the type of applicators (like tandem and ovoids, tandem and ring, tandem and ring with needles, etc.) and variation in cancer types. The diversity of cancer types, treatment volumes, CT scanner machines, image acquisition protocol, and standards in OARs contouring hampered meaningful comparison of our results with those reported in other studies. Overall, the deep learning algorithm exhibited superior performance compared to model-based, atlas-based, and intensity-based DIR approaches [34,61,62].

In this work, we investigated the performance of deep learning-assisted auto-contouring in gynecological cancer for patients treated with tandem and ovoids and multi-channel cylinders. The material of tandem and ovoids were titanium while the material of needles for multi-channel cylinders was plastic. Hence, CT markers (lead material) are needed to perform the reconstruction correctly. These CT markers amplify the artifacts visible on the images (Fig. 4). The adopted methodology may not be suitable for other situations where different techniques, such as interstitial and intracavitary plus interstitial are used for treating patients. It is more challenging for patients treated with interstitial brachytherapy since radiation oncologists commonly insert needles freehand and then CT markers are used to carry out CT-based planning and the reconstruction process correctly. It causes artifacts in the image, especially when a high number of needles are used. In addition, the location and number of needles vary between patients. Another challenge is related to the protocols used in the different centers. Some centers use a contrast agent inside the bladder instead of the balloon of Foley catheter to clear the boundary of the bladder. Other centers employ rectal tube or contrast agent inside the rectum. Therefore, a large dataset from different centers using different techniques can make our model more robust to improve performance further.

Conclusion

Delineation of OARs in brachytherapy is an important component of treatment planning that affects dose delivery to these organs, as well as the target volume. However, manual delineation is labor-intensive and time-consuming. In this work, a DCNN model for CT-based OARs delineation in locally-advanced brachytherapy planning was developed and evaluated. The performance of the DCNN was assessed taking manual contour delineated by an experienced radiation oncologist as standard of reference. The quantitative results demonstrated that the DCNN has a good agreement with the reference and could be used to segment the bladder, rectum, and sigmoid with high accuracy. However, outliers were observed in some cases. Yet, the DCCN can be used to improve reproducibility in brachytherapy contouring and speed-up the delineation of OARs.

Acknowledgements

This work was supported by the Swiss National Science Foundation under grant SNRF 320030_176052, the Eurostars programme of the European commission under grant E! 12326 ILLUMINUS and the Private Foundation of Geneva University Hospitals under Grant RC-06-01.

Disclosure

The authors have no conflict of interest to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2021.03.030.

References

[1] Saarnak AE, Boersma M, van Bunningen BNFM, Wolterink R, Steggerda MJ. Inter-observer variation in delineation of bladder and rectum contours for brachytherapy of cervical cancer. Radiother Oncol 2000;56:37–42. https://doi.org/10.1016/S0167-8140(00)00185-7.

[2] Haie-Meder C, Pötter R, Van Limbergen E, Briot E, De Brabandere M, Dimopoulos J, et al. Recommendations from Gynaecological (GYN) GEC-ESTRO Working Group☆(I): concepts and terms in 3D image based 3D treatment planning in cervix cancer brachytherapy with emphasis on MRI assessment of GTV and CTV. Radiother Oncol 2005;74:235–45. https://doi.org/10.1016/j.radonc.2004.12.015.

[3] Dimopoulos JCA, Petrow P, Tanderup K, Petric P, Berger D, Kirisits C, et al. Recommendations from Gynaecological (GYN) GEC-ESTRO Working Group (IV): Basic principles and parameters for MR imaging within the frame of image based adaptive cervix cancer brachytherapy. Radiother Oncol 2012;103:113–22. https://doi.org/10.1016/j.radonc.2011.12.024.

