

U-Net Architectures for Prostate Cancer Radiation Therapy: A Literature Review

Bruno Mendes^{1,2}, Inês Domingues^{2,3}, João Santos^{2,4}

¹Faculdade de Engenharia, Universidade do Porto (FEUP), Rua Dr. Roberto Frias, 4200-465 PORTO, Portugal (up200004330@up.pt) ORCID [0000-0002-7574-7630](https://orcid.org/0000-0002-7574-7630)

²Centro de Investigação do Instituto Português de Oncologia do Porto (CI-IPOP) - Física Médica, Radiobiologia e Protecção Radiológica, Rua Dr. António Bernardino de Almeida, 4200-072 PORTO, Portugal (joao.santos@ipoporto.min-saude.pt) ORCID [0000-0002-2334-7280](https://orcid.org/0000-0002-2334-7280)

³Politécnico de Coimbra, ISEC, DEIS, Rua Pedro Nunes - Quinta da Nora, 3030-199 COIMBRA, Portugal, (inesdomingues@gmail.com) ORCID [0000-0003-2465-5143](https://orcid.org/0000-0003-2465-5143)




⁴Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Rua Jorge de Viterbo Ferreira 228, 4050-313 PORTO, Portugal

Abstract

Silently asymptomatic, PCa is usually diagnosed with Digital Rectal Examination (DRE) and Prostate Specific Antigen (PSA) levels. Since the first treatment of an advanced prostatic malignancy with X-rays by Imbert and Imbert in 1904, External Beam Radiation Therapy (EBRT) is now a curative option for localised and locally advanced disease and a palliative option for the metastatic low-volume disease. Even with the introduction of computers in EBRT and better imaging techniques, volume delineation is still a very time-consuming task relying on manual or semi-automatic segmentation techniques. On the other hand, the U-Net architecture was specially designed for medical image segmentation presenting promising results. This literature review gathers work using U-Net architectures for PCa segmentation in an EBRT context. Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) standards, we outline methods, techniques and obtained outcomes as a potential foundation for an automated segmentation framework for PCa.

Author Keywords. U-Net, Prostate-Cancer, Segmentation, Radiation Therapy

Type: Research Article

 Open Access  Peer Reviewed  CC BY

1. Introduction

In 1853, John Adams, a surgeon at the London Hospital, diagnosed a cirrhosis of the prostate gland as an orphan disease, the first described case of Prostate Cancer (PCa). In 2020, it was responsible for 7.3% of all cancer deaths in men and was the second most frequent malignancy. Silently asymptomatic in an early stage, PCa is usually diagnosed by Digital Rectal Examination (DRE) and Prostate Specific Antigen (PSA) blood test. The prostate gland is about the size of a walnut, located in the pelvis surrounding the prostatic urethra and below the bladder. Usually, PCa originates from the peripheral zone of the prostate, adjacent to the rectum (Singh and Bolla 2019). **Error! Reference source not found. Error! Reference source not found. Error! Reference source not found.**

The typical PCa patient in the 1940's was a man in his seventies diagnosed with bone or soft tissue metastases. Such a poorly differentiated lesion was a death sentence. The first effective systematic treatment for any cancer was firstly proposed by Charles Huggins. He found that metastatic PCa responded well to androgen-ablation therapy (Denmeade and Isaacs 2002). Since then, cancer treatment techniques have evolved tremendously and so have diagnosing,

staging and grading PCa. Treatment options have therefore become more diverse and adjusted according to the aggressiveness and prognosis of PCa. Current guidelines suggest External Beam Radiotherapy (EBRT) as a curative option for localised and locally advanced disease and as a palliative option for metastatic low-volume disease (Parker et al. 2020).

In the EBRT workflow, patients usually do a Computed Tomography (CT) scan providing the anatomical basics for treatment planning and attenuation coefficients for dose estimations. In this stage, experts define tumour and tissue-related volumes aided by Magnetic Resonance Imaging (MRI), which presents better soft-tissue contrast. Manual volume delineation is definitively a time consuming and error-prone task. Although performed by highly trained medical experts, there is always the issue of inter and intra-observer variability.

EBRT Systems do offer some automated contouring solutions. Most of them are ATLAS based - a library of images, usually online, with multiple delineated organs of different sizes and shapes available to the EBRT system. These contours are then registered (rigid or elastic) to the patient's CT or MRI images but usually require manual corrections. EBRT systems also provide tools for manual segmentation along with semi-automatic methods. These can be isodensity threshold or region growing after a manually placed seed point. Besides, it is possible to propagate the contours to other slices in the volume. The delineated volumes will be the base for radiation dose calculations and optimizations. While the goal is to maximise the therapeutic dose to the Clinical Target Volume (CTV), radiation dose to Organs At Risk (OARs) must be kept to a minimum.

In a recent study, Zabel et al. (2021) compared the impact of deep learning and ATLAS based segmentation methods with the traditional manual segmentation of the bladder and rectum. The initial contours are obtained faster with an ATLAS based method but the editing time is much higher. Overall, ATLAS segmentation methods were consistently slower than manual and deep learning. Deep learning methods provided editing times similar to manual while obtaining initial contours much faster (Zabel et al. 2021). The lack of an automatic contouring solution in a clinical EBRT context is a major constraint increasing the time needed to efficiently plan a treatment. With the good results obtained by the U-Net for medical image segmentation, this work intends to provide a revision of U-Net based architectures for PCa automated segmentation in an EBRT context, enhancing the used methodologies and results. To the best of our knowledge, it is the first literature review focusing on the combination of U-Net architectures and PCa.

Following this introductory section, a brief review of the found U-Net architectures is performed. Section 3 presents the methodology and exclusion criteria following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) standards. In Section 4, the results were grouped considering the following segmentation pipeline: data augmentation, preprocessing techniques, training parameters and finally evaluation metrics. For each, a brief discussion is performed and a summary of the used methods is presented. Finally, Section 5 presents the main conclusions withdrawn and the final remarks.

2. Background

In 1981, David Hubel, Torsten Wiesel and Roger Sperry won the Nobel Prize for Physiology or Medicine for their work in visual neuroscience (Hubel and Wiesel 2012). They established the fundamental understanding of how human neurons extract information along the visual pathway to encode an image. With this in mind, Fukushima and Miyake (1982) created the first artificial neural network mimicking humans' simple and complex cells. In 1989, LeCun et al. (1989) implemented the first modern application of convolutional neural networks.

