Faculty of Medicine and Health Sciences

The application of a deep convolution neural network for the automated delineation of the target and organs at risk in High Dose Rate Cervical Brachytherapy.

Thesis submitted to the University of Stellenbosch in partial fulfilment of the requirements for the degree Master of Science in Medical Physics

Submitted by

Didier Raphael Roger Duprez

Supervisor:

Dr Christoph Trauernicht

DECLARATION

“By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third-party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.”

December 2022

Abstract

Low/middle income countries suffer from large deficits in experienced Radiation Oncologists, Medical Physicists and Radiation Therapists. Due to these deficits, the bottlenecks experienced in the High-dose rate (HDR) cervical brachytherapy treatment planning workflow are amplified. Image-guided HDR cervical brachytherapy is a complex, labour intensive, manual, time-consuming and expertise driven process. Automation in radiotherapy treatment planning, especially in brachytherapy, has the potential to substantially reduce the overall planning time however most of these algorithms require high level of expertise to develop. The aim of this study is to implement the out of the box self-configuring deep neural network package, known as No New U-Net (nnU-Net), for the task of automatically delineating the organs at risk (OARs) and high-risk clinical target volume (HR CTV) for HDR cervical brachytherapy.

The computed tomography (CT) scans of 100 previously treated patients were used to train and test three different nnU-Net configurations (2D, 3DFR and 3DCasc). The performance of the models was evaluated by calculating the Sørensen-Dice similarity coefficient, Hausdorff distance (HD), 95th percentile Hausdorff distance, mean surface distance (MSD) and precision score for 20 test patients. The dosimetric accuracy between the manual and predicted contours was assessed by looking at the various dose volume histogram (DVH) parameters and volume differences. Three different radiation oncologists (ROs) scored the predicted bladder, rectum and HR CTV contours generated by the best performing model. The manual contouring, prediction and editing times were recorded.

The mean DSC, HD, HD95, MSD and precision scores for our best performing model (3DFR) were 0.92/7.5 mm/3.0 mm/ 0.8 mm/0.91 for the bladder, 0.84/13.8 mm/5.2 mm/1.4 mm/0.84 for the rectum and 0.81/8.5 mm/6.0 mm/2.2 mm/0.80 for the HR CTV. Mean dose differences ($D_{2cc/90\%}$) and volume differences were 0.08 Gy/1.3 cm^3 for the bladder, 0.02 Gy/0.7 cm^3 for the rectum and 0.33 Gy/1.5 cm^3 for the HR CTV. On average, 65 % of the generated contours were clinically acceptable, 33 % requiring minor edits, 2 % required major edits and no contours were rejected. Average manual contouring time was 14.0 minutes, while the average prediction and editing times were 1.6 and 2.1 minutes respectively. Our best performing model (3DFR) provided fast

accurate auto generated OARs and HR CTV contours with a large clinical acceptance rate. Future work should focus on including larger datasets to eliminate inconsistencies, as well as focus on automating the generation of treatment plans.

Opsomming

Lae/middel-inkomste lande ly aan groot tekorte in ervare stralingsonkoloë, mediese fisici en radioterapeute. Hierdie tekorte lei tot verhoogde werksladings in die beplanning van hoë-dosis-tempo (HDT) servikale bragiterapie behandelings. Beeldgeleide HDT servikale bragiterapie is 'n komplekse-, arbeidsintensiewe-, tydrowende- en kundigheidgedrewe proses. Outomatisering in radioterapie-behandelingsbeplanning, veral in bragiterapie, het die potensiaal om die algehele beplanningstyd aansienlik te verkort, maar die meeste van hierdie algoritmes vereis gevorderde kundigheid om te ontwikkel. Die doel van hierdie studie is die implimentering van die self-konfigurerende diep neurale netwerk pakket, *No New U-Net (nnU-Net)*, om die krities belangrike organe (OARs) en hoë-risiko kliniese teiken-volumes (HR-CTV) outomaties vir HDT servikale bragiterapie te delinieer.

Drie verskillende nnU-Net-konfigurasies (2D, 3DFR en 3DCasc) is retrospektief op die rekenaartomografiese (CT) beelde van 100 behandelde pasiënte opgelei en getoets. Die doeltreffendheid van die modelle is evalueer deur die Sørensen-dobbelsteen-ooreenkomskoëffisiënt (DSC), Hausdorff-afstand (HD), 95ste persentiel Hausdorff-afstand (HD95), gemiddelde oppervlakafstand (MSD) en presisietelling vir 20 toets-pasiënte te bereken. Die verskillende dosis-volume-histogram (DVH) parameters en volume-verskille is gebruik om die dosimetriese akkuraatheid tussen die hand-gedelinieerde en voorspelde kontoere te beoordeel. Die beste presterende model se voorspelde blaas-, rektum- en HR-CTV-kontoere is daarna deur drie verskillende stralingsonkoloë beoordeel. Die totale tydsduur van hand-gedelinieerde kontoere, asook die voorspelde- en gewysigde kontoere is aangeteken.