[4] Pötter R, Haie-Meder C, Limbergen EV, Barillot I, Brabandere MD, Dimopoulos J, et al. Recommendations from gynaecological (GYN) GEC ESTRO working group (II): concepts and terms in 3D image-based treatment planning in cervix cancer brachytherapy-3D dose volume parameters and aspects of 3D image-based anatomy, radiation physics, radiobiology. Radiother Oncol 2006;78:67–77. https://doi.org/10.1016/j.radonc.2005.11.014.

[5] Hellebust TP, Kirisits C, Berger D, Pérez-Calatayud J, De Brabandere M, De Leeuw A, et al. Recommendations from Gynaecological (GYN) GEC-ESTRO Working Group: considerations and pitfalls in commissioning and applicator reconstruction in 3D image-based treatment planning of cervix cancer brachytherapy. Radiother Oncol 2010;96:153–60. https://doi.org/10.1016/j.radonc.2010.06.004.

[6] Swamidas J, Mahantshetty U. ICRU report 89: prescribing, recording, and reporting brachytherapy for cancer of the cervix. J Med Phys 2017;42:48. https://doi.org/10.1093/jicru/ndw027.

[7] Hellebust TP, Dale E, Skjønsberg A, Olsen DR. Inter fraction variations in rectum and bladder volumes and dose distributions during high dose rate brachytherapy treatment of the uterine cervix investigated by repetitive CT-examinations. Radiother Oncol 2001;60:273–80. https://doi.org/10.1016/S0167-8140(01)00386-3.

[8] Duane FK, Langan B, Gillham C, Walsh L, Rangaswamy G, Lyons C, et al. Impact of delineation uncertainties on dose to organs at risk in CT-guided intracavitary brachytherapy. Brachytherapy 2014;13:210–8. https://doi.org/10.1016/j.brachy.2013.08.010.

[9] Voet PWJ, Dirkx MLP, Teguh DN, Hoogeman MS, Levendag PC, Heijmen BJM. Does atlas-based autosegmentation of neck levels require subsequent manual contour editing to avoid risk of severe target underdosage? A dosimetric analysis. Radiother Oncol 2011;98:373–7. https://doi.org/10.1016/j.radonc.2010.11.017.

[10] Anders LC, Stieler F, Siebenlist K, Schäfer J, Lohr F, Wenz F. Performance of an atlas-based autosegmentation software for delineation of target volumes for radiotherapy of breast and anorectal cancer. Radiother Oncol 2012;102:68–73. https://doi.org/10.1016/j.radonc.2011.08.043.

[11] Daisne, J.-F. and A. Blumhofer, Atlas-based automatic segmentation of head and neck organs at risk and nodal target volumes: a clinical validation. Radiat Oncol (London, England), 2013. 8: 154-154. DOI: https://doi.org/10.1186/1748-717x-8-154.

[12] Teguh DN, Levendag PC, Voet PWJ, 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.

[13] Arabi H, Koutsouvelis N, Rouzaud M, Miralbell R, Zaidi H. Atlas-guided generation of pseudo-CT images for MRI-only and hybrid PET-MRI-guided radiotherapy treatment planning. Phys Med Biol 2016;61:6531–52. https://doi.org/10.1088/0031-9155/61/17/6531.

[14] Arabi H, Dowling JA, Burgos N, et al. Comparative study of algorithms for synthetic CT generation from MRI: Consequences for MRI-guided radiation planning in the pelvic region. Med Phys 2018;45:5218–33. https://doi.org/10.1002/mp.13187.

[15] Yeo UJ, Supple JR, Taylor ML, Smith R, Kron T, Franich RD. Performance of 12 DIR algorithms in low-contrast regions for mass and density conserving deformation. Med Phys 2013;40:101701. https://doi.org/10.1118/1.4819945.

[16] Zhong H, Kim J, Chetty IJJMp. Analysis of deformable image registration accuracy using computational modeling. Med Phys 2010;37:970–9. https://doi.org/10.1118/1.3302141.