Specially designed for medical image segmentation, Ronneberger, Fischer, and Brox (2015) won the ISBI 2012 EM segmentation challenge. The proposed architecture follows an encoder-decoder cascade structure with a U-shape. Since then it has been widely used in research especially concerning organ segmentation in cancer patients. One of the main advantages of the U-Net is that it is capable of outputting highly detailed segmentation maps with few images and much faster to train than other models since the learning is context-based (Siddique et al. 2021). In this chapter, we present the core concepts of the main variations in the U-Net architecture found in this review.

2.1. Base U-Net

The base U-Net, as proposed by Ronneberger, Fischer, e Brox 2015, follows an encoder-decoder cascade structure with a U-shape as illustrated in Figure 1, allowing fast and precise image segmentation.

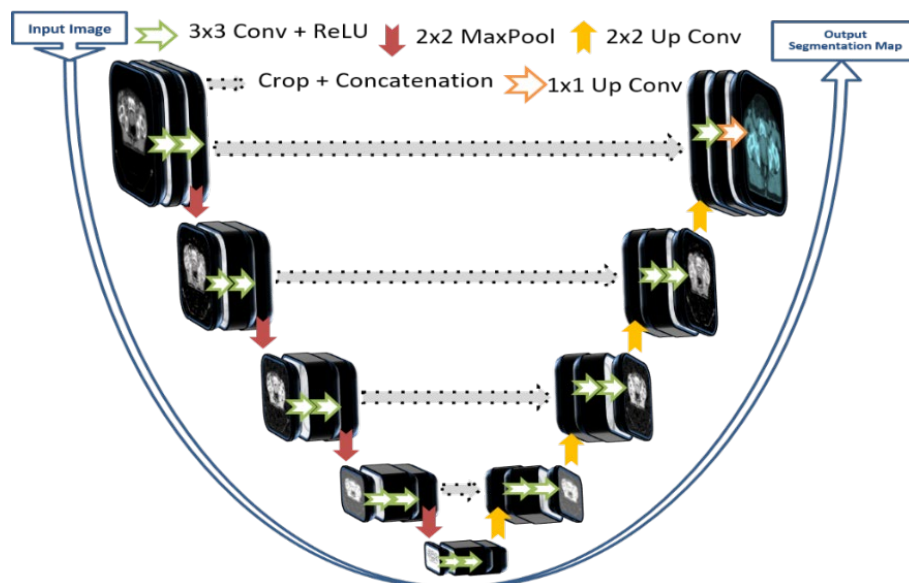


Figure 1: Schematic diagram of the U-Net architecture.

The contracting path consists of the successive application of double 3x3 convolution blocks, each followed by a Rectified Linear Unit (ReLU). Connecting the blocks is a 2x2 max-pooling operation with stride 2, and the number of feature extraction channels is doubled. The level of extracted features increases as the image goes deeper into the net. For example, texture and edges are extracted in the initial layers, while shape and categories are extracted much deeper.

The expansive path is very similar and composed of double convolution blocks. The connection between these blocks is now assured by a 2x2 up-convolution (halving the number of channels). At the same level of depth, the output of each convolution block from the contracting path (encoding arm) is cropped (to assure that sizes match) and concatenated with the corresponding output in the expansive path (decoding arm). These skip connections bring spatial information to a ready to use pixel-level segmentation output.

Attempting to improve performance, researchers have stacked multiple U-Net architectures. However, this raises issues of complexity and exploding gradients. In a novel configuration, Qin et al. (2020) introduced a two-level nested U-Net structure with specially designed residual blocks (U²-Net). This setup allows the network to go deeper, efficiently maintaining high-resolution feature maps (Qin et al. 2020).

2.2. D U-Net

Another downside of the U-Net is that it only allows a 2D input. In medical imaging, most data is volumetric, usually with anisotropic spacing. The 3D U-Net is simply an extension of the base U-Net where every convolution and pooling operation accepts a three-dimensional input. With this architecture, training becomes faster and with sparse labelled pixels (Siddique et al. 2021). A point worth mentioning is that this configuration requires the volume to be isotropic (Çiçek et al. 2016), which is usually not the case with CT or MRI.

A particular 3D U-net variation is the V-Net, introduced by Milletari, Navab, and Ahmadi (2016). The V-Net presents a customised Dice Loss layer that avoids re-weighting and, with a stage split configuration, learns residuals, improving results and convergence.

2.3. Adversarial U-Net

Let us imagine two networks competing against each other to improve the outcome. That is the principle behind Generative Adversarial Networks (GAN). One of the networks, the discriminator, acts as a lie detector. It receives the output of the other network, the generator, and classifies it as a real or fake image. The generator's goal is to produce images capable of deceiving the discriminator. Both gradient functions of the networks are connected. The generator adjusts its weight according to the response of the discriminator and adds random noise to produce a new image. The discriminator seeks to minimise its error rate and the generator to maximise the error rate of the discriminator. The battle ends when the discriminator can no longer distinguish fake from real images (Goodfellow et al. 2014). Applying this strategy to a U-Net requires the imposition of some conditions, restricting the band of synthetic images produced. The generator does not add noise but instead transforms the images mimicking human manual transformations (Siddique et al. 2021).

2.4. Attention U-Net

The Attention U-Net ignores unnecessary areas by placing an attention gate at the end of each skip connection (same depth level concatenation between encoding and decoding paths). Before the concatenation, the attention gate selects the more relevant features by filtering the neuron activations in the forward and backward passes. The Attention U-Net avoids the computational complexity by applying linear transformations without spatial support. Besides, it is deeply supervised to allow semantic discrimination of feature maps (Oktay et al. 2018). Increasing the ability of the U-Net to focus on more relevant features allows for localised classification and draws attention to the segmentation of different objects (Siddique et al. 2021).

Attention is commonly multiplicative or additive. While matrix multiplication provides faster and more memory efficient computations, additive attention provides the best results. Training the network is slightly more complex. The aggregation of feature maps at different scales affects learning and performance. A better approach is to train at each scale separately and then fine-tune (Schlemper et al. 2019).

