Deep Learning in Radiation Oncology Treatment Planning for Prostate Cancer: A Systematic Review

Gonçalo Almeida · João Manuel R.S.
Tavares

Abstract Radiation oncology for prostate cancer is important as it can decrease the morbidity and mortality associated with this disease. Planning for this modality of treatment is both fundamental, time-consuming and prone to human-errors, leading to potentially avoidable delays in start of treatment. A fundamental step in radiotherapy planning is contouring of radiation targets, where medical specialists contouring, i.e., segment, the boundaries of the structures to be irradiated. Automating this step can potentially lead to faster treatment planning without a decrease in quality, while increasing time available to physicians and also more consistent treatment results. This can be framed as an image segmentation task, which has been studied for many decades in the fields of Computer Vision and Machine Learning. With the advent of Deep Learning, there have been many proposals for different network architectures achieving high performance levels. In this review, we searched the literature for those methods and describe them briefly, grouping those based on Computed Tomography (CT) or Magnetic Resonance Imaging (MRI). This is a booming field, evidenced by the date of the publications found. However, most publications use data from a very limited number of patients, which presents an obstacle to deep learning models training. Although the performance of the models has achieved very satisfactory results, there is still room for improvement, and there is arguably a long way before these models can be used safely and effectively in clinical practice.

G. Almeida
Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Faculdade de Engenharia, Universidade do Porto, Porto, PORTUGAL
E-mail: galmeida@inegi.up.pt

J.M.R.S. Tavares
Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Departamento de Engenharia Mecânica, Faculdade de Engenharia, Universidade do Porto, Porto, PORTUGAL
E-mail: tavares@fe.up.pt

Fig. 1 Anatomical and CT slices of the same person (from the Visible Human Project).

Keywords Computer Vision · Medical Imaging · Machine Learning · Segmentation · Radiotherapy

1 Introduction

Prostate cancer is the most prevalent non-cutaneous cancer and the second leading cause of cancer death in men. It is estimated that about 1 in 9 men will be diagnosed with prostate cancer in their lifetime [1,2]. However, with appropriate treatment, the 5-year survival rate is 98.2% [3,4].

Radiation Therapy (RT) is a fundamental part of modern cancer care, serving more than 1.5 million patients worldwide [5], and the demand is projected to increase by 16% by 2025 [6]. It is considered that half of all cancer patients would benefit from radiotherapy in the course of their disease, but in Europe at least 25% of patients do not receive this treatment [7].

Radiotherapy planning is a sequence of events that lead from the initial patient encounter to the start of treatment. With the advancement of radiation delivery techniques, the number of treatment parameters increased so much that planning has become a very complex and time-consuming task, taking hours to days of human effort for each patient [8]. One of the most critical steps is the contouring of radiation targets which is akin to a segmentation task. It is just as important to ensure enough radiation dose in the target, as it is to stay below specific radiation absorption levels in the normal tissues [9].

It is fundamental that anatomical features are defined on Computed Tomography (CT) images, because radiation dose calculations depend on tissue electronic densities [10], which are not available on other imaging modalities. Due to their specific advantages, Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) scans can be, and are, also used in clinical practice to help with contouring; yet, CT is indispensable to radiation oncology (Figure 1) [11].

The output of this organ contouring task is a segmentation mask, as it can be thought of as an overlay onto the medical image (Figure 2). As any other task performed by human beings, this is subject to variability. Inter-observer variability for prostate manual segmentation has been estimated at 18-25%, using various different metrics [13,14]. These studies also attempted
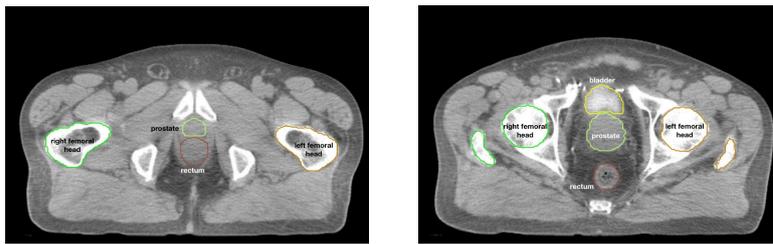
Fig. 2 Examples of segmentation masks for prostate cancer radiotherapy [12].

to find intra-observer variability, where the same physician contoured the same CT scan more than one time, finding variations of 1.5-9.0% [13] and 2-8% [14]. This variability has a significant impact on the final treatment plan and patient outcomes.

There are also small anatomic variations due to patient positioning during image acquisition, such as bladder and rectum filling, movement due to breathing, state of hydration, weight loss, muscle contraction due to anxiety and room temperature and even due to heart pumping. Radiotherapy centers follow specific protocols to minimize the impact produced by these factors, so that planning CT images are as reproducible as possible [15], but there will always be some residual uncertainty [16].

If the step of volume definition could be fully automated while ensuring perfect correlation with the anatomical structures, the gains would be immense: no uncertainties associated with manual delineation, ensuring perfect encompassing of the target volume, increasing treatment success and patient survival; decreasing the geometric margins used today to compensate for errors, leading to fewer radiation-induced complications as a smaller total volume would be irradiated; increase in physician time; less time between first patient encounter and treatment start, which also improves patient outcomes.

This review focuses on deep learning for segmentation of radiation targets and normal tissues during radiotherapy planning for prostate cancer. Although there is a considerable number of review articles on the use of deep learning for medical image analysis [17–19], to the best of the authors' knowledge, there is none on this particular topic. This is a nascent field, arguably in an early stage of development, but in a very important moment of increasing interest by the research community. The number of articles reviewed can be considered short, but this is due to novelty, and their very recent publication dates helps explain the importance that the subject has been gathering: there have been more than a dozen studies published in the last year, warranting the need for a review of this subject. Therefore, this study can be invaluable both for image analysis researchers as well as clinical practitioners who should recognise the potential benefits of automatic segmentation in their practice.

The male pelvic anatomy is composed of organs with similar density (soft-tissues), such as the prostate, bladder and rectum. The boundaries between

these organs are often very hard to distinguish, due to the small differences in contrast (especially in CT images). Also, this is one of the anatomical regions with the most variation between subjects but also within the same individual, owing to changes in content of the bladder and rectum. Furthermore, other structures are required to be segmented, which include the femoral heads and the penile bulb, offering additional challenges to designing models capable of performing automatic segmentation of all the structures needed for radiotherapy planning for prostate cancer. Unlike some other anatomical locations, RT for prostate cancer is performed by irradiating the whole organ, which must be the segmentation object, as opposed to segmentation of a tumor volume inside a given organ (e.g. lung, breast). This seems to make the task comparatively easier, but, in fact, increases the importance of an accurate segmentation because of the adjacent structures, not allowing for much margin of error.

This article is structured as follows: the next section presents a brief overview of deep learning as a tool for image analysis and the models that were developed for image segmentation. In the methods section, the search developed for this review is described. The results section presents the main findings achivied with the selected works grouped by the used imaging modality: MRI and CT. The final section provides a critical discussion of the results and draws the conclusions.

2 Machine Learning and Deep Learning

Deep learning is a subfield of Machine Learning (ML) which uses neural networks with many hidden layers to map a certain input to a predefined output. The used deep learning model has the intrinsic ability to learn useful features directly from the input data that are important to the task at hand. This may be classification, regression, clustering or segmentation, among others. When the output labels are available, the method of learning is called supervised learning [20].

