Deep CNN-based Fully Automated Segmentation of Pelvic Multi-Organ on CT Images for Prostate Cancer Radiotherapy

Bahram Mofid (MD)[1]ⓘ, Sayed Mohammad Modarres Mosalla (MD)[2], Masumeh Goodarzi (PhD)[3], Hassan Tavakoli (PhD)[3,4]*ⓘ

[1]Department of Radiation Oncology, Shohadae-Tajrish Medical Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[2]Department of Nuclear Medicine, Baqiyatallah University of Medical Sciences, Tehran, Iran
[3]Radiation Injuries Research Center, Systems Biology and Poisonings Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran
[4]Department of Physiology and Medical Physics, Baqiyatallah University of Medical Sciences, Tehran, Iran

ABSTRACT

Background: Manual delineation of volumes for prostate radiotherapy treatment is a time-consuming task for radiation oncologists and is also prone to variability. Deep learning-based auto-segmentation methods showed promising results with accurate and high-fidelity contours.

Objective: The objective of this study was to evaluate the feasibility of a Computed Tomography (CT)-based deep learning auto-segmentation algorithm for multi-organ delineation in prostate radiotherapy.

Material and Methods: In this single-institution retrospective study, a total of 118 patients with prostate cancer were included. We applied 3D nnU-net deep convolutional neural network architecture, a self-adapting ensemble method for simultaneous fast and reproducible multi-organ auto-contouring. The dataset was randomly divided into training and test sets from 95 and 23 patients, respectively. Intensity-modulated radiotherapy plans were generated for both manual and automatic delineations using identical optimization settings. Contours were assessed in terms of the Dice Similarity Coefficient (DSC), and average Hausdorff Distance (HD). Dose distributions were additionally evaluated using parameters derived from Dose-Volume Histograms (DVH).

Results: On the test set, 3D nnU-net achieved the best performance in the bladder (DSC:0.97, HD:4.13), right femur head (DSC:0.96, HD:3.58), left femur head (DSC:0.96, HD:3.95), rectum (DSC:0.9, HD:10.04), prostate (DSC:0.82, HD:3.68), lymph nodes (DSC:0.77, HD:15.5), and seminal vesicles (DSC:0.69, HD:10.95). DVH parameters of targets and Organ at Risks (OARs) were significantly different except for lymph nodes and femoral heads between treatment plans based on manual and automatic contours.

Conclusion: The 3D nnU-net architecture can be successfully used for multi-organ segmentation in the male pelvic area.

Introduction

Prostate cancer is the second most commonly identified form of cancer among men worldwide [1, 2]. Radiation Therapy (RT) is considered an integral component of the modern multidisciplinary approach to the management of prostate cancer [3]. Radiation therapy treatment aims to deliver a highly conformal radiation dose to the tumor, while at the same time sparing normal surrounding tissues (i.e.,

*Corresponding author:
Hassan Tavakoli
Radiation Injuries Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran
E-mail:
tavakoli@bmsu.ac.ir

Organs at Risk (OARs)), therefore minimizing acute and late radiation-induced toxicities [4]. During the RT treatment planning process, a prominent and critical step is an accurate and precise delineation of the target and OARs on the patient's Computed Tomography (CT) images [5]. In current clinical practice, manual segmentation performed by radiation oncologists or experienced planners is considered the gold standard, which is a tedious and time-consuming procedure [6, 7]. Furthermore, for prostate cancer, inter-and intra-rater variation in delineating target and OARs has been well documented due to differences in the level of expertise and preferences of the physicians over the past decades [8, 9]. Besides, in the pelvic region, using a simulation CT image for manual contouring Regions of Interest (ROIs) is challenging because of the poor soft tissue contrast of CT images.

Over the past decade, more sophisticated, advanced, and innovative radiation treatment technologies, including Volumetric- or Intensity Modulated Radiation Therapy (VMAT, IMRT), Stereotactic Body Radiotherapy (SBRT), and proton beam therapy have been widely adopted in clinical practice [10]. The use of these RT planning and delivery approaches leads to a decrease in the volume of normal tissues receiving moderate to high radiation doses, resulting in a reduction of radiation-induced side effects. Hence, the sharp dose gradient created by these techniques requires accurate delineation of target volumes and OARs to prevent large geometric misses [11]. It has been reported that even on-board imaging systems cannot eliminate systematic delineation errors [12]. As a consequence, it is needed to have accurate and precise contouring of the target and OARs.

Some research groups have attempted to develop efficient auto-segmentation tools in radiation oncology [13-16]. Automated medical image segmentation techniques, such as multi-atlas-based and hybrid methods have been previously considered state-of-the-art

[17]. More recently, with the rapid advances in Artificial Intelligence (AI), in particular deep learning algorithms, such as Convolutional Neural Networks (CNNs), a new generation of auto-segmentation models has been developed based on deep learning [18]. Deep CNN-based segmentation algorithms have gained more attention and resulted in improving the consistency and efficiency of contouring ROIs [19, 20]. To date, deep CNN models have been utilized for various tasks in medical imaging, including image registration, auto-segmentation, and classification [21-27]. Recently, deep learning-based auto-segmentation models outperformed older atlas-based methods due to their ability to learn complex sets of image features for their accurate performance of pixel-wise classification of images [28, 29].

The majority of previous studies used geometric accuracy parameters, such as the Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) to evaluate the performance of auto-segmentation approaches [30]. However, the accuracy of dose measurement and the quality of treatment plan quality, achieved based on the automatically generated contours, are very important in clinical practice [31]. To date, a limited number of studies have evaluated the dosimetric effect of CT organ segmentations for prostate cancer patients achieved from deep CNNs [13].

Therefore, the purpose of the current study was to evaluate the feasibility of a CT-based deep learning auto-segmentation algorithm for both OARs and target volume in radiotherapy treatment planning for prostate cancer. Geometric and Dose-Volume Histogram (DVH) metrics were applied to assess delineation accuracy between the auto-generated contours and the ground truths of clinicians' contouring as the standard of reference.

Material and Methods

Dataset
In this retrospective study, 118 prostate-

cancer patients participated, who received IMRT at Shohada-e-Tajrish Educational Hospital, Tehran, Iran between December 2021 and April 2023. Herein, patients with RT planning for prostate cancer were included, and those with prostatectomy and femoral implants were excluded. Raw CT scan images were used to train and test the proposed model. All CT images for the treatment planning were acquired using a 16-slice Siemens SOMATOM Sensation scanner (Siemens Medical Systems, Erlangen, Germany) at 110 KeV voltage. All images were obtained with a 512×512 matrix size and 3-mm slice thickness. In this study, the dataset was randomly split into 80% for training and 20% for testing the proposed model.

The target organs were the prostate, seminal vesicles, and lymph nodes and OARs included the rectum, bladder, and femoral heads. All contours were delineated by an experienced radiation oncologist with more than 25 years of experience in prostate radiotherapy according to international guidelines and recommendations. Figure 1 shows the flowchart of the proposed methodology for multi-organ segmentation.