3. Methodology

The research community has been highly engaged in studying and improving the U-Net architecture for specific use cases. In a quick search in PubMed with the keyword U-Net, we found 1894 results from 2015. On 2021 alone, we found 938 papers. If we narrow the search by adding the keyword Radiotherapy, we obtain only 113 results. Although growing, the need for automated segmentation solutions for EBRT urges to settle. This review intends to provide the baseline segmentation steps for a prostate cancer evaluation framework. This Section provides a systematic review of U-Net architectures used for prostate and OARs

segmentation, focusing on work using clinically confirmed PCa patients. Following the PRISMA standards, this review includes published works from 2018 onwards.

We carried this review in Scopus, PubMed (NIM), and ScienceDirect with the following search terms: “Prostate” And “Segmentation” And “U-Net” within the article title, abstract and keywords in February of 2022. A widely used metric for segmentation quality is the Dice Similarity Coefficient (DSC). For comparison purposes, we included DSC also as a search term. The search outputted 318 results. We removed 49 duplicate entries and carefully analysed the title and abstract of the remaining 268 unique records. Some were not PCa related. Others included ultrasound or microscope images and were excluded (219). From the 49 articles sought for retrieval, we did not have access to 13 of them, and one did not use a U-Net, leaving 35 eligible for this review. After a thorough reading of the full article, we ended up excluding 7 because they aimed at classification issues, 5 concerning the segmentation of prostate zones, 3 for dose optimization in EBRT planning and 4 did not specify DSC as an evaluation metric. [Figure 2](#) shows the PRISMA flow diagram, illustrating the selection steps.

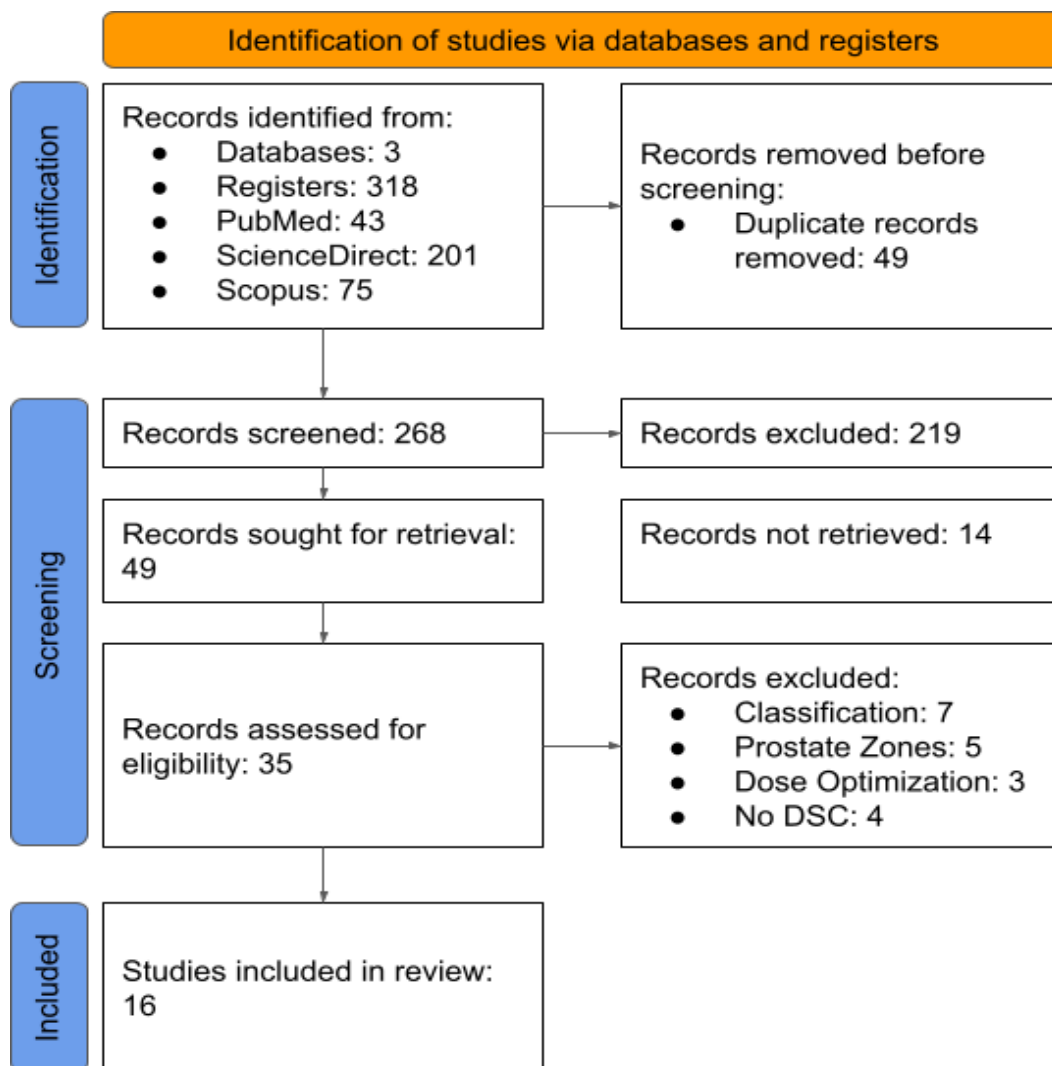


Figure 2: PRISMA Flow Diagram

Most of the work reviewed is based on MRI (10), while the remaining 6 are CT based. In EBRT planning for PCa, those are the preferred imaging modalities. CT provides anatomical background and linear attenuation coefficients for dose estimations and MRI provides soft-tissue contrast for volume delineation. A segmentation framework usually follows a pipeline.

The review process focused on preprocessing methods, augmentation techniques, class balance, training parameters, advantages and limitations of each study. Section 4 presents the results structured following this pipeline to enlighten which methods to use in each step. The main OARs for PCa are the bladder and the rectum. Although not the ultimate purpose, we also retrieved the DSC for those organs, if attempted by the authors. A multiclass strategy is also interesting since it mimics the human holistic volume delineation.

4. Results and Discussion

In all of the reviewed papers, authors attempted segmentation using a U-Net. For some, it was the basis for a fine-tuned network, but for others, it was a comparison baseline. Most of the datasets used are private. Chen et al. (2019) and Khan et al. (2020) used exclusively public images. In spite of this, 11 of them used external datasets for validation and testing purposes. The most used one was the Prostate MR Image Segmentation 2012 (PROMISE12) (Litjens et al. 2014) from the MICCAI Grand Challenge. As for data dimensionality, the reviewed articles are evenly split, with 8 using 2D and another 8 using 3D. Figure 3 shows example images from the public PROMISE12 dataset.