2.1 Convolutional Neural Networks

In 2012, in a landmark moment in the history of Machine Learning and Computer Vision, researchers from the University of Toronto achieved an impressive result in the ImageNet competition conquering the first place by a large margin to all other teams [21]. They used a Convolutional Neural Network (CNN) several layers deep and were able to nearly halve the error of the previous best result. CNNs had been first used over two decades before, when LeCun et al. applied them to the task of recognizing hand-written digits in 1990 [22].

In a CNN, at least one layer of the network performs a convolution operation on that layer's inputs. This is performed by means of a filter or kernel, which is translated across and down the input matrix to produce a representation map of the original image. The filter is composed of learnable parameters

which are updated through gradient descent. The filter is the same for every part of the image (at each convolution operation), such that any features can be extracted irrespective of their location in the image. This operation is very successful because it is capable of correlating a certain pixel's information with that contained by adjacent pixels.

2.2 From classification to segmentation with CNNs

It is clear nowadays that deep learning-based models have become the state of the art in medical image analysis, as most of the challenges are now populated with these methods in all the top positions. However, it is worth understanding that this only became true recently. It was only in 2017 that a fully Convolutional Neural Network architecture proposed by Yu et al.[23] captured the first place in the PROMISE12 prostate segmentation challenge. Since then, traditional ML models have been continually going down the leaderboard.

Image classification is performed with CNNs, where a full image is input into a network which returns a single one-hot vector as output, assigning the image to one of several classes. For this architecture to be used for segmentation, it had to be modified such that the output would consist of a segmentation mask (or could later be processed to become one). The first attempts had the network look at a small part of the image (a patch) and classify it as belonging to the object class or background. By dividing the image into patches and classifying all of them, one could then build a rough segmentation mask. However, the network never got to fully grasp the whole image, as it would only have access to a small region of the image at a time.

Shelhamer, Long and Darrel came up with the Fully Convolutional Network (FCN) architecture which produced a pixel-wise prediction all in one go [24]. This worked by appending a layer that performs upsampling through a transposed convolution, also called a deconvolution or up-convolution, decoding the information contained in the deepest layers of the CNN back into the full size of the original image. Ronneberger et al. designed the U-Net by dividing the network into two distinct parts: I) an encoding arm, progressively downsampling the input through convolutions and pooling (similarly to the traditional CNN); and II) a decoding arm, completely symmetric to the encoding part, where at each upsampling step they bring the spatial coordinate information of the image from the opposite side of the network and use concatenation to place it together with the result of the up-convolution operation [25]. This model outputs a segmentation mask, one value for each pixel of the original image: the segmentation task was transformed into a pixel-wise classification task.

U-Net became widely used for medical imaging segmentation and several improvements were soon made. Cicek et al. created a version of U-net capable of using 3D inputs instead of 2D images [26]. Similarly, Milletari et al. proposed V-Net, a volumetric version of U-Net and incorporated the Dice coefficient into the loss function [27]. The advantage of having a 3D architecture is that instead

of supplying a slice of a CT or MRI volume, one inputs the whole volume into the model, allowing for representation learning from all the data at once, akin to the evolution from patch-based to whole image input. Unlike a 2D model, which loses information situated between slices and is incapable of inferring surface continuation [28], a 3D model can comprehend these details, which are especially useful at the top and bottom ends of each structure. This increases both accuracy and ease of use, at the cost of computational capability.

Yu et al. added residual skip connections to the U-Net architecture, and it was the first time a fully convolutional neural network topped PROMISE12 [23]; Oktay et al. used a concept from Recurrent Neural Networks (RNN - used for sequential data such as text and speech), and added Attention gates to the concatenation step when bringing together the information about the spatial coordinates of the input with the resulting feature map from up-convolutions, with encouraging results for medical segmentation [29].

3 Methods

A systematic literature search was performed in Pubmed/Medline, ScienceDirect and Scopus databases with the following keywords in various combinations: "deep learning", "convolutional neural networks", "neural networks", "machine learning", "prostate segmentation", "prostate", "segmentation", "radiotherapy", "radiation therapy", "radiation oncology". This produced a total of 528 unique results, most of which were completely unrelated to the subject at hand, based on analysis of title and abstract. For further analysis, 47 articles were selected based on the following inclusion criteria: segmentation of the prostate, image analysis from studies of patients with prostate cancer, research related with radiation therapy. The exclusion criteria were: segmentation of structures for rectal cancer, inclusion of female anatomy images, use of ultrasound images (e.g., trans-rectal ultrasound), and not using deep learning segmentation models. Among the selected results, 11 were literature reviews which are detailed in the results section. Of the 36 original studies selected based on abstract, after careful review of the body of text, 28 were kept as they are specifically related to prostate segmentation using deep learning on CT or MR images. Figure 3 shows the performed search and the obtained results in the form of a PRISMA diagram.

4 Results

In a total of 28 selected articles, 19 performed segmentation on MRI images and 9 on CT images. As aforementioned, CT is essential for radiotherapy, but often radiation oncologists use MRI together with CT (registered or not) to help manual contouring. Thus, segmentations on MRI were also included, seeing as they can be useful in clinical practice. However, prostate segmentation in MRI is used with more goals in medicine than just for RT planning. Typically,
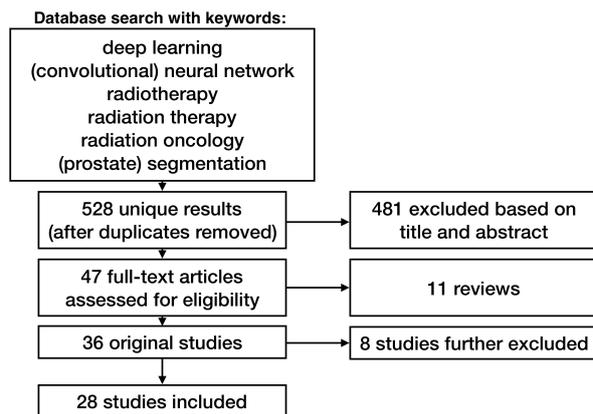
Database search with keywords:

deep learning
(convolutional) neural network
radiotherapy
radiation therapy
radiation oncology
(prostate) segmentation

528 unique results
(after duplicates removed) → 481 excluded based on title and abstract

47 full-text articles
assessed for eligibility → 11 reviews

36 original studies → 8 studies further excluded

28 studies included

Fig. 3 PRISMA diagram showing the results of the performed literature search.

segmentation on MRI is considered somewhat easier than on CT on account of higher contrast between organs and structures on the acquired MR images. This is observed in the performance of the reviewed articles.

In our search we found 11 reviews which at least briefly mention auto-segmentation of the prostate with deep learning, but none offers a systematic review of studies on this segmentation task. Among the studies on prostate segmentation, at most 6 articles are presented in one of the reviews [30], with the others merely mentioning in the text or presenting in a table fewer than those. The earliest of these reviews was published in 2017, with a comprehensive review of DL for many applications in medical imaging [31], and includes a table where 5 articles covering prostate segmentation are presented. The remaining range from focusing on the potential benefits of artificial intelligence (AI) and DL to cancer imaging as a whole [32], to medical image analysis [17, 33], to radiation therapy in all phases of the planning process [8,11,19,34], to covering ML applications for prostate cancer from diagnosis to treatment to follow-up [35,36].

Some articles mention if the ground truth segmentation was performed by a radiologist, radiation oncologist or other medical specialty. This is important because segmentation for prostate volume assessment (typically performed by radiologists in MRI) is quite different from segmentation for radiation therapy, where specific training is undertaken and millimetric accuracy is essential (detailed in Tables 1 and 4).