Preprocessing and data augmentation

In this study, several image preprocessing techniques were performed on collected CT images prior to network training. Initially, using the 3D-Slicer software, images, and segmentations for each patient were converted from the DICOM RT structure format into binary masks, utilized during the model training. Row CT scan images were cropped to the body contour to decrease the computational cost. Also, the image values were normalized to fit into a range of 0 to 1. Moreover, aggressive data augmentation techniques were employed to augment the training dataset, resulting in an increased number of data samples and the improvement of classification accuracy and the model's generalization ability.
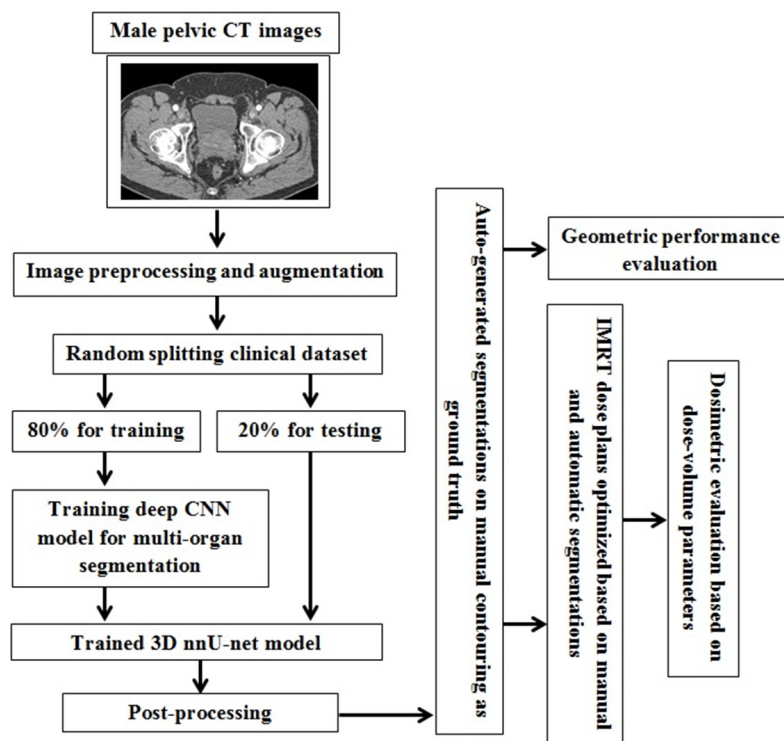
Figure 1: Flowchart of the proposed methodology for multi-organ segmentation (CT: Computed tomography; CNN: Convolutional neural network; IMRT: Intensity-modulated radiation therapy)

In addition, data augmentation techniques can reduce overfitting. Herein, data augmentation techniques were applied with a rotation range of 10 degrees, a zoom range of 0.1, a width shift range of 0.5, a height shift range of 0.5, horizontal/vertical filliping, scaling, brightness, and adding noise.

Deep CNN segmentation model

In the present study, the "no new U-net" or nnU-net algorithm, a deep learning CNN method, previously developed for auto-segmentation tasks in biomedical imaging, was adopted [32]. The architecture template of nnU-net follows a three-dimensional (3D) U-Net-like pattern with an encoder-decoder with skip connections. The nnU-net generates three various architectures based on the U-net backbone: a 2D U-net, a 3D U-net operating at full-image resolution, and a 3D U-net cascade network (3D-Cascade). The 3D-Cascade network consists of two U-nets: the first 3D U-net generates coarse segmentation maps on down-sampled images, and the second 3D U-net operates on full-resolution images to refine the segmentation map generated by the first one. The input channel size was selected 256×256×7 (7 i.e., prostate, seminal vesicle, lymph nodes, left femur head, right femur head, rectum, and bladder). To start with, the number of convolutional kernels was set to 16 in our configuration, which was doubled with each down-sampling up to a maximum of 320. The number of kernels in the decoder was set to mirror the number in the encoder. ReLu was used as the activation layer except for the final layer, for which we applied softmax.

Training details

The dataset consisting of CT scans from 118 prostate cancer patients was randomly divided into a training set from 95 patients and a test set from 23 patients. The nnU-net was trained on a training set along with the corresponding manually delineated contours by the radiation oncologist as the reference standard for learning. In this study, the proposed deep CNN model was trained using a Dice coefficient loss function and Adaptive Moment Estimation (Adam) optimizer with an initial learning rate of 0.001, batch size of 4, and epoch value of 150. Additionally, batch normalization was used to train the proposed deep CNN model faster and more stable. The model was implemented using Python 3.7. The training of the network was performed in a standard PC with a GeForce GTX 8 GB NVIDIA and 32 GB RAM. The training time for 3D nnU-net was about 11 hours.

Post-processing

In this study, post-processing techniques were utilized to process the segmentation map (i.e., the output of the proposed network) generated by the model to refine and improve the segmentation result. We used several post-processing algorithms, such as morphological operations, connected component analysis, and smoothing. One common post-processing technique is called morphological operations, involving applying mathematical morphology operations on the segmentation map, such as erosion, dilation, opening, and closing. Morphological operations were applied as a post-processing technique to refine the segmentation output generated by the proposed model. Specifically, a closing operation was used, which involves first performing a dilation operation on the binary image (i.e., the segmentation output), followed by an erosion operation. The purpose of this operation is to fill in any gaps or holes within the segmented regions. Hence, it smooths out any irregularities or noise in the segmentation output. The findings show this closing operation after the initial segmentation step significantly improved the overall accuracy and robustness of our segmentation results for male pelvic organs, particularly for structures, such as the prostate gland, seminal vesicles, and bladder, which can have complex shapes and exhibit significant anatomical variation across patients.

In the present study, connected component analysis was used as a critical step in the post-processing of male pelvic organ segmentation. Connected component analysis is a technique that identifies and labels distinct regions or components within a binary image based on pixel connectivity. By applying this method to the segmentation output, the pelvic organs of interest were separated from the background and other non-relevant structures, leading to isolating and quantifying each organ's volume accurately, which is crucial for diagnostic or treatment planning purposes. Consequently, incorporating connected component analysis into our segmentation pipeline helped to improve the accuracy and consistency of our segmentation results across different patients. Additionally, a smoothing method was employed as part of the post-processing of our segmentation results. The smoothing method involves applying a low-pass filter or convolution kernel to the binary image output generated by our segmentation model. This operation serves to reduce noise and sharp edges, resulting in a more continuous and visually pleasing representation of the segmented organs. Incorporating this smoothing step into our pipeline helped to reduce false positive and false negative segmentation errors, particularly in regions, where the organ boundaries were less well-defined. Additionally, the smoothed segmentation output can improve subsequent analysis steps, such as surface rendering or volume quantification, which rely on accurate and smooth representations of the segmented structures. Then, the predicted NumPy arrays as the output of semantic segmentation nnU-net model were converted into DICOM images [33].