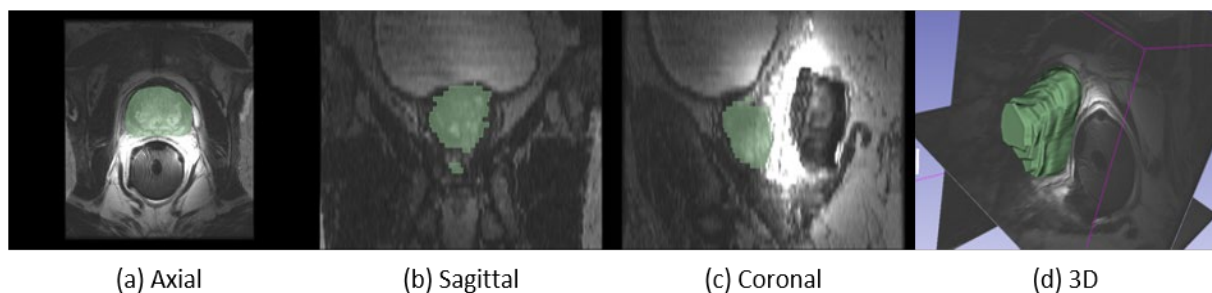


Figure 3: PROMISE12 (MRI) (Litjens et al. 2014). In green: the prostate.

The smallest dataset was from Khan et al. (2020) with 11 patients and the largest from Almeida et al. (2022) with 2226 patients. Tian et al. (2018) showed that fine-tuning only the last three layers of a model may be sufficient to improve the segmentation accuracy of a U-Net and V-Net. Sultana et al. (2020) included GAN in their architecture and Cem Birbiri et al. (2020) introduced a custom conditional GAN (cGAN), an extended version of the GAN that imposes a condition in the learning map output, while Dong et al. (2019), Almeida et al. (2022), and Duran et al. (2022) used Attention gates. Almeida et al. (2022) also added an active contour method and a level-set fine-tuning to improve segmentation outcomes. Dong et al. (2019) introduced the DAUNet that uses deep supervision (D) and attention gates (A) with a U-Net base. Dai et al. (2020) proposed the use of a mask region-based convolutional neural network (R-CNN) model that provides a bounding box where the prostate lesion has a higher probability before segmenting the prostate with a 2D and 3D U-Net. In a pioneer work, Hambarde et al. (2020) used radiomic features as a deeply-supervision agent of a 2D U-Net, improving prostate segmentation results. With a direct application in EBRT, Kawula et al. (2022) compared manually segmented Volumetric Modulated Arc Therapy (VMAT) treatment plans with automatically segmented, obtaining interesting results. Table 1 shows information from the datasets and the U-Net based architectures used by authors conveniently ordered by year of publication, modality and data dimensionality.

Year	Authors	Modality	# Patients	Training/Validation/Testing	Data	Architecture
2018	Tian et al.	MRI	50	Public + PROMISE12	3D	U-Net + V-Net + PSNet
		MRI	49	Public + ISBI2013	3D	
		MRI	41	Private + InHouse	3D	
2019	A. Chen et al.	MRI	50	Public + PROMISE12	2D	U-Net
	Elguindi et al.	MRI	100	Private + PROMISE12	2D	U-Net
	Dong et al.	CT	140	Private	2D	DAUNet
	Wang et al.	CT	313	Private	2D	U-Net
	Liu et al.	CT	1114	Private	3D	V-Net
2020	Cem Birbiri et al.	MRI	40	Private + PROMISE12	2D	U-Net + cGAN
	Dai et al.	MRI	42	Private + PROSTATEX-2	2D	R-CNN U-Net
	Hambarde et al.	MRI	50	Private + PROMISE12	2D	U-Net
		MRI	40	Public + NCI-IBSI2013	2D	U-Net
	Khan et al.	MRI	11	Public + UKMMC	2D	U-Net
	Sultana et al.	CT	115	Private	3D	U-Net + GAN
2021	Barra et al.	MRI	100	Private	3D	U-Net
	Meyer et al.	MRI	89	Private + ProstateX	3D	U-Net
2022	Duran et al.	MRI	219	Private	3D	ProstAttention-Net
	Almeida et al.	CT	2226	Private	3D	U ² -Net
	Kawula et al.	CT	69	Private	3D	U-Net

Table 1: Summary results. Proposed architectures, imaging modalities and datasets ordered by year of publication.

4.1. Data Augmentation

When working with private clinical datasets, the number of images is usually limited. Neural networks have millions of parameters, and to achieve good performance, the number of examples required needs to be proportional. However, this also increases complexity. Therefore, if we need more data, there are two solutions: search for more data or generate new ones. Data augmentation is a technique that allows the creation of distinct images or data from the dataset. Figure 4 illustrates some of those techniques.

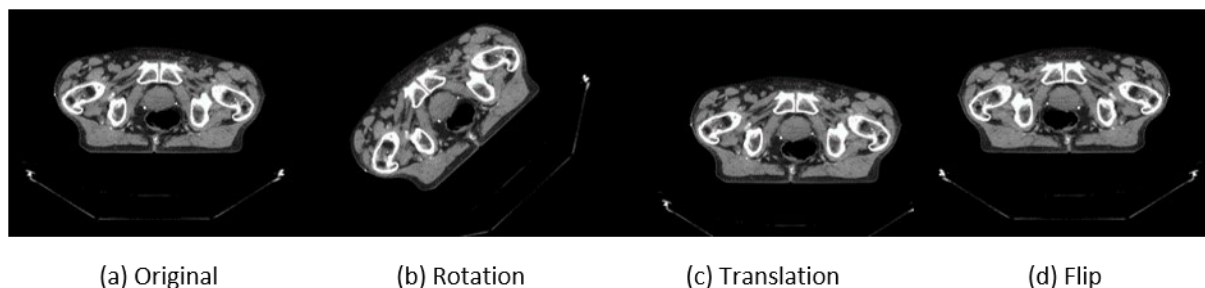


Figure 4: Geometrical Transformations: examples from a private CT dataset.