For the development of automated segmentation methods to be possible, there have to be ways to correctly compare validated manual segmentations to those performed by computational algorithms. There is no single best metric for this purpose, as some give more importance to volume differences while others impose a higher weight on boundary differences. Important to note, nevertheless, that most published works proposing computational segmentation methods report Dice Similarity Coefficient (DSC) and some measure of

surface distance, be it Average Boundary Distance (ABD, eq.2) or Hausdorff Distance (HD). Particularly, 95%HD is the 95th percentile of the Hausdorff Distance, seen widely in the literature. The Dice Similarity Coefficient is calculated as:

$$Dice\ Similarity\ Coefficient = \frac{2|A \cap B|}{|A| + |B|}, \qquad (1)$$

where A and B are the segmented volumes to be compared. The Average Boundary Distance and Hausdorff distances are, respectively:

$$ABD = \frac{1}{|A_s| + |B_s|} \left(\sum_{a \in A_s} min_{b \in B_s} ||a - b|| + \sum_{b \in B_s} min_{a \in A_s} ||a - b|| \right), \quad (2)$$

$$HD = max_{a \in A_s} (min_{b \in B_s} ||a - b||), \qquad (3)$$

where A_s and B_s are the surfaces of the segmented volumes to be compared, and $||a - b||$ is the Euclidean distance between two points on A and B.

4.1 Magnetic Resonance Imaging

Among the articles reviewed, 19 proposed deep learning models for prostate segmentation on MRI. Nearly half (9) used a 3D pipeline. The largest used dataset had 958 patient scans [37]. The use of public datasets was found in 10 articles, with data from PROMISE12 challenge [38], ASPS13 [39], BWH [40] and ProstateX [41].

Although tempting to make comparisons between prostate DSC achieved by the different methods, it would be unwise to do so, because the datasets used and the data handling varies significantly. It is only fair to make direct comparisons among those studies which were submitted to a standardised competition. Some articles present their score at the PROMISE12 challenge, which uses a combination of metrics for more robust comparisons - this is indicated in Table 1 if included in the article. Table 2 presents the models used by the various authors, their contributions, benefits and limitations of their approaches. In the following paragraphs, there is an in-depth description of the methods used by the various authors and how some innovated in specific features.

From the publishing dates one can see that it was really in 2018 that the use of deep learning (DL) for prostate segmentation started to gain widespread interest, and really took off in 2019. However, some authors had experimented with DL methods before: instead of end-to-end learning, they used deep networks for feature extraction through representation learning and applied those features in traditional methods [44, 45]. Although primitive, these served as examples to the potential of DL for automated segmentation in the male pelvic anatomy. Interestingly, some authors have been more recently again combining DL with traditional machine learning methods: Tan et al. combined a

Table 1 Works found addressing prostate segmentation on MRI based on deep learning models.

Authors	Year	Data Dimension	Segmented OARs	Dataset Size	Dataset Source	Ground-truth by	Prostate DSC	Score on PROMISE12
Yan et al. [42]	2019	2D	-	80	public[a]	-	84.13%	-
Tian et al. [43]	2018	2D	-	140	private + public[a,b]	radiologist	85.0%	-
Liao et al. [44]	2013	2D	-	30	private	not mentioned	86.7%	-
Guo et al. [45]	2017	2D	-	66	private	"physician"	87.1%	-
Drozdzal et al. [46]	2018	2D	-	80	public[a]	-	87.4%	83.02
Zhu et al. [47]	2017	2D	-	81	private	radiologist	88.5%	-
Cheng et al. [48]	2017	2D	-	250	private	radiologist	89.77%	-
Zhu et al. [49]	2018	2D	-	163	private	2 "experts"	92.7%	-
Zabihollahy et al. [50]	2019	2D	-	225	private	4 radiologists	92.96%	-
Geng et al. [51]	2019	2D	-	130	public[a,b]	-	95.4%	-
Tan et al. [52]	2019	3D	-	85	public[a,c]	-	64.6%	-
Zhu et al. [53]	2018	3D	-	81	private	not mentioned	82.1%	-
Karimi et al. [54]	2018	3D	-	75	private + public[a]	"medical expert"	88%	-
Feng et al. [55]	2018	3D	bladder, rectum	70	private	"physician"	90.3%	-
Jia et al. [56]	2019	3D	-	80+60[e]	public[a,b]	-	90.6%	88.59
Taghanaki et al. [37]	2019	3D	-	958	private + public[d]	not mentioned	91%	-
Nie et al. [57]	2019	3D	bladder, rectum	50	private	radiation oncologist	91.4%	86.15
Zhu et al. [58]	2019	3D	-	146	private + public[a,c]	not mentioned	92.54%	89.59
To et al. [59]	2018	3D	-	200+80[e]	private + public[a]	radiologist	95.11% - 89.11%[f]	-

[a] PROMISE12 dataset [38]
[b] ASPS13 dataset [39]
[c] BWH dataset [40]
[d] ProstateX dataset [41]
[e] trained and tested separately
[f] each DSC value for each separate dataset

Table 2 Features, contributions, benefits and limitations of the works which used MRI.

Authors	Model Features	Contributions/Benefits	Limitations
Yan et al. [42]	2D FCN-16, propagation layer based on superpixels	DL plus hand-crafted features. Loss based on F-score, deals with class imbalance	Limited ability to explore 3D. Superpixels increase computational cost.
Tian et al. [43]	2D FCN, transfer learning from the PASCAL VOC dataset	Transfer learning from a dataset unrelated to medical images improves simple model	Very simple model, no additional features.
Liao et al. [44]	Stacked Independent Subspace Analysis Network with Sparse Label propagation	First study employing a DL network to prostate segmentation, used for automatic feature extraction	Dependent on small image patches. Does not use end-to-end learning.
Guo et al. [45]	Stacked Sparse Auto-encoder network with Deformable Model	Deep Auto-encoder network able of hierarchical feature extraction	Does not use end-to-end learning. Dependent on availability of atlases.
Drozdzal et al. [46]	2D FCN, residual connections, additional FCN for preprocessing	Preprocessing fully integrated into low-capacity FCN, acts as data normaliser	Oversegmentation at base/apex of prostate, in part due to being 2D architecture.
Zhu et al. [47]	2D U-net-like with deep supervision	Deep supervision: regularisation and efficient gradient propagation	Only 4 MRI studies used for testing, results have poor statistical power.
Cheng et al. [48]	2D Holistically Nested Network, deep supervision	Consistent segmentations without trimming contours at base or apex of prostate	Requires additional preprocessing, handles one slice at a time.
Zhu et al. [49]	Two 2D classic U-nets, in sequence in a cascade	Multi-structure segmentation, improves compared to similar single network	2D architecture lacks comprehension of spatial continuity.
Zabihollahy et al. [50]	2D classic U-net	2 different MRI modalities; segmentation of prostate substructures	Requires post-processing; images from a single institution retrospective cohort.
Geng et al. [51]	2D encoder-decoder, dense dilated spatial pyramid pooling	Dilated convolutions increase receptive field and process information at different scales	Removed slices with unclear structures and without prostate: overestimates results.
Tan et al. [52]	3D V-net used together with variational methods	Variational methods improve CNN avoiding boundaries in regions of similar intensity	Used low quality and low resolution images: potentially underestimates results.
Zhu et al. [53]	3D encoder-decoder CNN with dense blocks, long connections	Dense blocks allow deeper network without increasing parameters, avoid overfitting	Cropped sub-volumes instead of whole-volume inputs.
Karimi et al. [54]	3D CNN with fully connected layers at the top, predicting a point cloud	Instead of segmentation mask, network predicts organ center coordinates and point cloud, with comparable results to others	Network has many parameters which makes its training difficult requiring strong regularisation.