Treatment planning

In this study, the impact of automatically segmented contours on target and OARs dosimetry was also evaluated, and for all test cases (n=23 patients), IMRT treatment plans were generated using the Eclipse v.13.0 (Varian Medical System Inc, Palo Alto, CA, USA) Treatment Planning Software (TPS). Herein, a pair of treatment plans was created using the identical planning CT image, one based on the physician segmentation as the standard ground truth and one based on the 3D nnU-net segmentation. In both scenarios, a 7-millimeter Planning Target Volume (PTV) margin was used around the prostate and seminal vesicles. Manually and automatically delineated contours were subjected to the same optimization settings, which encompassed identical objectives and weights for the target and OARs. All IMRT plans were generated in the Varian Eclipse TPS, using the Anisotropic Analytical Algorithm (AAA) algorithm. All patients were treated with image-guided IMRT using a 6 MV photon beam (Varian Clinac 600C linear accelerator (Varian Medical Systems, Palo Alto, CA, USA)). A prescription dose of 54.6-74.5 Gy was considered. For the target (i.e., prostate, seminal vesicles, and pelvic lymph nodes), values of D_{mean}, D98%, D2%, and V95% were calculated. For OARs (i.e., rectum, bladder, and femoral heads), D_{mean}, D_{max}, D_{min}, and V50/65/70Gy were determined.

Performance evaluation

In the current study, 23 patients as the clinical test dataset were used to evaluate the segmentation performance of the nnU-net model. The performance of the proposed auto-segmentation model was evaluated with the DSC, Jaccard index (JI), HD, 95th percentile HD (95% HD), Average Symmetric Surface Distance (ASSD), precision, and recall (sensitivity). The DSC measures the overlap between predicted contours by deep CNN models and the ground truth contours, and its value ranges from 0 (no overlap) to 1 (perfect overlap). The JI determines the similarity of radiologist-drawn contours as ground truth and the auto-segmented contours. The HD is the maximum surface distance between the surfaces of two contours. Herein, due to the sensitivity of HD

to outliers, 95% HD was also computed, which is the 95th percentile of the distances between the surfaces of the contours. The ASSD is the average difference of all the distances between two contours. The ASSD value of 0 indicates perfect segmentation.

$$DSC(A,B) = \frac{2 \times |A \cap B|}{A + B} \qquad (1)$$

$$JI(A,B) = \frac{|A \cap B|}{|A \cup B|} \qquad (2)$$

$$ASSD(A,B) = \frac{1}{N_A + N_B}\left\{\sum_{a \in A} min(d(a,b)) + \sum_{b \in B} min(d(a,b))\right\} \qquad (3)$$

$$HD(A,B) = max(h(A,B), h(B,A)) \qquad (4)$$

$$95\%HD(A,B) = 95^{th} \; percentile \; of \; (h(A,B), h(B,A)) \qquad (5)$$

$$Precision = \frac{|A \cap B|}{|B|} \qquad (6)$$

$$Sensitivity = \frac{|A \cap B|}{|A|} \qquad (7)$$

Where, A and B indicate the manually delineated mask and the auto-segmented mask, respectively. Further, a and b are individual voxels for ground truth and predicted contours, respectively, and h(A, B) is the directed Hausdorff distance from A to B.

Statistical analysis

SPSS Statistics V.22.0 software (SPSS Inc., IBM, Chicago, USA) was used for statistical analysis. The normal distribution of data was investigated by the Kolmogorov–Smirnov test. The Wilcoxon signed-rank test was used to assess the statistical differences between DVH parameters for plans generated manually and the nnU-net delineated contour. Statistical significance was determined when the *P*-value was below 0.05 (*P*-value <0.05).

Results

The segmentation using the 3D nnU-net took an average of 100 s per input volume on the standard PC with a GeForce GTX 8 GB NVIDIA and 32 GB RAM. Figure 2 represents manually delineated contours as ground

| Axial | Coronal | Sagittal |
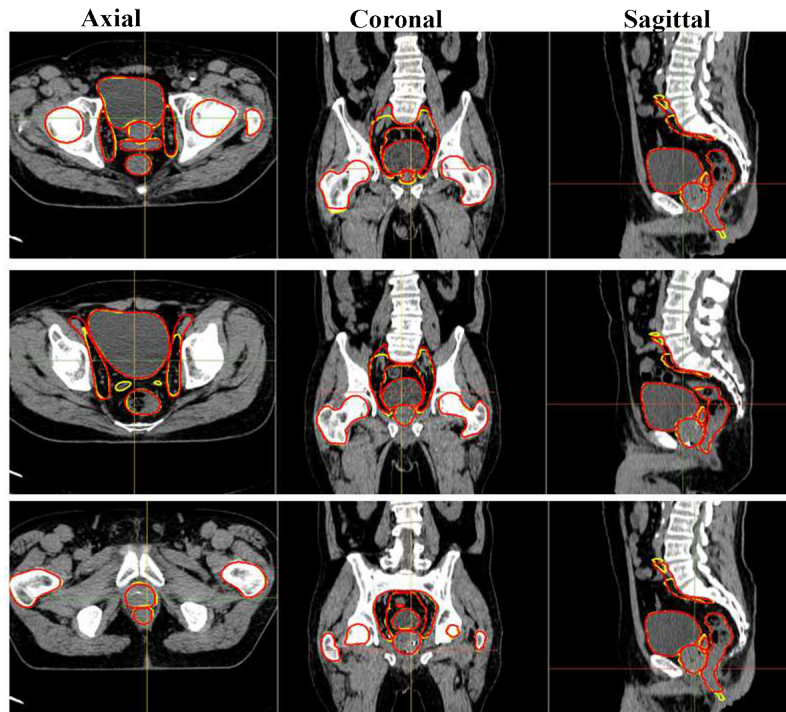
Figure 2: Axial, coronal, and sagittal slices of CT images for three test cases indicating radiation oncologist delineated contours as the ground truth (red) and automatically segmented contours by the nnU-net (yellow).

truth and automatically generated delineations of the prostate, seminal vesicle, lymph nodes, femoral heads, rectum, and bladder for three test patients. The comparison of organ volumes delineated by radiation oncologists as the ground truth and automatic segmentation based on our proposed deep CNN method is presented in Table 1. It is obvious that the volumes defined by the nnU-net model are similar to the volumes defined by manual segmentation as the standard ground truth. Table 1 shows a statistically significant difference in seminal vesicle volumes between manual and automatic delineation.