[17] Mohammadi R, Mahdavi S, Jaberi R, Siavashpour Z, Janani L, Meigooni A, et al. Evaluation of deformable image registration algorithm for determination of accumulated dose for brachytherapy of cervical cancer patients. J Contemp Brachyther 2019;11:469–78. https://doi.org/10.5114/jcb.2019.88762.

[18] Greenham S, Dean J, Fu CKK, Goman J, Mulligan J, Tune D, 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.

[19] 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/10.1016/j.ejmp.2015.10.020.

[20] Van de Velde J, Wouters J, Vercauteren T, De Gersem W, Achten E, De Neve W, et al. Optimal number of atlases and label fusion for automatic multi-atlas-based brachial plexus contouring in radiotherapy treatment planning. Radiat Oncol 2016;11. https://doi.org/10.1186/s13014-015-0579-1.

[21] Arabi H, Zaidi H. One registration multi-atlas-based pseudo-CT generation for attenuation correction in PET/MRI. Eur J Nucl Med Mol Imaging 2016;43:2021–35. https://doi.org/10.1007/s00259-016-3422-5.

[22] Teguh DN, Levendag PC, Voet PWJ, 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.

[23] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 2017;60:84–90. https://doi.org/10.1145/3065386.

[24] Chen H, Qi X, Yu L, et al. DCAN: deep contour-aware networks for accurate gland segmentation. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. https://doi.org/10.1016/j.media.2016.11.004.

[25] Liu Y, Stojadinovic S, Hrycushko B, Wardak Z, Lau S, Lu W, et al. A deep convolutional neural network-based automatic delineation strategy for multiple brain metastases stereotactic radiosurgery. PLoS ONE 2017;12:e0185844. https://doi.org/10.1371/journal.pone.0185844.

[26] de Vos BD, Berendsen FF, Viergever MA, et al. End-to-end unsupervised deformable image registration with a convolutional neural network. In: Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer; 2017. p. 204–12.

[27] Chougrad H, Zouaki H, Alheyane OJCm, et al., Deep convolutional neural networks for breast cancer screening, 2018; 157: 19-30. DOI: https://doi.org/10.1016/j.cmpb.2018.01.011.

[28] Moradi S, Oghli MG, Alizadehasl A, Shiri I, Oveisi N, Oveisi M, et al. MFP-Unet: A novel deep learning based approach for left ventricle segmentation in echocardiography. Phys Med 2019;67:58–69. https://doi.org/10.1016/j.ejmp.2019.10.001.

[29] Shiri I, Akhavanallaf A, Sanaat A, Salimi Y, Askari D, Mansouri Z, et al. Ultra-low-dose chest CT imaging of COVID-19 patients using a deep residual neural network. Eur Radiol 2021;31:1420–31. https://doi.org/10.1007/s00330-020-07225-6.

[30] Shiri I, Arabi H, Geramifar P, Hajianfar G, Ghafarian P, Rahmim A, et al. Deep-JASC: joint attenuation and scatter correction in whole-body 18 F-FDG PET using a deep residual network. Eur J Nucl Med Mol Imaging 2020;47:2533–48. https://doi.org/10.1007/s00259-020-04852-5.