Table 3 Table 2 continued.

Authors	Model Features	Contributions/Benefits	Limitations
Feng et al. [55]	3D FCN, multi-task learning, regression on intensity maps of each organ, semi-supervised learning	Regression task improves organ boundary distinctions. Semi-supervised learning allows the use of unlabelled data successfully	Requires handcrafted methods for validation of artificial segmentations of unlabelled data.
Jia et al. [56]	3D ResNet encoder-decoder, anisotropic convolutional blocks, pyramid convolutional blocks, adversarial training	Anisotropic resolution can exploit the 3D context information, adversarial learning confers regularisation and spatial consistency in the segmentations	Authors refer need to perform semi- or weakly supervised learning and use GANs for data augmentation to overcome limitation of small number of training samples.
Taghanaki et al. [37]	3D auto-encoder, new "Combo" loss function	Loss leverages cross-entropy and dice coefficient, weights can be varied	More hyperparameters, in loss function: fine-tuning required.
Nie et al. [57]	3D U-net, stochastic residual units, dilated convolutions, spatially varying convolutional layer, adversarial training	Combining useful network modules increase receptive field, adapt network to regions of interest, and increase consistency in segmentations by adversarial training	Network has only three downsampling levels.
Zhu et al. [58]	3D SNet, encoder-decoder, with densely connected residual blocks, boundary-weighted knowledge transfer, boundary-weighted loss	Adversarial learning to enhance transfer learning, propose a new loss function focused on object edges, which improves performance	Requires additional outside data for training.
To et al. [59]	3D U-net like, dense connections in the encoder, multi-path residual blocks in the decoder	Combining dense, residual and multi-path connections, keep the number of parameters low and fast training time	Extensive post-processing, potentially hindering generalisation ability.

deep neural network with variational methods and improved the CNN-only segmentation on low-resolution, low-quality images [52].

Data preprocessing, including cropping, rescaling and histogram normalisation, is commonly used and is generally beneficial [42, 48–50, 59]. Instead of handcrafted methods, one study attempted a learning-based approach for this preprocessing step by training a neural network, achieving positive results [46]. It would be interesting to see more experiments to validate this approach.

Although the Cross-entropy and Dice loss functions have somewhat been established as the most stable and are used by most authors [48, 50, 55, 57, 59], design of a loss function is shown to be significant by some authors who use new functions tailored to organ edges [58] or with tunable hyperparameters which can be adapted to the dataset at hand [37, 42]. This, however, can make generalisation more difficult to achieve.

In order to fit the model to the training data, end-to-end learning through gradient descent-based optimisation is the process most widely used (Stochastic Gradient Descent [53], Adam [55], Adadelta [37, 50]), but some authors experimented with adversarial learning where a discriminator network is used to distinguish ground-truth from artificial model segmentations [56, 57]. By optimising both networks in tandem, when the discriminator is no longer able to make correct distinctions, the segmentation network is properly fitted to the training dataset and able to perform very accurate segmentations. Although hard to train and computationally demanding, the improvements are significant.

Transfer learning, where the network is initialised with pre-trained weights from a different domain dataset, has been applied to prostate segmentation by Tian et al. with success [43]. This was done from a network trained on a dataset of natural images, which was shown to be able to adapt well to MRI images. Zhu et al. also employed knowledge transfer but through a different technique, by using a discriminator network to enforce convergence between feature vectors produced by the segmentation network from two different datasets [58]. They argue that this method allows for more effective knowledge transfer, especially using images of the same imaging modality and similar domain, but with different details such as scanner specifications. It can be a successful strategy to overcome data scarcity typical of medical imaging problems.

Another strategy to tackle dataset limitations is semi-supervised learning, which was employed in one study [55]. This allows for the use of images for which expert segmentations are not available, effectively increasing the amount of data the model is trained on, leading to better results than using only the data for which ground-truth is available. The problem with this approach is that it requires manual validation to determine when the artificial segmentations are good enough to be added to the training set.

The use of deep supervision where the ground-truth labels are provided to the model at various levels of the network - as opposed to only at the output - was shown to improve prostate edge detection and spatial continuity by facilitating gradient propagation [47, 48]. This is presented as a significant

improvement whilst being relatively easy to implement as side-outputs in the segmentation network [48].

Only two studies on MR images attempted segmentation of the organs-at-risk (OARs) besides the prostate. Both of them used a similar strategy: a softmax layer at the top of the convolutional network for multi-class voxel-wise classification [55,57]. One study offers a distinct approach for multi-class, which despite being used for segmentation of a substructure of the prostate (peripheral zone), could potentially be used for OARs [49]. This strategy consists of cascaded U-nets, one network for each structure to be segmented, where the output of the preceding network is used to make the input to the next. Besides, the authors show that using two networks trained together in a cascade also improves the segmentation ability of the first network, due to the features learned by the second network which are propagated backwards.

The use of post-processing rules is explored by some authors such as eliminating segmentations smaller than a certain volume to avoid false positives and filling holes inside larger segmentations to deal with false negatives [50, 59]. Although this strategy can be tuned to the specific data at hand with positive results, it precludes generalisation ability as different datasets have different field-of-view and resolution, and at the base and apex there can be slices where the prostate has a very small volume which would be missed by using such rules.

It is likely that a combination of many of these advances could be employed together in a single framework based on fully convolutional neural networks to improve the state-of-the-art performance. This seems possible because many of the various contributions are not mutually exclusive in their implementation, but might even be synergistic in improving segmentation performance, as already shown for some features such as dense, residual and long skip connections [57–59]. In fact, the study with the best result on the PROMISE12 challenge achieved this by combining several of the improvements that have been employed separately by other authors, and using a new loss function focused on organ boundaries [58].

Arguably the improvement that provides the best performance gain is having a fully three-dimensional architecture. It is intuitive that by only looking at 2D slices, a model would have difficulty understanding the spatial continuity of an object. In theory, a 3D deep learning model can perform better segmentation of volumes, especially at the top and bottom of approximately round objects such as the prostate. This is in general reflected in the scores of the studies reviewed, as can be seen in Table 1.

The most successful studies build upon a common stem - fully convolutional neural network with encoder-decoder architecture similar to the classic U-net - and each provide some improvement by adding new network modules or training strategies. However, there are some studies that seem to deviate and present radically different approaches which arguably deserve more attention: instead of a voxel-wise classification, Karimi et al. designed a neural network that predicts the coordinates of a point cloud in 3D space to predict the prostate volume [54]. Yet, their result is not significantly better and training

is difficult due to the high number of parameters. Also, in contrast to other implementations, Feng et al. employed multi-task learning with a regression task besides voxel classification for segmentation: the regression component predicts the intensity map and boundary location for each organ. Information gained from the regression tasks helped guide the segmentation in the low contrast boundary regions [55]. More than alternate ways to connect neural network components and small improvements, these provide new avenues of research, even if their results are not the best.

4.2 Computed Tomography

Of all reviewed articles, 9 used computed tomography scans as the domain for prostate and OAR segmentation. Two-thirds (six studies) used a 3D pipeline while only 3 used 2D patches or the whole slices for segmentation. No articles made use of public datasets (to the best of the authors' knowledge there are none available), and the largest dataset was comprised of 1114 scans. Contrary to what happened with MRI, more authors attempted to segment both prostate and OARs, including bladder, rectum, femoral heads and penile bulb. This is probably due to the fact that prostate segmentation in CT is more specific to radiation oncology, where OAR segmentation is essential. The main findings are summarized in Table 4 and a brief description of the models, their benefits and limitations are given in Table 5.