Table 2 summarizes the DSC, JI, HD, 95% HD, ASSD, precision, and sensitivity of the proposed model on the testing set. Mean

(standard deviation (SD)) DSCs of 0.82 (0.09), 0.69 (0.17), 0.77 (0.07), 0.90 (0.03), 0.97 (0.02), 0.96 (0.03), and 0.96 (0.01) were achieved for the prostate, seminal vesicle, lymph nodes, rectum, bladder, left femur head, and right femur head, respectively. As shown in Table 2, the highest performance was found for the bladder. As can be seen in Table 2, the nnU-net achieved a slightly worse performance for prostate, lymph nodes, and subsequently seminal vesicle. Figure 3 shows overall box and whisker plots of DSC, JI, ASSD, 95% HD, and HD for prostate, seminal vesicle, lymph nodes, rectum, bladder, left femur, and right femur.

Tables 3 and 4 outline the dosimetric accuracy of the proposed deep CNN-based segmentation approach in comparison with results achieved from manually delineated contours, as reference. Figure 4 shows the DVH of a test case for plans generated by the nnU-net model and ground truth contours. No statistically significant difference was observed for prostate and seminal vesicles considering all dose-volume parameters. There was a significant difference for the pelvic lymph nodes considering all dose-volume parameters (Table 3). There was no significant difference for OARs, except femoral heads considering all dose-volume parameters, as shown in Table 4.

Table 5 compares the performance of our

Table 1: Comparison of organ volumes between manual and automatic segmentations on the 23 testing patients (mean±SD)

Site	Ground truth	nnU-net	P-value
Prostate	27.5±7.4	26.1±5.6	0.249
Seminal vesicle	11.6±6.2	9.1±4.0	0.002
Lymph nodes	289.4±90.2	312.3±78.3	0.199
Rectum	69.5±23.4	66.8±24.6	0.108
Bladder	237.9±128.3	237.8±129.1	0.548
Left femur	175.5±29.6	174.9±30.1	0.884
Right femur	174.5±28.6	173.5±26.1	0.148

Table 2: Performance metrics of nnU-Net model for organ segmentation on testing set (mean±SD)

Site	DSC	Jaccard	ASSD	HD	95% HD	Precision	Sensitivity
Prostate	0.82±0.09	0.70±0.12	1.14±0.70	3.68±5.57	6.50±0.63	0.87±0.11	0.80±0.12
Seminal vesicle	0.69±0.17	0.57±0.16	2.05±2.72	10.95±19.68	8.06±18.68	0.79±0.17	0.63±0.19
Lymph nodes	0.77±0.07	0.64±0.08	2.76±1.03	15.5±7.79	4.93±4.73	0.71±0.09	0.87±0.05
Rectum	0.90±0.03	0.82±0.05	1.09±0.58	10.04±6.87	3.16±3.03	0.91±0.04	0.90±0.05
Bladder	0.97±0.02	0.94±0.04	0.40±0.16	4.13±8.60	1.99±7.16	0.97±0.02	0.96±0.02
Left femur	0.96±0.03	0.92±0.05	0.51±0.13	3.95±3.36	1.67±1.45	0.96±0.05	0.95±0.02
Right femur	0.96±0.01	0.93±0.02	0.58±0.03	3.85±2.61	1.53±1.29	0.96±0.01	0.97±0.01

DSC: Dice Similarity Coefficient; ASSD: Average Symmetric Surface Distance; HD: Hausdorff Distance
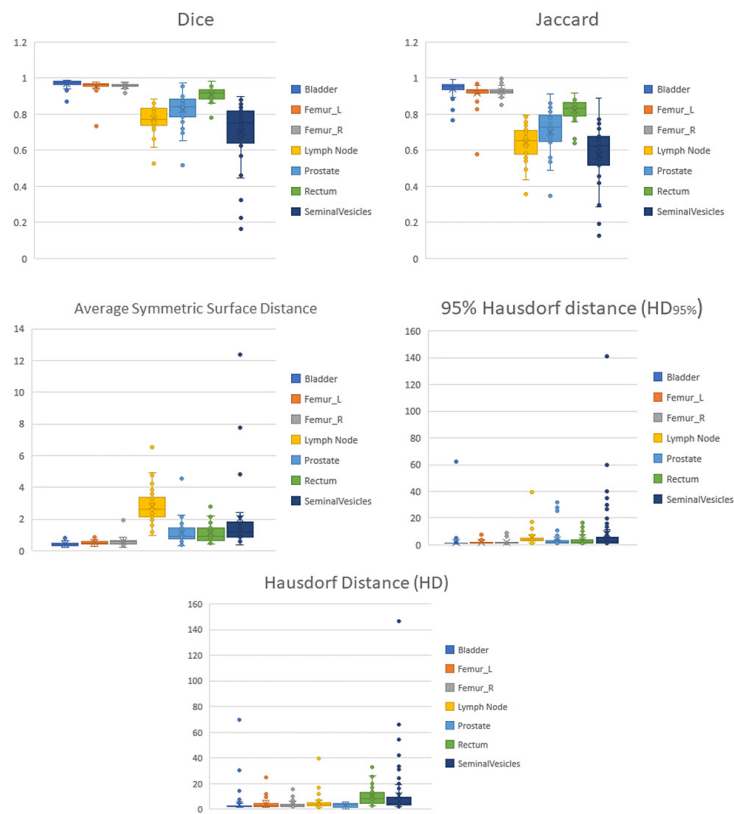
Figure 3: Box plots of quantitative metrics for nnU-Net model for prostate, seminal vesicle, lymph nodes, rectum, bladder, left femur, and right femur segmentation. In each panel, the bold line represents the median, the boxes represent the 25th and 75th percentiles, and whiskers represent ranges not including outliers. Individual point is considered as outlier.

Table 3: Comparison of the dose-volume parameters of target volumes between the two treatment plans optimized using nnU-net- and manually-delineated contours (mean±SD)

Organ	ΔD_{mean} (Gy)	$\Delta D_{98\%}$ (Gy)	$\Delta D_{2\%}$ (Gy)	$\Delta V_{95\%}$ (%)
Prostate	-0.025±0.269	-0.089±4.356	-0.007±0.134	-0.281±0.841
Pelvic lymph nodes	-1.573±2.923	-2.136±5.670	-1.133±3.191	-1.698±3.509
Seminal vesicles	-1.167±11.65	2.934±20.305	-4.633±15.282	-5.545±30.638

Bold values indicate that there is a statistically significant difference in lymph node dose-volume parameters between the manual segmentation (ground truth) and the automatically generated contours ($P<0.05$)

proposed model with other state-of-the-art models reported in the literature for pelvis organ segmentation on the CT images according to the DSC. Table 5 presents that our proposed architecture has comparable performance with other state-of-the-art models for multi-organ segmentation in the pelvic area.