[31] Shiri I, Sabet KA, Arabi H, et al., Standard SPECT myocardial perfusion estimation from half-time acquisitions using deep convolutional residual neural networks. J Nucl Cardiol, 2020. DOI: https://doi.org/10.1007/s12350-020-02119-y.
[32] Akhavanallaf A, Shiri I, Arabi H, et al. Whole-body voxel-based internal dosimetry using deep learning. Eur J Nucl Med Mol Imaging 2020:1–13. https://doi.org/10.1007/s00259-020-05013-4.
[33] Hu Y, Xie C, Yang H, et al. Computed tomography-based deep-learning prediction of neoadjuvant chemoradiotherapy treatment response in esophageal squamous cell carcinoma. Radiother Oncol 2021;154:6–13. https://doi.org/10.1016/j.radonc.2020.09.014.
[34] van Dijk LV, Van den Bosch L, Aljabar P, et al. Improving automatic delineation for head and neck organs at risk by Deep Learning Contouring. Radiother Oncol 2020;142:115–23. https://doi.org/10.1016/j.radonc.2019.09.022.
[35] Zhang L, Dong D, Zhang W, et al. A deep learning risk prediction model for overall survival in patients with gastric cancer: A multicenter study. Radiother Oncol 2020;150:73–80. https://doi.org/10.1016/j.radonc.2020.06.010.
[36] Xiao H, Ren G, Cai J. A review on 3D deformable image registration and its application in dose warping. Radiat Med Protect 2020;1:171–8. https://doi.org/10.1016/j.radmp.2020.11.002.
[37] Sharp G, Fritscher KD, Pekar V, et al. Vision 20/20: perspectives on automated image segmentation for radiotherapy. Med Phys 2014;41:050902. https://doi.org/10.1118/1.4871620.
[38] Stapleford LJ, Lawson JD, Perkins C, et al. Evaluation of automatic atlas-based lymph node segmentation for head-and-neck cancer. Int J Radiat Oncol Biol Phys 2010;77:959–66. https://doi.org/10.1016/j.ijrobp.2009.09.023.
[39] van Baardwijk A, Bosmans G, Boersma L, et al. Pet-ct–based auto-contouring in non–small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumor and involved nodal volumes. Int J Radiat Oncol Biol Phys 2007;68:771–8. https://doi.org/10.1016/j.ijrobp.2006.12.067.
[40] Zhang D, Yang Z, Jiang S, et al. Automatic segmentation and applicator reconstruction for CT-based brachytherapy of cervical cancer using 3D convolutional neural networks. J Appl Clin Med Phys 2020;21:158–69. https://doi.org/10.1002/acm2.v21.1010.1002/acm2.13024.
[41] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. https://doi.org/10.1109/CVPR.2016.90.
[42] Ronneberger O, Fischer P, Brox T, U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention. 2015. Springer. DOI: https://doi.org/10.1007/978-3-319-24574-4_28.
[43] Lorensen WE, Cline HE. Marching cubes: A high resolution 3D surface construction algorithm. Comput Graph Interact Tech 1987;21:163–9. https://doi.org/10.1145/37401.37422.
[44] Arabi H, Zaidi H. Applications of artificial intelligence and deep learning in molecular imaging and radiotherapy. Euro J Hybrid Imaging 2020;4:17. https://doi.org/10.1186/s41824-020-00086-8.
[45] Bahrami A, Karimian A, Fatemizadeh E, Arabi H, Zaidi H. A new deep convolutional neural network design with efficient learning capability: Application to CT image synthesis from MRI. Med Phys 2020;47:5158–71. https://doi.org/10.1002/mp.v47.1010.1002/mp.14418.
[46] Jamema SV, Mahantshetty U, Tanderup K, et al. Inter-application variation of dose and spatial location of D(2cm(3)) volumes of OARs during MR image based cervix brachytherapy. Radiother Oncol 2013;107:58–62. https://doi.org/10.1016/j.radonc.2013.01.011.
[47] Kazemifar S, Balagopal A, Nguyen D, et al. Segmentation of the prostate and organs at risk in male pelvic CT images using deep learning. Biomed Phys Eng Express 2018;4:. https://doi.org/10.1016/j.media.2019.03.003055003.
[48] Balagopal A, Kazemifar S, Nguyen D, Lin M-H, Hannan R, Owrangi A, 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.