Similarly to the studies focused on MRI, the earlier articles which performed prostate segmentation on CT used a combination of DL and more traditional methods (multi-atlas fusion [60] and level-set method [62]). Their findings established the usefulness of Deep Learning for this task and lead the way to more advanced, better performance models.

Computed tomography images have significantly less contrast than magnetic resonance images, particularly for soft-tissue organs such as those in the pelvic cavity. The boundaries between the prostate, bladder and rectum are sometimes difficult to define, especially where they touch or the edges seem to vanish. In order to overcome this problem, Zhou et al. employed several long pathways between the encoder and decoder arms of their U-net-like model making use of dilated convolutions, a distilling path and also a regression task focused on the boundaries. They were able to extract information at different resolution scales. For the same problem, some authors adopted Attention gates commonly used in Recurrent neural networks [69] and incorporated them into the long skip connections of their segmentation models [63,68]. It is difficult to compare these approaches as their models have other differences and were not applied to the same dataset, but it would be useful to have direct qualitative comparisons.

To further address the low contrast of CT, in clinical practice MRIs are often used to help with manual contouring, mainly after registration with the CT images. Dong et al. designed a Cycle Generative Adversarial Network (CycleGAN) to generate synthetic MR images (sMRI) from the input CT

Table 4 Works found addressing prostate segmentation on CT images based on deep learning.

Authors	Year	Data Dimension	Segmented OARs	Dataset Size	Dataset Source	Ground-truth by	Prostate DSC
Ma et al. [60]	2017	2D	-	92	private	radiologist	86.8%
Zhou et al. [61]	2019	2D	bladder, rectum	339	private	not mentioned	88.4%
Shi et al. [62]	2017	2D	-	22	private	"physicians"	88.8%
Dong et al. [63]	2019	3D	bladder, rectum	140	private	"physicians"	87%
Liu et al. [64]	2019	3D	-	1114	private	5 radiation oncologists[a]	88%
He et al. [65]	2018	3D	bladder, rectum, femoral heads	313	private	2 radiation oncologists[b]	89%
Wang et al. [66]	2019	3D	bladder, rectum, femoral heads	313	private	2 radiation oncologists[b]	89%
Balagopal et al. [67]	2018	3D	bladder, rectum, femoral heads	136	private	radiation oncologist	90%
Kearney et al. [68]	2019	3D	bladder, rectum, penile bulb	120	private	radiation oncologist	90.02%

[a] for the test set, every expert contoured every scan and a consensus segmentation was calculated
[b] every expert contoured every scan and an average segmentation was calculated

Table 5 Features, contributions, benefits and limitations of the works which used CT.

Authors	Model Features	Contributions/Benefits	Limitations
Ma et al. [60]	2D simple CNN for feature extraction from patches, to input into atlas-based model	Combining CNN for feature extraction with atlas-based model is better than either separately	Very small 2D patches, very simple CNN only used for feature extraction, dependent on existing atlases.
Zhou et al. [61]	2D U-net-like, long skip connections, distilling path, dilated convolutions, regression task on organ boundaries	Combination of pathways and difficulty-aware custom loss function improve segmentation of touching, blurry and vanishing organ boundaries	Poor results in images with noise from low dose CT, metal or motion artefacts.
Shi et al. [62]	2D simple FCN based on LeNet for feature extraction from patches, to input into level-set method, domain adaptation	Novel method for domain adaptation, transferring knowledge from original CT images to planning CTs; Patch-to-patch achieves better results than patch-to-scalar	Very small FCN, with only two levels; Low number of training samples; No end-to-end learning.
Dong et al. [63]	3D U-net with deep supervision and Attention, plus a CycleGAN to generate synthetic MRI	Generate synthetic MR images from CT; attention gates capture nuanced organ boundaries	Requires co-registered CT and MR images for training, errors in registration are propagated to the segmentation model.
Liu et al. [64]	3D V-net; patch-based data augmentation technique for initial prostate localisation	Largest dataset; manual contours by several experts; decrease inter-observer variation in ground-truth segmentations	Simple V-net; poor results in images with metallic artefacts; did not attempt multi-organ segmentation.
He et al. [65]	3D U-net-like FCN; distinctive curves for organ surface localization; multi-task learning	One network for each organ, weighted majority voting to solve overlaps; distinct curves to overcome organ shape variations	Often has undersegmentation in small indents and outdents in the organs.
Wang et al. [66]	3D modified U-net; multi-task learning; small FCN for initial organ localisation; boundary sensitive representation model	Each voxel is given a soft-label according to proximity to boundary - easy and hard voxels; multi-label custom loss function	For unseen images, soft-labels can not be calculated as the boundaries are not yet known (no manual segmentation available).
Balagopal et al. [67]	3D modified U-net, backbone based on ResNeXt. 2D U-net used for initial localisation	ResNeXt blocks achieved higher DSC than DenseNet, ResNet or classic U-net backbones	Arbitrary cropping margin of 15% specific to the dataset used, restricts generalisation; lower DSC for rectum.
Kearney et al. [68]	3D FCN, Attention gates, deep supervision, incremental channel boosting, data augmentation	Combination of modules achieves very good performance in easy to train model; only article to segment penile bulb	Uses manual contours from a single observer; predictions are negatively impacted by metal artefacts,

[63]. The organ segmentations were then performed on the sMRI, with better results than applying the same model to the CTs of the same patients. A drawback of this method is that it requires previously co-registered CT and MR images for training the CycleGAN, but it is impressive that this model can overcome low contrast and vanishing boundaries by synthesising simulated Magnetic Resonance images.

Besides low contrast and blurry edges between organs, CT of the male pelvis is notorious for large organ shape and appearance variations. In theory, by accurately identifying whole-organ shape variations one would already solve the vanishing boundaries problem. He et al. propose a solution based on distinctive curves, a morphological representation of the surface of each organ, which are incorporated into an FCN to guide the segmentation [65]. This method is successful in round or tubular structures as are the prostate, bladder and rectum. However, by using specific points to find the distinctive curve for each organ (the most anterior, posterior, left, right and center points), they often run into undersegmentation if there are small indents or outdents not captured by in the curve model.

To perform multi-organ segmentation, and similarly to what was seen in MRI studies, most authors use a softmax function at the top of network [61, 63]. One study used this strategy to also segment the penile bulb, a notoriously difficult structure to localise on CT due to its small size and very low contrast difference to the background - achieved a DSC of 72.21% for this structure [68]. Some authors achieve segmentation of the various organs by training three separate networks [65,66]. This often leads to overlaps in some voxels, which Wang et al. tackle by using a max function getting the classification from the highest confidence model, but He et al. use a weighted majority voting method: a weighting factor is multiplied with the network probability. The prostate segmentation network has larger weighting than bladder or rectum. Thus, if the network confidences for a given voxel is the same for the three networks, the voxel will be classified as prostate. This has the potential drawback of oversegmentation of the prostate, which the authors try to overcome by fine-tuning the weighting of each network [65].

Data handling for training, validation and testing is crucial and impacts the performance of deep learning models greatly. Likewise, any model is only as good as the quality of the underlying ground-truth labels used for its training. In this regard, most of the studies in this review are hindered by using manual contours mostly from a single expert, since this will imprint their personal preferences and biases into the models. Some studies were able to overcome this problem by having manual contours performed by different physicians in the training set, helping cancel out inter-observer variation [64–66]. Liu et al. quantified the benefit of this approach by comparing model performance on a test-set with manual segmentations by one single expert to another test-set with a consensus segmentation from 5 experts [64]. The same trained model performed better on the consensus test-set with smaller standard deviation (DSC of 0.88 ± 0.03 vs 0.85 ± 0.04). This shows that the model was able to combine knowledge from the set of different specialists, thus achieving better

generalisation ability. Any medical imaging dataset to be used for training robust deep learning models for segmentation should take this finding in consideration.