Discussion

Accurate delineation of targets and OARs is known as a critical task in the radiotherapy of prostate cancer. However, the quality of the manual contouring depends on the radiation oncologist's knowledge and experience. Also, manual segmentation is prone to

Table 4: Comparison of the dose-volume parameters of the organs at risks between manual and automatic segmentation (mean±SD)

Organ	ΔD_{min} (Gy)	ΔD_{max} (Gy)	ΔD_{mean} (Gy)	$\Delta V_{50\%}$ (%)	$\Delta V_{65\%}$ (%)	$\Delta V_{70\%}$ (%)
Rectum	0.304±0.900	-0.009±0.329	0.330±1.642	0.071±3.011	0.840±2.048	0.838±2.073
Bladder	0.032±0.428	0.050±0.158	0.251±0.887	0.618±2.096	0.142±1.569	0.084±1.404
Left femur	0.084±0.233	-0.249±0.979	0.202±0.456	0.158±0.408	0.016±0.098	-0.010±0.028
Right femur	0.098±0.206	-0.016±1.714	0.198±0.311	0.175±0.821	0.126±0.505	0.075±0.354

Bold values indicate that there is a statistically significant difference in lymph node dose-volume parameters between the manual segmentation (ground truth) and the automatically generated contours ($P<0.05$)
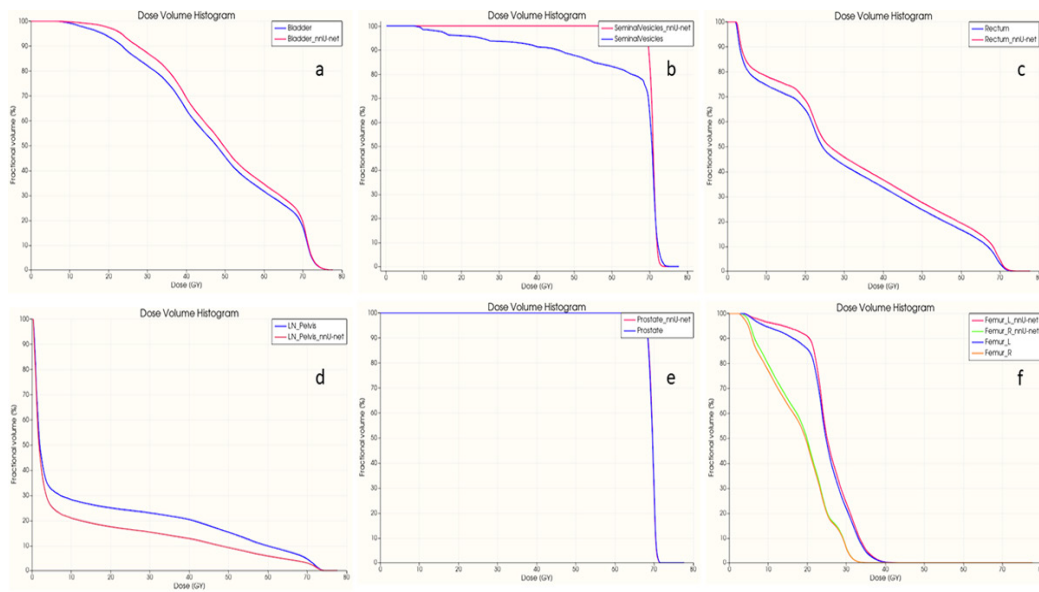
Figure 4: Dose-Volume Histogram (DVH) of a test case for plans generated by nnU-Net model and ground truth contours. **a**) Bladder, **b**) Seminal vesicle, **c**) Rectum, **d**) Pelvis lymph nodes, **e**) Prostate, **f**) left femur, and right femur.

inter- and intra-rater variability. Besides, CT images suffer from poor soft tissue contrast; therefore, the contouring process can be more challenging. To address the aforementioned challenges, deep learning methods, a branch of AI research, has emerged as a new promising approach to directly generate contours over the recent years. Deep learning-based contouring can generate accurate and reproducible delineation of structures in planning CT images. Auto-contouring of targets and OARs can reduce radiation-induced normal tissue toxicity, leading to dose escalation in prostate radiotherapy. Therefore, this study aimed to assess the feasibility of a deep CNN model for simultaneous multi-structure contouring in prostate radiotherapy. Here, training of a 3D nnU-net model has been accomplished with success and used for auto-contouring on male pelvic CT images. Also, the model's performance was evaluated based on both geometric metrics and clinically relevant dose-volume parameters. The obtained data showed that the contours generated using the nnU-net model had high overlap/agreement with manually delineated contours as the standard ground truth.

The nnU-net model, a self-adapting ensemble method, for simultaneous multi-structure

Table 5: Performance comparison of proposed model with the state-of-the-art models for pelvis organs segmentation on the CT images according to the Dice similarity coefficient

Study/ year	Method	Cases	Bladder	Rectum	Prostate	Lymph nodes	Seminal vesicles	Left femur	Right femur
Kawula et al./ 2022 [13]	3D U-net	69	0.97	0.89	0.87	-	-	-	-
Balagopal et al./ 2018 [34]	2D-3D U-net	136	0.95	0.84	0.90	-	-	0.96	0.95
Tong et al./ 2021 [35]	MTER-Net	200	0.96	0.86	0.86	-	-	-	-
Sultana et al./ 2020 [36]	3D UNet-GAN	290	0.95	0.90	0.91	-	-	-	-
Wang et al./ 2021 [37]	3D BCnet	313	0.93	0.92	0.89	-	-	-	-
Kiljunen et al./ 2020 [38]	Commercial DL-based AST	900	0.93	0.84	0.82	0.80	0.72	0.68	0.69
Present study	3D nnU-net	118	0.97	0.90	0.82	0.77	0.69	0.96	0.96

DL: Deep Learning; AST: Automatic Segmentation Tool

segmentation, was employed in prostate radiotherapy. Moreover, the nnU-net shows a streamlined workflow and eliminates the need for time-consuming fine-tuning with accurate and reproducible segmentation; therefore, it has great potential for widespread use in the clinic. In this study, the 3D nnU-net achieved a satisfactory performance regarding the geometric accuracy of the organ segmentations, i.e., a high degree of similarity was observed between manually and automatically delineated contours. According to the Dice metric, a performance comparison was conducted between our proposed auto-segmentation model and other state-of-the-art models reported for pelvis organ segmentation on the CT images, as outlined in Table 5. It is also worth noting that it is not a precise way to directly compare our proposed deep learning model with the reviewed methods for automated multi-organ segmentation in the male pelvic area, as different CT databases with different dataset sizes, various CT scanners, different cancer types, and different OAR contouring standards were used in each study. Nevertheless, the