[49] Rigaud B, Anderson BM, Yu ZH, Gobeli M, Cazoulat G, Söderberg J, et al. Automatic segmentation using deep learning to enable online dose optimization during adaptive radiation therapy of cervical cancer. Int J Radiat Oncol Biol Phys 2021;109:1096–110. https://doi.org/10.1016/j.ijrobp.2020.10.038.
[50] Sanders JW, Lewis GD, Thames HD, Kudchadker RJ, Venkatesan AM, Bruno TL, et al. Machine segmentation of pelvic anatomy in MRI-assisted radiosurgery (MARS) for prostate cancer brachytherapy. Int J Radiat Oncol Biol Phys 2020;108:1292–303. https://doi.org/10.1016/j.ijrobp.2020.06.076.
[51] Zhou X, Yamada K, Kojima T, et al. Performance evaluation of 2D and 3D deep learning approaches for automatic segmentation of multiple organs on CT images. Med Imaging 2018. https://doi.org/10.1117/12.2295178.
[52] Ma X, Hadjiiski LM, Wei J, et al. U-Net based deep learning bladder segmentation in CT urography. Med Phys 2019;46:1752–65. https://doi.org/10.1002/mp.13438.
[53] Zhou Y, Huang W, Dong P, et al., D-UNet: a dimension-fusion U shape network for chronic stroke lesion segmentation. IEEE/ACM transactions on computational biology and bioinformatics, 2019: p. 1-1. DOI: https://doi.org/10.1109/TCBB.2019.2939522
[54] Orlando N, Gillies DJ, Gyacskov I, Romagnoli C, D'Souza D, Fenster A. Automatic prostate segmentation using deep learning on clinically diverse 3D transrectal ultrasound images. Med Phys 2020;47:2413–26. https://doi.org/10.1002/mp.v47.610.1002/mp.14134.
[55] Ayadi M, Delpon G, Lafay F, et al. Evaluation of ABAS^{TM} : multi-center study in the case of prostate cancer. Phys Med 2011;27:S14–5. https://doi.org/10.1016/j.ejmp.2011.06.032.
[56] Hernandez S, Hoggarth MA, Albuquerque K, Comsia N, Quinn M, Roeske JC. An analysis of the efficiency of autosegmentation software in contouring normal tissues during 3-dimensional image-based brachytherapy planning. Brachytherapy 2010;9:S67–8. https://doi.org/10.1016/j.brachy.2010.02.107.
[57] Swamidas J, Kirisits C, De Brabandere M, Hellebust TP, Siebert F-A, Tanderup K. Image registration, contour propagation and dose accumulation of external beam and brachytherapy in gynecological radiotherapy. Radiother Oncol 2020;143:1–11. https://doi.org/10.1016/j.radonc.2019.08.023.
[58] Hayashi K, Isohashi F, Akino Y, et al., Estimation of the total rectal dose of radical external beam and intracavitary radiotherapy for uterine cervical cancer using the deformable image registration method. J Radiat Res, 2015. 56 (3): p. 546-552. DOI: https://doi.org/10.1093/jrr/rru127.
[59] Jamema SV, Mahantshetty U, Andersen E, Noe KØ, Sørensen TS, Kallehauge JF, et al. Uncertainties of deformable image registration for dose accumulation of high-dose regions in bladder and rectum in locally advanced cervical cancer. Brachytherapy 2015;14:953–62. https://doi.org/10.1016/j.brachy.2015.08.011.
[60] Flower E, Do V, Sykes J, Dempsey C, Holloway L, Summerhayes K, et al. Deformable image registration for cervical cancer brachytherapy dose accumulation: Organ at risk dose–volume histogram parameter reproducibility and anatomic position stability. Brachytherapy 2017;16:387–92. https://doi.org/10.1016/j.brachy.2016.12.006.
[61] Wong J, Fong A, McVicar N, Smith S, Giambattista J, Wells D, et al. Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning. Radiother Oncol 2020;144:152–8. https://doi.org/10.1016/j.radonc.2019.10.019.
[62] Zhu J, Zhang J, Qiu B, Liu Y, Liu X, Chen L. Comparison of the automatic segmentation of multiple organs at risk in CT images of lung cancer between deep convolutional neural network-based and atlas-based techniques. Acta Oncol 2019;58:257–64. https://doi.org/10.1080/0284186X.2018.1529421.