From the reviewed works, 8 authors used data augmentation techniques. The preferred methods are geometrical transformations. A. Chen et al. (2019) also attempted colour augmentation with the adjustment of brightness, contrast and saturation. Cem Birbiri et al. (2020) added Gaussian noise and grouped pixels sharing similar visual features such as pixel intensity, location or texture, a technique called Super-pixel. Kawula et al. (2022) also attempted B-spline deformations besides 3D rotations around the image centre. Elguindi et al. (2019) spiced the rotations using 5 random permutations per axial image. Liu et al. (2019) applied shifts in the 3 orthogonal directions with amplitudes ranging from 1 to 10 mm. Khan et al. (2020) used random reflections and translation by a fixed amount in the x and y-axis and Tian et al. (2018) used translations and horizontal reflections. Table 2 summarises the techniques used by the authors for data augmentation.

Authors	Techniques
Tian et al. (2018)	Translations and horizontal reflections
A. Chen et al. (2019)	Horizontal and vertical flips; Colour augmentation
Elguindi et al. (2019)	Scaling, cropping, rotations using 5 random permutations per image
Liu et al. (2019)	Shift in the 3 orthogonal directions (1 to 10 mm)
Cem Birbiri et al. (2020)	Super-pixel; Gaussian noise; Moving Mean
Hambarde et al. (2020)	Geometric transformations
Khan et al. (2020)	Random reflection and Translation in the x and y-axis
Kawula et al. (2022)	3D rotations; B-spline deformations; Zooming

Table 2: Data augmentation techniques.

4.2. Preprocessing

For medical imaging, an important aspect is anonymization. Personal information regarding patients and physicians is removed for ethical reasons. Aside from that, the efficiency of a neural network is highly dependent on input data quality. Besides, several methods will increase training speed and minimise losses. When working with images, the preprocessing step can consume a lot of time. Nevertheless, it is crucial to format and standardise data before feeding the network. From the reviewed articles, only Dong et al. (2019) does not mention any preprocessing technique. The most mentioned is the centre crop, usually obtaining a squared batch. A. Chen et al. (2019), for example, did a random crop to 256x256 and then resize the images to 300x300, changing the pixel size. Cem Birbiri et al. (2020) also changed the pixel depth, downsampling from 16 bits to 8 bits. Besides, he also applied a contrast limited adaptive histogram equalisation on the axial image. A common problem with neural networks is sampling. With 3D volumes, voxel size is rarely isotropic. Liu et al. (2019), Wang et al. (2019), Sultana et al. (2020), and Meyer et al. (2021) addressed this issue by resampling to isotropic spacing. Le Cun, Kanter, and Solla (1991) showed that standardisation improves convergence properties by forcing the neurons output to the linear region of the activation functions. With this in mind, Wang et al. (2019), Hambarde et al. (2020), Khan et al. (2020), Almeida et al. (2022), Barra et al. (2021), Meyer et al. (2021), Duran et al. (2022), and Kawula et al. (2022) normalised the images to zero mean and unit variance. Almeida et al. (2022) and Kawula et al. (2022) also adjusted the window width and centre to a specific range. This tuning allows better contrast and visualisation of soft tissues. Table 3 summarises the preprocessing techniques used by the authors in this review.

Authors	Techniques
Tian et al. (2018)	Isotropic sampling
A. Chen et al. (2019)	Resize to 300x300; Random Crop of 256x256
Elguindi et al. (2019)	Downsample 16 bits to 8 bits; False colour images; Contrast Limited Adaptive; Histogram Equalisation on the axial image; Crop to 256x256
Liu et al. (2019)	Resample to 128x128x64 and spatial resolution 1x1x1.5
Wang et al. (2019)	Isotropic sampling; Truncate intensity values; Normalisation
Cem Birbiri et al. (2020)	Resize to 128x128
Dai et al. (2020)	Resize to 384x384
Hambarde et al. (2020)	Standardisation with z score; Non-linear median filter; Resize to 256x256
Khan et al. (2020)	Resize 320x320; Centre Crop 256x256; Intensity normalisation
Sultana et al. (2020)	Downsample to 5x5x5 voxel size and a resolution of 118x118
Barra et al. (2021)	Resize to 256x256; Standardisation with z score
Meyer et al. (2021)	Resampling to a common coordinate system; Crop in the intersection of the three scans; Volume is further cropped to 184x184; Intensity normalisation
Almeida et al. (2022)	Centre Crop: 192x352x192; Window width: -140 to 210 HU; Standardisation with z score; Downscale to half-size in each axis
Duran et al. (2022)	Resample to 1x1x3 voxel size; Centre Crop of 96x96; Intensity normalisation
Kawula et al. (2022)	Pixel intensities truncated to the soft tissue window; Rescale; Centre Crop 128x128x128

Table 3: Preprocessing techniques.

4.3. Training Parameters

In a neural network, certain parameters allow optimising the weights and bias. These parameters are learned and tuned in the training stage. Typically, authors mention the number of epochs, batch size, learning rate monitored by the optimizer and the losses, as summarised in Table 3.4. One can see that the preferred optimizer from the articles in this review is the Adam optimizer, an extension to Stochastic Gradient Descent (SGD) algorithm. Kingma and Ba (2014) mention that Adam is computationally efficient and with low memory requirements, invariant to diagonal rescale of the gradients, highly appropriate for problems with very noisy/or sparse gradients and requires little tuning. Liu et al. (2019), Wang et al. (2019), Dai et al. (2020), Sultana et al. (2020), do not mention the optimizer they used, and A. Chen et al. (2019) presented a custom one. The remaining articles used Adam with initial learning rates ranging from 1×10^{-9} to 1×10^{-2} . The batch size varies from 1 to a maximum of 20, and the number of epochs from 40 to 1300.

A neural network will produce a prediction. Loss functions allow the quantification of the difference between the prediction and the expected outcome, deriving the gradients and updating the weights. The goal is to minimise the loss function. From the reviewed papers, the most used one is the Cross-Entropy function. It computes the difference between two probability distributions using the maximum likelihood estimation. To overcome class imbalance, some authors used a weighted loss function (Tian et al., 2018, A. Chen et al., 2019; Khan et al., 2020; Elguindi et al., 2019; Wang et al., 2019; Hambarde et al., 2020; Sultana et al., 2020; Barra et al., 2021; Duran et al., 2022). Another approach we found was the Dice loss that seeks to balance foreground and background. Dong et al. (2019), Liu et al. (2019), Meyer et al. (2021), and Kawula et al. (2022) used only the Dice loss function but Barra et al. (2021) combined it with the binary focal loss and Almeida et al. (2022) and Duran et al. (2022) with the cross-entropy. Table 4 summarises the training parameters used by authors in this review.