comparison summary proved that our proposed architecture has similar or even better performance for multi-organ segmentation in the pelvic area. In this study, the most promising findings were observed for bladder and femoral head contouring, followed by the rectum, and prostate. More recently, Kawula et al. applied 3D U-net for CT-based multi-organ segmentation in prostate radiotherapy [13]. The results of their study showed that 3D U-net achieved the best results for bladder segmentation with a mean Dice value of 0.97, followed by the rectum (mean DSC: 0.89), and prostate (mean DSC: 0.87) [13]. The highest segmentation performance of the nnU-net model for the bladder can be attributed to the simplicity of its geometry and its considerable size. Regarding bladder auto-segmentation, some previous studies reported a wide range of Dice values (67-93%) due to filling level and shape [39-42]. Contouring of femoral heads generally is highly concordant, with a DSC range of 90–95% [39-41, 43], which is in agreement with our data (mean: 96%). One possible reason for this overlap can be explained

by the good contrast of femoral heads with the surrounding structures and a well-defined regular shape. The low contrast of the lymph nodes, seminal vesicles, prostate, and rectum on the CT images results in their contouring being most challenging. In the present study, automated segmentation of seminal vesicles and pelvic lymph nodes had the lowest Dice values of 69% and 77%, respectively. The seminal vesicles have a small volume and are placed in regions with high anatomical variation. The lymph nodes have more variety in size, boundaries, and localization; hence, their segmentation is very challenging. Balagopal et al. proposed a 2D–3D hybrid network for fully automated multi-organ delineation in male pelvic CT images and obtained a DSC of 0.90, 0.95, and 0.84 for prostate, bladder, and rectum, respectively [34]. In another study, Tong et al. applied the edge-calibrated multi-task network for male pelvic multi-structure contouring on CT images, achieving an overall Dice score of 0.89 for bladder, rectum, and prostate segmentation [35]. The UNet-GAN hybrid model developed by Sultana et al. obtained a Dice score of 0.91 for prostate [36]. As observable in various studies, the bladder obtained the highest segmentation accuracy.

The novelty of this study lies in the following: 1) the proposed model contoured both OARs and target volumes. As observable in Table 5, most existing studies have only contoured the prostate, rectum, and bladder. Here, not only the prostate as the target volume, but also seminal vesicles and lymph nodes were contoured. In addition, all OARs (i.e., rectum, bladder, and femoral heads) were contoured; 2) both geometric and DVH were applied to assess delineation accuracy between the auto-generated contours and the ground truths of clinicians' contouring as the standard of reference. In general, the evaluation of the performance of auto-segmentation models is based on commonly used geometric metrics, such as DSC and HD. Geometric metrics are not directly associated with the treatment plan

dosimetry; therefore, it is challenging to assess the accuracy and effectiveness of automated segmentations in dose optimization and plan evaluation. The performance of the nnU-net model was assessed by considering not only geometric metrics but also clinically relevant dose-volume parameters. In overall, a high level of agreement was observed between the results obtained from the planning based on manually and automatically segmented contours. A statistically significant difference was observed for the femoral heads and lymph nodes. Few studies have investigated the dosimetric effect of auto-generated contours [13].

It is worthwhile to mention that auto-contouring algorithms can generate reliable segmentations in a short time, as compared to manual delineation, which may take 20-30 min. Previous studies highlighted the time-saving benefit of the nnU-net model [44-46]. A previous study found that the nnU-net model required approximately 20s and 15s to segment the whole breast and the fibro-glandular tissue under the dynamic contrast-enhanced magnetic resonance images, respectively [46]. In the current study, the 3D nnU-net model took approximately 100s to segment an input CT volume using a standard PC with a Ge-Force GTX 8 GB NVIDIA and 32 GB RAM.

The present study has some limitations, as follows: 1) the utilization of a restricted dataset, so that deep learning models require a large number of subjects for training. A more robust and stable deep CNN model should be trained and tested on large multi-center datasets and 2) memory and computation power. The proposed network's performance can be improved with advanced memory and computational power. Besides, a single radiation oncologist delineated all contours as ground truth. It is important to point out that in real-world clinics inter-observer variations are known to exist among physicians, and no 100% gold standard is found. Currently, the proposed models can be applied as a supportive tool for radiation oncologists.

Conclusion

In this study, the feasibility of auto-contouring was evaluated using a deep CNN approach for CT-based targets and OAR segmentation. The performance of the deep CNN auto-segmentation tool was assessed using manually delineated contours by an experienced radiation oncologist as ground truth. Our results indicated that the 3D nnU-net architecture can be successfully used for multi-organ segmentation in the male pelvic area. No statistically significant difference was in dosimetric endpoints of targets and OARs except for lymph nodes and femoral heads between treatment plans based on manually and automatically segmented contours. The proposed model has the potential to decrease radiation oncologist's workload by reducing the segmentation time required to generate acceptable contours. The integration of the proposed framework into current clinical practice may increase the efficiency of the RT workflow. The proposed model shows promise as an automated tool for further auto-segmentation workflow studies in RT.

Authors' Contribution

B. Mofid contributed to the study conception and design, supervision of the study, data analyses, and writing the first draft of the manuscript; M. Mosalla contributed to data collection, data analyses, writing the first draft of the manuscript; M. Goodarzi contributed to data collection, data analyses: writing the first draft of the manuscript; H. Tavakoli contributed to the study design, data analyses, and interpretation and revision of the manuscript. All the authors read, modified, and approved the final version of the manuscript.

Ethical Approval

This study was approved by the ethical committee of Baqiyatallah University of Medical Sciences, Tehran, Iran. Due to the retrospective nature of the data used in this study and the absence of any human or animal intervention, the research proposal for the aforementioned project has been approved by the Ethics Committee of the University's Radiation Injury Research Center. No ethical code has been assigned to this proposal according to the committee's discretion.

Conflict of Interest

None

References

1. Wang L, Lu B, He M, Wang Y, Wang Z, Du L. Prostate Cancer Incidence and Mortality: Global Status and Temporal Trends in 89 Countries From 2000 to 2019. *Front Public Health*. 2022;**10**:811044. doi: 10.3389/fpubh.2022.811044. PubMed PMID: 35252092. PubMed PMCID: PMC8888523.

2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021;**71**(3):209-49. doi: 10.3322/caac.21660. PubMed PMID: 33538338.

3. Gay HA, Michalski JM. Radiation Therapy for Prostate Cancer. *Mo Med*. 2018;**115**(2):146-50. PubMed PMID: 30228707. PubMed PMCID: PMC6139853.

4. Baskar R, Lee KA, Yeo R, Yeoh KW. Cancer and radiation therapy: current advances and future directions. *Int J Med Sci*. 2012;**9**(3):193-9. doi: 10.7150/ijms.3635. PubMed PMID: 22408567. PubMed PMCID: PMC3298009.

5. Kosmin M, Ledsam J, Romera-Paredes B, Mendes R, Moinuddin S, De Souza D, et al. Rapid advances in auto-segmentation of organs at risk and target volumes in head and neck cancer. *Radiother Oncol*. 2019;**135**:130-40. doi: 10.1016/j.radonc.2019.03.004. PubMed PMID: 31015159.