Although one of the advantages of deep learning models for segmentation is that they are fully automatic, requiring no intervention, it seems that performance can be improved by having some method to improve initial organ localisation [64,65]. This allows for a smaller, higher-resolution sub volume to be extracted and used by the segmentation network instead of the whole imaging volume where the vast majority of the image is useless for the segmentation task. To overcome this problem, Liu et al. experiment with two possible solutions: a patch-based data augmentation method and using a smaller localisation CNN preceding the larger segmentation network (this latter method is also used by other authors [65–67]). The results are similar for both approaches, but the 2-network method shows higher variation in distance metrics, particularly in the superior-inferior direction [64].

5 Discussion and conclusion

As evidenced in the Results section, there are several studies proposing different deep learning architectures for radiotherapy planning for prostate cancer. Research is booming in this field, proven by the fact that most of the articles reviewed were published in the last year. Although CT is of utmost importance for RT, it is MRI which has the most attention by the research community. This is perhaps explained by acknowledging that segmentation on MR images has more uses, from diagnosis to follow-up, and for more medical specialties than the niche use of CT segmentation for radiation oncology. Likewise, there are several publicly available datasets for prostate segmentation in MRI but none in CT. This presents an obvious hindrance to research and should be addressed as soon as possible.

The published articles in CT show smaller variance in prostate DSC than those in MRI. Although at first this seems noteworthy, it is most likely coincidental. The various models are widely different, and have been applied to different datasets, in number of patients, image quality and resolution. If all models were applied on the same standardised dataset, it is expected that a model such as a combination of a very simple CNN with traditional ML [62] would perform much worse than a complex three-dimensional fully convolutional network with improved modules [67]. In fact, Zhou et al. present a comparison study showing that a classic U-net applied to the same dataset as their proposed model approaches 60% as to DSC in the harder cases [61]. Likewise, a classic U-net was also implemented by Kearney et al. achieving a mean DSC of 84.13% on the prostate (versus 90.02% of the proposed model) [68]. This is another evidence that direct quantitative comparisons between models applied on different datasets should be made carefully until the models can be tested on the same dataset.

Segmentation of the prostate alone has gained attraction by the research community, evidenced by the importance of some older challenges still active nowadays, but there has been a low investment in segmentation of both radiation targets and OARs, specifically for RT planning. Possible reasons for this may be lack of access to RT planning images (as opposed to diagnostic images), and difficulty in getting expert segmentations of every structure required (as opposed to only the prostate itself).

In line with semantic segmentation of natural images for tasks other than those related to the medical field, it can be seen that researchers are following the most recent advances in deep learning with techniques such as dense connections, dilated convolutions, spatially varying convolutions, anisotropic convolutions, attention gates, multi-task learning, cost function engineering, generative networks and adversarial training applied to the encoder-decoder architecture that has become the standard for medical imaging segmentation. Combination of several of these methods in a single model designed for end-to-end learning seems promising as these may be synergistic in the performance gains.

These methods are usually compared through geometric comparisons in the form of overlap and boundary metrics, but the dosimetric impact of these differences in the radiotherapy treatment plan is essential for future incorporation of these computational models in clinical practice, with the overarching goal of improving quality and efficiency in healthcare.

The review by Moore [8] about automated treatment planning in radiation therapy ends with a vision of the future where computers automatically calculate a deliverable dose distribution within a second after the physician completes their last contour. In the current review, we have realized that in that future the physician might not even need to contour the volumes at all, instead performing expert validation of the segmentations and the radiation doses all in one go, right after the planning CT is acquired, allowing the treatment to start earlier.

References

1. American Cancer Society. Facts & Figures 2019. Technical report, American Cancer Society, 2019.
2. Islam Reda, Ashraf Khalil, Mohammed Elmogy, Ahmed Abou El-Fetouh, Ahmed Shalaby, Mohamed Abou El-Ghar, Adel Elmaghraby, Mohammed Ghazal, and Ayman El-Baz. Deep learning role in early diagnosis of prostate cancer. *Technol Cancer Res Treat*, 17:1–11, 2018.
3. AM Noone, N Howlader, M Krapcho, D Miller, A Brest, M Yu, J Ruhl, Z Tatalovich, A Mariotto, DR Lewis, HS Chen, EJ Feuer, and KA Cronin. SEER cancer statistics review. Technical report, National Cancer Institute, 2017.
4. Milad Zafar Nezhad, Najibesadat Sadati, Kai Yang, and Dongxiao Zhu. A Deep Active Survival Analysis approach for precision treatment recommendations: Application of prostate cancer. *Expert Syst Appl*, 115:16–26, 2019.
5. Joanna Kazmierska, Núria Jornet Sala, Michelle Leech, Barbara Alicja Jereczek-Fossa, Yolande Lievens, and John Yarnold. Radiotherapy: Seizing the opportunity in cancer care. Technical report, ESTRO Cancer Foundation, 2018.
6. Josep M. Borras, Yolande Lievens, Michael Barton, Julieta Corral, Jacques Ferlay, Freddie Bray, and Cai Grau. How many new cancer patients in Europe will require radiotherapy by 2025? An ESTRO-HERO analysis. *Radiother Oncol*, 119(1):5–11, 2016.
7. Josep M. Borras, Yolande Lievens, Peter Dunscombe, Mary Coffey, Julian Malicki, Julieta Corral, Chiara Gasparotto, Noemie Defourny, Michael Barton, Rob Verhoeven, Liesbeth Van Eycken, Maja Primic-Zakelj, Maciej Trojanowski, Primoz Strojan, and Cai Grau. The optimal utilization proportion of external beam radiotherapy in European countries: An ESTRO-HERO analysis. *Radiother Oncol*, 116(1):38–44, 2015.
8. Kevin L. Moore. Automated Radiotherapy Treatment Planning. *Semin Radiat Oncol*, 29(3):209–218, 2019.
9. B Emami, J Lyman, A Brown, L Cola, M Goitein, J. E Munzenrider, B Shank, L. J Solin, and M Wesson. Tolerance of normal tissue to therapeutic irradiation. *Int J Radiat Oncol Biol Phys*, 21(1):109–122, 1991.
10. Gisele C. Pereira, Melanie Traughber, and Raymond F. Muzic. The role of imaging in radiation therapy planning: Past, present, and future. *Biomed Res Int*, 2014(2), 2014.
11. Philippe Meyer, Vincent Noblet, Christophe Mazzara, and Alex Lallement. Survey on deep learning for radiotherapy. *Comput Biol Med*, 98(May):126–146, 2018.
12. HA Gay, HJ Barthold, E O'Meara, WR Bosch, IE Naga, R Al-Lozi, SA Rosenthal, C Lawton, WR Lee, H Sandler, A Zietman, R Myerson, LA Dawson, C Willett, LA Kachnic, A Jhingran, L Portelance, J Ryu, W Small, D Gaffney, AN Viswanathan, and JF Michalski. Male Pelvis Normal Tissue - RTOG Consensus Contouring Guidelines. Technical report, Radiation Therapy Oncology Group Foundation, 2007.
13. Claudio Fiorino, Michele Reni, Angelo Bolognesi, Giovanni Mauro Cattaneo, and Riccardo Calandrino. Intra- and inter-observer variability in contouring prostate and seminal vesicles: Implications for conformal treatment planning. *Radiother Oncol*, 47(3):285–292, 1998.
14. Zhanrong Gao, David Wilkins, Libni Eapen, Christopher Morash, Youssef Wassef, and Lee Gerig. A study of prostate delineation referenced against a gold standard created from the visible human data. *Radiother Oncol*, 85(2):239–246, 2007.
15. JL Mohler, S Srinivas, ES Antonarakis, AJ Armstrong, AV D'Amico, BJ Davis, and T Dorff. Prostate Cancer NCCN Guidelines Version 4.1029. Technical report, National Comprehensive Cancer Network, 2019.
16. The Royal College of Radiologists, Society of Radiographers, College, Institute of Physics in Medicine, and Engineering. On target: ensuring geometric accuracy in radiotherapy. Technical report, The Royal College of Radiologists, 2008.
17. Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep Learning in Medical Image Analysis. *Annu Rev Biomed Eng*, 19:221–248, 2017.
18. Rodrigo Suarez-Ibarrola, Simon Hein, Gerd Reis, Christian Gratzke, and Arkadiusz Miernik. Current and future applications of machine and deep learning in urology: a review of the literature on urolithiasis, renal cell carcinoma, and bladder and prostate cancer. *World J Urol*, (Ml), 2019.