Authors	# Epochs	Batch Size	Optimizer	LR	Loss
Tian et al. (2018)	N.A.	N.A.	SGD	1x10 ⁹	Weighted cross-entropy
A. Chen et al. (2019)	N.A.	N.A.	Custom	N.A.	Adversarial. Weighted cross-entropy
Dong et al. (2019)	200	20	Adam	1x10 ⁻³	
Elguindi et al. (2019)	200	16	Adam	1x10 ⁻⁴	Dice Loss
Liu et al. (2019)	1300	8	N.A.	N.A.	Cross-entropy
Wang et al. (2019)	N.A.	20	N.A.	1x10 ⁻³	Multi-label cross-entropy
Cem Birbiri et al. (2020)	150	1	Adam	1x10 ⁻²	Cross-entropy
Hambarde et al. (2020)	N.A.	32	Adam	2x10 ⁻⁴	Cross-entropy
Khan et al. (2020)	N.A.	16	Adam	1x10 ⁻⁴	Weighted cross-entropy
Barra et al. (2021)	40	4	Adam	1x10 ⁻³	Dice + Binary focal
Meyer et al. (2021)	270	1	Adam	1x10 ⁻³	Dice Loss
Almeida et al. (2022)	100	2	Adam	3x10 ⁻⁴	Dice + Cross-entropy
Duran et al. (2022)	25	N.A.	Adam	1x10 ⁻³	Dice + Cross-entropy
Kawula et al. (2022)	N.A.	N.A.	Adam	1x10 ⁻⁷	Dice Loss

Table 4: Training Parameters. LR: Learning Rate. N.A.: Not Available.

4.4. Evaluation

The DSC is a metric that quantifies the degree of similarity between two regions (ground truth and prediction) and is highly used to measure the performance of an image segmentation method. In this review, we sought articles reporting the DSC for the prostate. Some authors also attempted segmentation of other organs, namely the bladder and rectum. This approach is also interesting since it mimics human holistic contouring. Dong et al. (2019) attempted a multi-organ segmentation framework to apply in an EBRT planning workflow. The obtained

DSC for the prostate was the lowest in this review (0.52), but the results for the bladder and rectum were interesting. Dong et al. (2019) also mentions that the manual ground truths used were not perfect, and one must recall that they did not use any preprocessing technique. Another multi-organ attempt was from Wang et al., 2019 building a model capable of capturing shape with less noise. Sultana et al. (2020) added a GAN and claimed an improvement in segmentation accuracy by allowing the fine segmentation networks to focus only on the region of each organ. Almeida et al. (2022) introduced an active contour method and an attention gate to a traditional U-Net, improving the segmentation outcomes of deep learning. Kawula et al. (2022) also attempted multi-organ, claiming that treatment plans generated with automatic contours proved sufficient for treatment target volume coverage. The highest DSC was from Meyer et al. (2021) (0.94), using multi-planar strategies. In summary, the best results were from 3D approaches, and for multiclass attempts, Sultana et al. (2020) achieved a DSC > 0.90 for the prostate, bladder and rectum. Table 5 summarises the obtained results.

Authors	Prostate	Bladder	Rectum
Tian et al. (2018)	0.85	N.A.	N.A.
A. Chen et al. (2019)	0.88	N.A.	N.A.
Dong et al. (2019)	0.52	0.95	0.84
Elguindi et al. (2019)	0.75	0.90	0.78
Liu et al. (2019)	0.88	N.A.	N.A.
Wang et al. (2019)	0.89	0.94	0.89
Cem Birbiri et al. (2020)	0.83	N.A.	N.A.
Dai et al. (2020)	0.88	N.A.	N.A.
Hambarde et al. (2020)	0.92	N.A.	N.A.
Khan et al. (2020)	0.88	N.A.	N.A.
Sultana et al. (2020)	0.90	0.96	0.91
Barra et al. (2021)	0.90	N.A.	N.A.
Meyer et al. (2021)	0.94	N.A.	N.A.
Almeida et al. (2022)	0.86	0.96	0.84
Duran et al. (2022)	0.88	N.A.	N.A.
Kawula et al. (2022)	0.87	0.97	0.89

Table 5: DSC Values.

5. Final Remarks

Manual segmentation is the preferred method for EBRT planning. Although time-consuming, the available automatic alternatives in EBRT planning systems are far from optimal, accurate or fast. ATLAS based methods require long manual adjustment times. While deep learning models are not yet widely available, research shows promising results. For Prostate cancer, the main segmentation challenges may arise from the wide diversity in patients Hounsfield Unit (HU) from the CT, even for the same organ. A young prostate does not have the same attenuation coefficient as an older prostate. The same for other organs. Besides, the bladder may be filled with urine and the rectum with gas. A previously treated organ, with either surgery or radiotherapy, also has a different attenuation coefficient than an untreated one. There are also patients with hip prostheses that will cause artefacts in the CT image. This diversity may provide challenges to overcome, searching for an accurate segmentation framework.

When performing manual segmentation, medical experts rely on MRI images, to increase soft-tissue contrast. Besides, their expertise and anatomical knowledge allows them to draw a contour based on a 3D mental representation of the volume. For some 2D images, it may seem impossible to segment any structure, at least for an untrained eye. Our visual perception is

sometimes improved by the awareness of other structures. From this review, the best results were from 3D approaches with Meyer et al. (2021) achieving 0.94 with a multi-planar strategy. The addition of GAN to a U-Net also seems to improve segmentation outcomes, even for multi-organ. Sultana et al. (2020) obtained a DSC of 0.90 for the prostate, 0.96 for the bladder and 0.91 for the rectum. Barra et al. (2021) achieved a DSC of 0.90 for the prostate with a more traditional approach. The worst results (not shown) were in the apex and base, where the segmented ground truth is small and presents a challenge for the network. In summary, from the reviewed articles, it seems that the application of state-of-the-art techniques such as GAN, proper preprocessing methods such as isotropic resampling, cropping or resizing, and training with weighted functions to address class imbalance, is a good strategy. Although the U-Net architecture offers structural simplicity and overall good accuracy, results are always highly dependent on the ground truth segmentations and the dataset image quality.