6. Deeley MA, Chen A, Datteri R, Noble JH, Cmelak AJ, Donnelly EF, et al. Comparison of manual and automatic segmentation methods for brain structures in the presence of space-occupying lesions: a multi-expert study. *Phys Med Biol*. 2011;**56**(14):4557-77. doi: 10.1088/0031-9155/56/14/021. PubMed PMID: 21725140. PubMed PMCID: PMC3153124.

7. Chen MY, Woodruff MA, Dasgupta P, Rukin NJ. Variability in accuracy of prostate cancer segmentation among radiologists, urologists, and scientists. *Cancer Med*. 2020;**9**(19):7172-82. doi: 10.1002/cam4.3386. PubMed PMID: 32810385. PubMed PMCID: PMC7541146.

8. Valicenti RK, Sweet JW, Hauck WW, Hudes RS, Lee T, Dicker AP, et al. Variation of clinical target volume definition in three-dimensional conformal radia-

tion therapy for prostate cancer. *Int J Radiat Oncol Biol Phys.* 1999;**44**(4):931-5. doi: 10.1016/s0360-3016(99)00090-5. PubMed PMID: 10386652.

9. Fiorino C, Reni M, Bolognesi A, Cattaneo GM, Calandrino R. Intra- and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. *Radiother Oncol.* 1998;**47**(3):285-92. doi: 10.1016/s0167-8140(98)00021-8. PubMed PMID: 9681892.

10. West CM, Huddart RA. Biomarkers and Imaging for Precision Radiotherapy. *Clin Oncol (R Coll Radiol).* 2015;**27**(10):545-6. doi: 10.1016/j.clon.2015.06.021. PubMed PMID: 26173954.

11. Chen AM, Chin R, Beron P, Yoshizaki T, Mikaeilian AG, Cao M. Inadequate target volume delineation and local-regional recurrence after intensity-modulated radiotherapy for human papillomavirus-positive oropharynx cancer. *Radiother Oncol.* 2017;**123**(3):412-8. doi: 10.1016/j.radonc.2017.04.015. PubMed PMID: 28511960.

12. Njeh CF. Tumor delineation: The weakest link in the search for accuracy in radiotherapy. *J Med Phys.* 2008;**33**(4):136-40. doi: 10.4103/0971-6203.44472. PubMed PMID: 19893706. PubMed PMCID: PMC2772050.

13. Kawula M, Purice D, Li M, Vivar G, Ahmadi SA, Parodi K, et al. Dosimetric impact of deep learning-based CT auto-segmentation on radiation therapy treatment planning for prostate cancer. *Radiat Oncol.* 2022;**17**(1):21. doi: 10.1186/s13014-022-01985-9. PubMed PMID: 35101068. PubMed PMCID: PMC8805311.

14. Elguindi S, Zelefsky MJ, Jiang J, Veeraraghavan H, Deasy JO, Hunt MA, Tyagi N. Deep learning-based auto-segmentation of targets and organs-at-risk for magnetic resonance imaging only planning of prostate radiotherapy. *Phys Imaging Radiat Oncol.* 2019;**12**:80-6. doi: 10.1016/j.phro.2019.11.006. PubMed PMID: 32355894. PubMed PMCID: PMC7192345.

15. Duan J, Bernard M, Downes L, Willows B, Feng X, Mourad WF, St Clair W, Chen Q. Evaluating the clinical acceptability of deep learning contours of prostate and organs-at-risk in an automated prostate treatment planning process. *Med Phys.* 2022;**49**(4):2570-81. doi: 10.1002/mp.15525. PubMed PMID: 35147216.

16. Sharp G, Fritscher KD, Pekar V, Peroni M, Shusharina N, Veeraraghavan H, Yang J. Vision 20/20: perspectives on automated image segmentation for radiotherapy. *Med Phys.* 2014;**41**(5):050902. doi: 10.1118/1.4871620. PubMed PMID: 24784366. PubMed PMCID: PMC4000389.

17. Heilemann G, Buschmann M, Lechner W, Dick V, Eckert F, Heilmann M, et al. Clinical Implementation and Evaluation of Auto-Segmentation Tools for Multi-Site Contouring in Radiotherapy. *Phys Imaging Radiat Oncol.* 2023;**28**:100515. doi: 10.1016/j.phro.2023.100515. PubMed PMID: 38111502. PubMed PMCID: PMC10726238.

18. Savjani RR, Lauria M, Bose S, Deng J, Yuan Y, Andrearczyk V. Automated Tumor Segmentation in Radiotherapy. *Semin Radiat Oncol.* 2022;**32**(4):319-29. doi: 10.1016/j.semradonc.2022.06.002. PubMed PMID: 36202435.

19. Lustberg T, Van Soest J, Gooding M, Peressutti D, Aljabar P, Van Der Stoep J, et al. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. *Radiother Oncol.* 2018;**126**(2):312-7. doi: 10.1016/j.radonc.2017.11.012. PubMed PMID: 29208513.

20. Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in Auto-Segmentation. *Semin Radiat Oncol.* 2019;**29**(3):185-97. doi: 10.1016/j.semradonc.2019.02.001. PubMed PMID: 31027636.

21. Ghaffari H, Tavakoli H, Pirzad Jahromi G. Deep transfer learning-based fully automated detection and classification of Alzheimer's disease on brain MRI. *Br J Radiol.* 2022;**95**(1136):20211253. doi: 10.1259/bjr.20211253. PubMed PMID: 35616643. PubMed PMCID: PMC10162060.

22. Mohammadi R, Salehi M, Ghaffari H, Reiazi R. Transfer Learning-Based Automatic Detection of Coronavirus Disease 2019 (COVID-19) from Chest X-ray Images. *J Biomed Phys Eng.* 2020;**10**(5):559-68. doi: 10.31661/jbpe.v0i0.2008-1153. PubMed PMID: 33134214. PubMed PMCID: PMC7557468.

23. Salehi M, Ardekani MA, Taramsari AB, Ghaffari H, Haghparast M. Automated deep learning-based segmentation of COVID-19 lesions from chest computed tomography images. *Pol J Radiol.* 2022;**87**:e478-86. doi: 10.5114/pjr.2022.119027. PubMed PMID: 36091652. PubMed PMCID: PMC9453472.

24. Salehi M, Mohammadi R, Ghaffari H, Sadighi N, Reiazi R. Automated detection of pneumonia cases using deep transfer learning with paediatric chest X-ray images. *Br J Radiol.* 2021;**94**(1121):20201263. doi: 10.1259/bjr.20201263. PubMed PMID: 33861150. PubMed PMCID: PMC8506182.

25. Showkatian E, Salehi M, Ghaffari H, Reiazi R, Sadighi N. Deep learning-based automatic detection of tuberculosis disease in chest X-ray images. *Pol J Radiol.* 2022;**87**:e118-24. doi: 10.5114/pjr.2022.113435. PubMed PMID: 35280947. PubMed PMCID: PMC8906182.