19. Luca Boldrini, Jean-Emmanuel Bibault, Carlotta Masciocchi, Yanting Shen, and Martin-Immanuel Bittner. Deep Learning: A Review for the Radiation Oncologist. *Front Oncol*, 9(October), 2019.
20. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
21. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *Adv Neural Inf Process Syst*, 15:474–483, 2014.
22. Yann LeCun, Bernhard E Boser, John S Denker, Donnie Henderson, R E Howard, Wayne E Hubbard, and Lawrence D Jackel. Handwritten Digit Recognition with a Back-Propagation Network. In D S Touretzky, editor, *Adv Neural Inf Process Syst 2*, pages 396–404. Morgan-Kaufmann, 1990.
23. Lequan Yu, Xin Yang, Hao Chen, Jing Qin, and Pheng-Ann Heng. Volumetric ConvNets with Mixed Residual Connections for Automated Prostate Segmentation from 3D MR Images. In *Thirty-First AAAI Conf Artif Intell*, pages 66–72, 2017.
24. Evan Shelhamer, Jonathon Long, and Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans Pattern Anal Mach Intell*, 39:1, 2016.
25. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Nassir Navab, Joachim Hornegger, William M Wells, and Alejandro F Frangi, editors, *Med Image Comput Comput Interv – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.
26. Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-net: Learning dense volumetric segmentation from sparse annotation. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*, 9901 LNCS:424–432, 2016.
27. Fausto Milletari, Nassir Navab, and Seyed Ahmad Ahmadi. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. *Proc - 2016 4th Int Conf 3D Vision, 3DV 2016*, pages 565–571, 2016.
28. Samaneh Kazemifar, Anjali Balagopal, Dan Nguyen, Sarah McGuire, Raquibul Hannan, Steve Jiang, and Amir Owrangi. Segmentation of the prostate and organs at risk in male pelvic CT images using deep learning. *Biomed Phys Eng Express*, 4(5):55003, 2018.
29. Ozan Oktay, Jo Schlemper, Loic Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention U-Net: Learning Where to Look for the Pancreas. *1st Conf Med Imaging with Deep Learn (MIDL 2018)*, 2018.
30. Carlos E. Cardenas, Jinzhong Yang, Brian M. Anderson, Laurence E. Court, and Kristy B. Brock. Advances in Auto-Segmentation. *Semin Radiat Oncol*, 29(3):185–197, 2019.
31. Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Med Image Anal*, 42(December 2012):60–88, 2017.
32. Wenya Linda Bi, Ahmed Hosny, Matthew B. Schabath, Maryellen L. Giger, Nicolai J. Birkbak, Alireza Mehrtash, Tavis Allison, Omar Arnaout, Christopher Abbosh, Ian F. Dunn, Raymond H. Mak, Rulla M. Tamimi, Clare M. Tempany, Charles Swanton, Udo Hoffmann, Lawrence H. Schwartz, Robert J. Gillies, Raymond Y. Huang, and Hugo J. W. L. Aerts. Artificial Intelligence in Cancer Imaging: Clinical Challenges and Applications. *CA Cancer J Clin*, 0(0):1–31, 2019.
33. Berkman Sahiner, Aria Pezeshk, Lubomir M. Hadjiiski, Xiaosong Wang, Karen Drukker, Kenny H. Cha, Ronald M. Summers, and Maryellen L. Giger. Deep learning in medical imaging and radiation therapy. *Med Phys*, 46(1):e1–e36, 2019.
34. Ian Boon, Tracy Au Yong, and Cheng Boon. Assessing the Role of Artificial Intelligence (AI) in Clinical Oncology: Utility of Machine Learning in Radiotherapy Target Volume Delineation. *Medicines*, 5(4):131, 2018.
35. Renato Cuocolo, Maria Brunella Cipullo, Arnaldo Stanzione, Lorenzo Ugga, Valeria Romeo, Leonardo Radice, Arturo Brunetti, and Massimo Imbriaco. Machine learning applications in prostate cancer magnetic resonance imaging. *Eur Radiol Exp*, 3(1), 2019.
36. S. Larry Goldenberg, Guy Nir, and Septimiu E. Salcudean. A new era: artificial intelligence and machine learning in prostate cancer. *Nat Rev Urol*, 16(7):391–403, 2019.

37. Saeid Asgari Taghanaki, Yefeng Zheng, S. Kevin Zhou, Bogdan Georgescu, Puneet Sharma, Daguang Xu, Dorin Comaniciu, and Ghassan Hamarneh. Combo loss: Handling input and output imbalance in multi-organ segmentation. *Comput Med Imaging Graph*, 75:24–33, 2019.

38. Geert Litjens, Robert Toth, Wendy van de Ven, Caroline Hoeks, Sjoerd Kerkstra, Bram van Ginneken, Graham Vincent, Gwenael Guillard, Neil Birbeck, Jindang Zhang, Robin Strand, Filip Malmberg, Yangming Ou, Christos Davatzikos, Matthias Kirschner, Florian Jung, Jing Yuan, Wu Qiu, Qinquan Gao, Philip Eddie Edwards, Bianca Maan, Ferdinand van der Heijden, Soumya Ghose, Jhimli Mitra, Jason Dowling, Dean Barratt, Henkjan Huisman, and Anant Madabhushi. Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge. *Med Image Anal*, 18(2):359–373, 2014.

39. Nicholas Bloch, Anant Madabhushi, Henkjan Huisman, John Freymann, Justin Kirby, Michael Grauer, Andinet Enquobahrie, Carl Jaffe, Larry Clarke, and Keyvan Farahani. NCI-ISBI 2013 Challenge: Automated Segmentation of Prostate Structures, 2015.

40. The Brigham Hospital and Women's. BWH Prostate MR Image Database, 2008.

41. Geert Litjens, Oscar Debats, Jelle Barentsz, Nico Karssemeijer, and Henkjan Huisman. ProstateX Challenge data, 2017.

42. Ke Yan, Xiuying Wang, Jinman Kim, Mohamed Khadra, Michael Fulham, and Dagan Feng. A propagation-DNN: Deep combination learning of multi-level features for MR prostate segmentation. *Comput Methods Programs Biomed*, 170:11–21, 2019.

43. Zhiqiang Tian, Lizhi Liu, Zhenfeng Zhang, and Baowei Fei. PSNet: prostate segmentation on MRI based on a convolutional neural network. *J Med Imaging*, 5(02):1, 2018.