References

- Almeida, Gonçalo, Ana Rita Figueira, Joana Lencart, e João Manuel R.S. Tavares. 2022. «Segmentation of Male Pelvic Organs on Computed Tomography with a Deep Neural Network Fine-Tuned by a Level-Set Method». *Computers in Biology and Medicine* 140 (janeiro): 105107. <https://doi.org/10.1016/j.combiomed.2021.105107>.
- Barra, Davide, Giulia Nicoletti, Arianna Defeudis, Simone Mazzetti, Jovana Panic, Marco Gatti, Riccardo Faletti, Filippo Russo, Daniele Regge, e Valentina Giannini. 2021. «Deep learning model for automatic prostate segmentation on bicentric T2w images with and without endorectal coil». Em *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 3370–73. Mexico: IEEE. <https://doi.org/10.1109/EMBC46164.2021.9630792>.
- Cem Birbiri, Ufuk, Azam Hamidinekoo, Amélie Grall, Paul Malcolm, e Reyer Zwiggelaar. 2020. «Investigating the Performance of Generative Adversarial Networks for Prostate Tissue Detection and Segmentation». *Journal of Imaging* 6 (9): 83. <https://doi.org/10.3390/jimaging6090083>.
- Chen, Ailian, Leilei Zhu, Huaijuan Zang, Zhenglong Ding, e Shu Zhan. 2019. «Computer-Aided Diagnosis and Decision-Making System for Medical Data Analysis: A Case Study on Prostate MR Images». *Journal of Management Science and Engineering* 4 (4): 266–78. <https://doi.org/10.1016/j.jmse.2020.01.002>.
- Chen, Tong, Mengjuan Li, Yuefan Gu, Yueyue Zhang, Shuo Yang, Chaogang Wei, Jiangfen Wu, Xin Li, Wenlu Zhao, e Junkang Shen. 2019. «Prostate Cancer Differentiation and Aggressiveness: Assessment With a Radiomic-Based Model vs. PI-RADS V2». *Journal of Magnetic Resonance Imaging* 49 (3): 875–84. <https://doi.org/10.1002/jmri.26243>.
- Çiçek, Özgün, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, e Olaf Ronneberger. 2016. «3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation». Em *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, editado por Sebastien Ourselin, Leo Joskowicz, Mert R. Sabuncu, Gozde Unal, e William Wells, 9901:424–32. *Lecture Notes in Computer Science*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-46723-8_49.
- Cun, Yann Le, Ido Kanter, e Sara A. Solla. 1991. «Eigenvalues of Covariance Matrices: Application to Neural-Network Learning». *Physical Review Letters* 66 (18): 2396–99. <https://doi.org/10.1103/PhysRevLett.66.2396>.
- Dai, Zhenzhen, Eric Carver, Chang Liu, Joon Lee, Aharon Feldman, Weiwei Zong, Milan Pantelic, Mohamed Elshaikh, e Ning Wen. 2020. «Segmentation of the Prostatic Gland and the

- Intraprostatic Lesions on Multiparametric Magnetic Resonance Imaging Using Mask Region-Based Convolutional Neural Networks». *Advances in Radiation Oncology* 5 (3): 473–81. <https://doi.org/10.1016/j.adro.2020.01.005>.
- Denmeade, Samuel R., e John T. Isaacs. 2002. «A History of Prostate Cancer Treatment». *Nature Reviews Cancer* 2 (5): 389–96. <https://doi.org/10.1038/nrc801>.
- Dong, Xue, Yang Lei, Sibotian, Tonghe Wang, Pretesh Patel, Walter J. Curran, Ashesh B. Jani, Tian Liu, e Xiaofeng Yang. 2019. «Synthetic MRI-Aided Multi-Organ Segmentation on Male Pelvic CT Using Cycle Consistent Deep Attention Network». *Radiotherapy and Oncology* 141 (dezembro): 192–99. <https://doi.org/10.1016/j.radonc.2019.09.028>.
- Duran, Audrey, Gaspard Dussert, Olivier Rouvière, Tristan Jaouen, Pierre-Marc Jodoin, e Carole Lartzien. 2022. «ProstAttention-Net: A Deep Attention Model for Prostate Cancer Segmentation by Aggressiveness in MRI Scans». *Medical Image Analysis* 77 (abril): 102347. <https://doi.org/10.1016/j.media.2021.102347>.
- Elguindi, Sharif, Michael J. Zelefsky, Jue Jiang, Harini Veeraraghavan, Joseph O. Deasy, Margie A. Hunt, e Neelam Tyagi. 2019. «Deep Learning-Based Auto-Segmentation of Targets and Organs-at-Risk for Magnetic Resonance Imaging Only Planning of Prostate Radiotherapy». *Physics and Imaging in Radiation Oncology* 12 (outubro): 80–86. <https://doi.org/10.1016/j.phro.2019.11.006>.
- Fukushima, Kunihiko, e Sei Miyake. 1982. «Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Visual Pattern Recognition». Em *Competition and Cooperation in Neural Nets*, editado por Shun-ichi Amari e Michael A. Arbib, 45:267–85. *Lecture Notes in Biomathematics*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-46466-9_18.
- Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, e Yoshua Bengio. 2020. «Generative adversarial networks». *Communications of The Acm* 63 (11): 139–44. <https://doi.org/10.1145/3422622>.
- Hambarde, Praful, Sanjay Talbar, Abhishek Mahajan, Satishkumar Chavan, Meenakshi Thakur, e Nilesh Sable. 2020. «Prostate Lesion Segmentation in MR Images Using Radiomics Based Deeply Supervised U-Net». *Biocybernetics and Biomedical Engineering* 40 (4): 1421–35. <https://doi.org/10.1016/j.bbe.2020.07.011>.
- Hubel, David, e Torsten Wiesel. 2012. «David Hubel and Torsten Wiesel». *Neuron* 75 (2): 182–84. <https://doi.org/10.1016/j.neuron.2012.07.002>.
- Kawula, Maria, Dinu Purice, Minglun Li, Gerome Vivar, Seyed-Ahmad Ahmadi, Katia Parodi, Claus Belka, Guillaume Landry, e Christopher Kurz. 2021. «Dosimetric Impact of Deep Learning-Based CT Auto-Segmentation on IMRT Treatment Planning for Prostate Cancer». Preprint. In Review. <https://doi.org/10.21203/rs.3.rs-718965/v1>.
- Khan, Zia, Norashikin Yahya, Khaled Alsaih, Syed Saad Azhar Ali, e Fabrice Meriaudeau. 2020. «Evaluation of Deep Neural Networks for Semantic Segmentation of Prostate in T2W MRI». *Sensors* 20 (11): 3183. <https://doi.org/10.3390/s20113183>.
- Kingma, Diederik P., e Jimmy Ba. 2014. «Adam: A Method for Stochastic Optimization». <https://doi.org/10.48550/ARXIV.1412.6980>.
- LeCun, Yann, Bernhard Boser, John Denker, Donnie Henderson, R. Howard, Wayne Hubbard, e Lawrence Jackel. 1989. «Handwritten Digit Recognition with a Back-Propagation Network». Em *Advances in Neural Information Processing Systems*. Vol. 2. Morgan-Kaufmann.