26. Xiao H, Ren G, Cai J. A review on 3D deformable image registration and its application in dose warping. *Radiat Med Protect.* 2020;**1**(4):171-8. doi: 10.1016/j.radmp.2020.11.002.

27. Van Dijk LV, Van Den Bosch L, Aljabar P, Peressutti D, Both S, et al. Improving automatic delineation for head and neck organs at risk by Deep Learning Contouring. *Radiother Oncol.* 2020;**142**:15-123.

doi: 10.1016/j.radonc.2019.09.022. PubMed PMID: 31653573.

28. Guo Y, Gao Y, Shen D. Deformable MR Prostate Segmentation via Deep Feature Learning and Sparse Patch Matching. *IEEE Trans Med Imaging.* 2016;**35**(4):1077-89. doi: 10.1109/TMI.2015.2508280. PubMed PMID: 26685226. PubMed PMCID: PMC5002995.

29. Liang S, Tang F, Huang X, Yang K, Zhong T, Hu R, et al. Deep-learning-based detection and segmentation of organs at risk in nasopharyngeal carcinoma computed tomographic images for radiotherapy planning. *Eur Radiol.* 2019;**29**(4):1961-7. doi: 10.1007/s00330-018-5748-9. PubMed PMID: 30302589.

30. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging.* 2015;**15**:29. doi: 10.1186/s12880-015-0068-x. PubMed PMID: 26263899. PubMed PMCID: PMC4533825.

31. Guo H, Wang J, Xia X, Zhong Y, Peng J, Zhang Z, Hu W. The dosimetric impact of deep learning-based auto-segmentation of organs at risk on nasopharyngeal and rectal cancer. *Radiat Oncol.* 2021;**16**(1):113. doi: 10.1186/s13014-021-01837-y. PubMed PMID: 34162410. PubMed PMCID: PMC8220801.

32. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods.* 2021;**18**(2):203-11. doi: 10.1038/s41592-020-01008-z. PubMed PMID: 33288961.

33. Anderson BM, Wahid KA, Brock KK. Simple Python Module for Conversions Between DICOM Images and Radiation Therapy Structures, Masks, and Prediction Arrays. *Pract Radiat Oncol.* 2021;**11**(3):226-9. doi: 10.1016/j.prro.2021.02.003. PubMed PMID: 33607331. PubMed PMCID: PMC8102371.

34. Balagopal A, Kazemifar S, Nguyen D, Lin MH, Hannan R, Owrangi A, Jiang S. Fully automated organ segmentation in male pelvic CT images. *Phys Med Biol.* 2018;**63**(24):245015. doi: 10.1088/1361-6560/aaf11c. PubMed PMID: 30523973.

35. Tong N, Gou S, Chen S, Yao Y, Yang S, Cao M, et al. Multi-task edge-recalibrated network for male pelvic multi-organ segmentation on CT images. *Phys Med Biol.* 2021;**66**(3):035001. doi: 10.1088/1361-6560/abcad9. PubMed PMID: 33197901.

36. Sultana S, Robinson A, Song DY, Lee J. Automatic multi-organ segmentation in computed tomography images using hierarchical convolutional neural network. *J Med Imaging (Bellingham).* 2020;**7**(5):055001. doi: 10.1117/1.JMI.7.5.055001. PubMed PMID: 33102622. PubMed PMCID: PMC7554423.

37. Wang S, Liu M, Lian J, Shen D. Boundary Coding Representation for Organ Segmentation in Prostate Cancer Radiotherapy. *IEEE Trans Med Imaging.* 2021;**40**(1):310-320. doi: 10.1109/TMI.2020.3025517. PubMed PMID: 32956051.

38. Kiljunen T, Akram S, Niemelä J, Löyttyniemi E, Seppälä J, Heikkilä J, et al. A Deep Learning-Based Automated CT Segmentation of Prostate Cancer Anatomy for Radiation Therapy Planning-A Retrospective Multicenter Study. *Diagnostics (Basel).* 2020;**10**(11):959. doi: 10.3390/diagnostics10110959. PubMed PMID: 33212793. PubMed PMCID: PMC7697786.

39. Men K, Dai J, Li Y. Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks. *Med Phys.* 2017;**44**(12):6377-89. doi: 10.1002/mp.12602. PubMed PMID: 28963779.

40. Liu Z, Liu X, Xiao B, Wang S, Miao Z, Sun Y, Zhang F. Segmentation of organs-at-risk in cervical cancer CT images with a convolutional neural network. *Phys Med.* 2020;**69**:184-91. doi: 10.1016/j.ejmp.2019.12.008. PubMed PMID: 31918371.

41. Hwee J, Louie AV, Gaede S, Bauman G, D'Souza D, Sexton T, Lock M, Ahmad B, Rodrigues G. Technology assessment of automated atlas based segmentation in prostate bed contouring. *Radiat Oncol.* 2011;**6**:110. doi: 10.1186/1748-717X-6-110. PubMed PMID: 21906279. PubMed PMCID: PMC3180272.

42. Li D, Zang P, Chai X, Cui Y, Li R, Xing L. Automatic multiorgan segmentation in CT images of the male pelvis using region-specific hierarchical appearance cluster models. *Med Phys.* 2016;**43**(10):5426. doi: 10.1118/1.4962468. PubMed PMID: 27782723. PubMed PMCID: PMC5035314.

43. Sartor H, Minarik D, Enqvist O, Ulén J, Wittrup A, Bjurberg M, Trägårdh E. Auto-segmentations by convolutional neural network in cervical and anorectal cancer with clinical structure sets as the ground truth. *Clin Transl Radiat Oncol.* 2020;**25**:37-45. doi: 10.1016/j.ctro.2020.09.004. PubMed PMID: 33005756. PubMed PMCID: PMC7519211.

44. Mihelic SA, Sikora WA, Hassan AM, Williamson MR, Jones TA, Dunn AK. Segmentation-Less, Automated, Vascular Vectorization. *PLoS Comput Biol.* 2021;**17**(10):e1009451. doi: 10.1371/journal.pcbi.1009451. PubMed PMID: 34624013. PubMed PMCID: PMC8528315.

45. Zhu Y, Chen L, Lu W, Gong Y, Wang X. The application of the nnU-Net-based automatic segmentation model in assisting carotid artery stenosis and carotid atherosclerotic plaque evaluation. *Front Physiol.* 2022;**13**:1057800. doi: 10.3389/fphys.2022.1057800. PubMed PMID: 36561211. PubMed PMCID: PMC9763590.

46. Huo L, Hu X, Xiao Q, Gu Y, Chu X, Jiang L. Segmentation of whole breast and fibroglandular tissue using nnU-Net in dynamic contrast enhanced MR images. *Magn Reson Imaging.* 2021;**82**:31-41. doi: 10.1016/j.mri.2021.06.017. PubMed PMID: 34147598.