44. Shu Liao, Yaozong Gao, Aytekin Oto, and Dinggang Shen. Representation learning: A unified deep learning framework for automatic prostate MR segmentation. *Med Image Comput Comput Assist Interv*, 16(02):254–261, 2013.

45. Yanrong Guo, Yaozong Gao, and Dinggang Shen. Deformable MR Prostate Segmentation via Deep Feature Learning and Sparse Patch Matching. *Deep Learn Med Image Anal*, 0062(c):197–222, 2017.

46. Michal Drozdzal, Gabriel Chartrand, Eugene Vorontsov, Mahsa Shakeri, Lisa Di Jorio, An Tang, Adriana Romero, Yoshua Bengio, Chris Pal, and Samuel Kadoury. Learning normalized inputs for iterative estimation in medical image segmentation. *Med Image Anal*, 44:1–13, 2018.

47. Qikui Zhu, Bo Du, Baris Turkbey, Peter L. Choyke, and Pingkun Yan. Deeply-supervised CNN for prostate segmentation. *Proc Int Jt Conf Neural Networks*, 2017-May:178–184, 2017.

48. Ruida Cheng, Holger R. Roth, Nathan Lay, Le Lu, Baris Turkbey, William Gandler, Evan S. McCreedy, Peter Choyke, Ronald M. Summers, and Matthew J. McAuliffe. Automatic MR prostate segmentation by deep learning with holistically-nested networks. *Med Imaging 2017 Image Process*, 10133(4):101332H, 2017.

49. Yi Zhu, Rong Wei, Ge Gao, Lian Ding, Xiaodong Zhang, Xiaoying Wang, and Jue Zhang. Fully automatic segmentation on prostate MR images based on cascaded fully convolution network. *J Magn Reson Imaging*, 49(4):1149–1156, 2018.

50. Fatemeh Zabihollahy, Nicola Schieda, Satheesh Krishna Jeyaraj, and Eranga Ukwatta. Automated segmentation of prostate zonal anatomy on T2-weighted (T2W) and apparent diffusion coefficient (ADC) map MR images using U-Nets. *Med Phys*, 46(7):3078–3090, 2019.

51. Lei Geng, Jia Wang, Zhitao Xiao, Jun Tong, Fang Zhang, and Jun Wu. Encoder-decoder with dense dilated spatial pyramid pooling for prostate MR images segmentation. *Comput Assist Surg*, 24(sup2):13–19, 2019.

52. Lu Tan, Antoni Liang, Ling Li, Wanquan Liu, Hanwen Kang, and Chao Chen. Automatic prostate segmentation based on fusion between deep network and variational methods. *J Xray Sci Technol*, pages 1–17, 2019.

53. Qikui Zhu, Bo Du, Jia Wu, and Pingkun Yan. A Deep Learning Health Data Analysis Approach: Automatic 3D Prostate MR Segmentation with Densely-Connected Volumetric ConvNets. *Proc Int Jt Conf Neural Networks*, 2018-July, 2018.

54. Davood Karimi, Golnoosh Samei, Claudia Kesch, Guy Nir, and Septimiu E. Salcudean. Prostate segmentation in MRI using a convolutional neural network architecture and training strategy based on statistical shape models. *Int J Comput Assist Radiol Surg*, 13(8):1211–1219, 2018.

55. Zishun Feng, Dong Nie, Li Wang, and Dinggang Shen. Semi-supervised learning for pelvic MR image segmentation based on multi-task residual fully convolutional networks. *Proc - Int Symp Biomed Imaging*, 2018-April(Isbi):885–888, 2018.
56. Haozhe Jia, Yong Xia, Yang Song, Donghao Zhang, Heng Huang, Yanning Zhang, and Weidong Cai. 3D APA-Net: 3D Adversarial Pyramid Anisotropic Convolutional Network for Prostate Segmentation in MR Images. *IEEE Trans Med Imaging*, PP(c):1–1, 2019.
57. Dong Nie, Li Wang, Yaozong Gao, Jun Lian, and Dinggang Shen. STRAINet: Spatially Varying sTochastic Residual AdversarIal Networks for MRI Pelvic Organ Segmentation. *IEEE Trans Neural Networks Learn Syst*, 30(5):1552–1564, 2019.
58. Qikui Zhu, Bo Du, and Pingkun Yan. Boundary-weighted Domain Adaptive Neural Network for Prostate MR Image Segmentation. *IEEE Trans Med Imaging*, pages 1–1, 2019.
59. Minh Nguyen Nhat To, Dang Quoc Vu, Baris Turkbey, Peter L. Choyke, and Jin Tae Kwak. Deep dense multi-path neural network for prostate segmentation in magnetic resonance imaging. *Int J Comput Assist Radiol Surg*, 13(11):1687–1696, 2018.
60. Ling Ma, Rongrong Guo, Guoyi Zhang, Funmilayo Tade, David M. Schuster, Peter Nieh, Viraj Master, and Baowei Fei. Automatic segmentation of the prostate on CT images using deep learning and multi-atlas fusion. *Med Imaging 2017 Image Process*, 10133:101332O, 2017.
61. Sihang Zhou, Dong Nie, Ehsan Adeli, Jianping Yin, Jun Lian, and Dinggang Shen. High-Resolution Encoder-Decoder Networks for Low-Contrast Medical Image Segmentation. *IEEE Trans Image Process*, 29(X):461–475, 2019.
62. Yinghuan Shi, Wanqi Yang, Yang Gao, and Dinggang Shen. Does Manual Delineation only Provide the Side Information in CT Prostate Segmentation? In M. Descoteaux, editor, *MICCAI 2017, Part III, LNCS 10435*, volume 10435, pages 692–700. Springer International Publishing, 2017.
63. Xue Dong, Yang Lei, Sibo Tian, Tonghe Wang, Pretesh Patel, Walter J. Curran, Ashesh B. Jani, Tian Liu, and Xiaofeng Yang. Synthetic MRI-aided multi-organ segmentation on male pelvic CT using cycle consistent deep attention network. *Radiother Oncol*, pages 1–8, 2019.
64. Chang Liu, Stephen J. Gardner, Ning Wen, Mohamed A. Elshaikh, Farzan Siddiqui, Benjamin Movsas, and Indrin J. Chetty. Automatic Segmentation of the Prostate on CT Images Using Deep Neural Networks (DNN). *Int J Radiat Oncol Biol Phys*, 104(4):924–932, 2019.
65. Kelei He, Xiaohuan Cao, Yinghuan Shi, Dong Nie, Yang Gao, and Dinggang Shen. Pelvic Organ Segmentation Using Distinctive Curve Guided Fully Convolutional Networks. *IEEE Trans Med Imaging*, 38(2):585–595, 2019.
66. Shuai Wang, Kelei He, Dong Nie, Sihang Zhou, Yaozong Gao, and Dinggang Shen. CT male pelvic organ segmentation using fully convolutional networks with boundary sensitive representation. *Med Image Anal*, 54:168–178, 2019.
67. Anjali Balagopal, Samaneh Kazemifar, Dan Nguyen, Mu-Han Lin, Raquibul Hannan, Amir Owrangi, and Steve Jiang. Fully automated organ segmentation in male pelvic CT images. *Phys Med Biol*, 63(24), 2018.
68. Vasant Kearney, Jason W Chan, Tianqi Wang, Alan Perry, Sue S Yom, and Timothy D Solberg. Attention-enabled 3D boosted convolutional neural networks for semantic CT segmentation using deep supervision. *Phys Med Biol*, 64(13):135001, 2019.
69. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.