<https://proceedings.neurips.cc/paper/1989/hash/53c3bce66e43be4f209556518c2fcb54-Abstract.html>.

- Litjens, Geert, Robert Toth, Wendy van de Ven, Caroline Hoeks, Sjoerd Kerkstra, Bram van Ginneken, Graham Vincent, et al. 2014. «Evaluation of Prostate Segmentation Algorithms for MRI: The PROMISE12 Challenge». *Medical Image Analysis* 18 (2): 359–73. <https://doi.org/10.1016/j.media.2013.12.002>.
- Liu, Chang, Stephen J. Gardner, Ning Wen, Mohamed A. Elshaikh, Farzan Siddiqui, Benjamin Movsas, e Indrin J. Chetty. 2019. «Automatic Segmentation of the Prostate on CT Images Using Deep Neural Networks (DNN)». *International Journal of Radiation Oncology* Biology* Physics* 104 (4): 924–32. <https://doi.org/10.1016/j.ijrobp.2019.03.017>.
- Meyer, Anneke, Grzegorz Chlebus, Marko Rak, Daniel Schindele, Martin Schostak, Bram van Ginneken, Andrea Schenk, et al. 2021. «Anisotropic 3D Multi-Stream CNN for Accurate Prostate Segmentation from Multi-Planar MRI». *Computer Methods and Programs in Biomedicine* 200 (março): 105821. <https://doi.org/10.1016/j.cmpb.2020.105821>.
- Milletari, Fausto, Nassir Navab, e Seyed-Ahmad Ahmadi. 2016. «V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation». Em 2016 Fourth International Conference on 3D Vision (3DV), 565–71. Stanford, CA, USA: IEEE. <https://doi.org/10.1109/3DV.2016.79>.
- Oktaç, Ozan, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, et al. 2018. «Attention U-Net: Learning Where to Look for the Pancreas». <https://doi.org/10.48550/ARXIV.1804.03999>.
- Parker, C., E. Castro, K. Fizazi, A. Heidenreich, P. Ost, G. Procopio, B. Tombal, e S. Gillissen. 2020. «Prostate Cancer: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up». *Annals of Oncology* 31 (9): 1119–34. <https://doi.org/10.1016/j.annonc.2020.06.011>.
- Qin, Xuebin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R. Zaiane, e Martin Jagersand. 2020. «U2-Net: Going Deeper with Nested U-Structure for Salient Object Detection». *Pattern Recognition* 106 (outubro): 107404. <https://doi.org/10.1016/j.patcog.2020.107404>.
- Ronneberger, Olaf, Philipp Fischer, e Thomas Brox. 2015. «U-Net: Convolutional Networks for Biomedical Image Segmentation». Em *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, editado por Nassir Navab, Joachim Hornegger, William M. Wells, e Alejandro F. Frangi, 9351:234–41. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28.
- Schlemper, Jo, Ozan Oktaç, Michiel Schaap, Mattias Heinrich, Bernhard Kainz, Ben Glocker, e Daniel Rueckert. 2019. «Attention Gated Networks: Learning to Leverage Salient Regions in Medical Images». *Medical Image Analysis* 53 (abril): 197–207. <https://doi.org/10.1016/j.media.2019.01.012>.
- Siddique, Nahian, Sidike Paheding, Colin P. Elkin, e Vijay Devabhaktuni. 2021. «U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications». *IEEE Access* 9: 82031–57. <https://doi.org/10.1109/ACCESS.2021.3086020>.
- Singh, Omesh, e Srinivasa Rao Bolla. 2022. «Anatomy, Abdomen and Pelvis, Prostate». Em *StatPearls. Treasure Island (FL): StatPearls Publishing*. <http://www.ncbi.nlm.nih.gov/books/NBK540987/>.
- Sultana, Sharmin, Adam Robinson, Daniel Y. Song, e Junghoon Lee. 2020. «CNN-based hierarchical coarse-to-fine segmentation of pelvic CT images for prostate cancer

- radiotherapy». Em *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*, editado por Baowei Fei e Cristian A. Linte, 53. Houston, United States: SPIE. <https://doi.org/10.1117/12.2549979>.
- Sung, Hyuna, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, e Freddie Bray. 2021. «Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries». *CA: A Cancer Journal for Clinicians* 71 (3): 209–49. <https://doi.org/10.3322/caac.21660>.
- Tian, Zhiqiang, Lizhi Liu, Zhenfeng Zhang, e Baowei Fei. 2018. «PSNet: prostate segmentation on MRI based on a convolutional neural network». *Journal of Medical Imaging* 5 (02): 1. <https://doi.org/10.1117/1.JMI.5.2.021208>.
- Wang, Yueyue, Liang Zhao, Manning Wang, e Zhijian Song. 2019. «Organ at Risk Segmentation in Head and Neck CT Images Using a Two-Stage Segmentation Framework Based on 3D U-Net». *IEEE Access* 7: 144591–602. <https://doi.org/10.1109/ACCESS.2019.2944958>.
- Zabel, W. Jeffrey, Jessica L. Conway, Adam Gladwish, Julia Skliarenko, Giulio Didiodato, Leah Goorts-Matthews, Adam Michalak, et al. 2021. «Clinical Evaluation of Deep Learning and Atlas-Based Auto-Contouring of Bladder and Rectum for Prostate Radiation Therapy». *Practical Radiation Oncology* 11 (1): e80–89. <https://doi.org/10.1016/j.prro.2020.05.013>.