Die gemiddelde DSC, HD, HD95, MSD en presisietellings vir die beste model (3DFR) is 0.92/7.5 mm/3.0 mm/0.8 mm/0.91 vir die blaas, 0.84/13.8 mm/5.2 mm/1.4 mm/0.84 vir die rektum en 0.81/8.5 mm/6.0 mm/2.2 mm/0.80 vir die HR-CTV. Die gemiddelde dosis- ($D_{2cc/90\%}$) en volume-verskille is 0.08 Gy/1.3 cm^3 vir die blaas, 0.02 Gy/0.7 cm^3 vir die rektum en 0.33 Gy/1.5 cm^3 vir die HR-CTV. Oor die algemeen is 65% van die genereerde kontoere klinies aanvaarbaar, terwyl slegs 33% klein veranderinge benodig het. Verder, het slegs 2% hiervan groot veranderinge benodig, terwyl geen kontoere afgekeur is nie. Die gemiddelde hand-gedelinieerde kontoere het 14.0

minute geneem, terwyl die gemiddelde voorspelde kontoere 1.6 minute geneem het. Dit het slegs 2.1 minute geneem om die voorspelde kontoere te wysig totdat dit klinies aanvaarbaar is. Die beste model (3DFR) het vinnige, akkurate outomaties-gegenereerde kontoere met 'n groot kliniese aanvaardingskoers verskaf. Groter datastelle sal in die toekoms ingesluit word om teenstrydighede uit te skakel en outomatiese behandelingsbeplanning sal ook deel vorm van toekomstige navorsing.

Acknowledgments

I would like to acknowledge the support of my supervisor Dr Christoph Trauernicht, as well as my colleagues in the Medical Physics department at Tygerberg Hospital (TBH). I would also like to thank the clinicians and staff involved in brachytherapy at TBH, as well as those clinicians that gave up their time to evaluate the generated contours.

Table of Contents

List of Figures

List of Tables

1. Introduction and Background

1.1 Introduction

Cancer is an ever-growing burden in Africa with the International Atomic Energy Agency (IAEA) reporting an estimated 1.06 million new cancer cases per year, with this number expected to increase to 2.12 million by 2040.[1] In many sub-Saharan African (SSA) countries, including South Africa, cervical carcinoma is one of the most diagnosed malignancies in women.[2] The human papilloma virus (HPV) is the cause of most cervical cancers with vaccines in high-income countries being well received leading to an already low incidence rate. However, in low/middle income countries, incidence rates are higher and due to lack of funding for healthcare, implementation of large-scale HPV vaccination programmes are hindered.[3] According to the National Centre for Communicable Diseases (NCID) national cancer registry, in 2017, 6 600 new cervical cancer cases were recorded in South Africa alone. This accounted for 15.85 % of all new cancer cases among women[4]. In 2018 the World Health Organisation (WHO) registered 13 004 new cervical cancer cases, accounting for 12.1 % of all cancer cases recorded for men and women combined.[5] These numbers are expected to only increase over the coming years. It is therefore crucial to ensure that patients have access to the latest treatment techniques and for professionals to continuously improve all aspects of the radiotherapy treatment process.

1.2 Thesis structure

This thesis is divided into five main sections. Section one focuses on the introduction, background, and literature with a brief introduction to brachytherapy, calculation of dose distribution around a brachytherapy source, followed by a description of automation techniques and the application deep learning in radiotherapy. Section two outlines the problem statement and study objectives. Section three discusses the research methodology as well as brief explanations of the performance metrics used to evaluate the accuracy of the automatic contouring models. Section four includes the

complete article that has been submitted to the American Association of Physicists in Medicine (AAPM) journal of Medical Physics and finally section five concludes with the summary of findings and future work.

1.3 HDR Brachytherapy

Brachytherapy (sometimes referred to as curie therapy) is a term used to describe radiation therapy in which small, encapsulated radionuclide sources in needles, seeds, wires, or catheters are placed directly into or near a tumour. During intracavitary brachytherapy, the sources are placed inside body cavities in close proximity to the tumour volume using a site-specific applicator. Intracavitary brachytherapy is commonly carried out using either low dose rate brachytherapy (dose rate of 0.4-2 Gy/hr), or a High-dose rate (HDR) brachytherapy afterloader unit (dose rate > 10 Gy/hr). HDR brachytherapy provides several advantages over LDR brachytherapy, such as: reduced treatment time, more flexibility in customizing treatment plans, as well as reduced exposure to physicians and staff.[6]

HDR intracavitary brachytherapy is one of the primary treatment modalities in the fight against cervical cancer and is delivered as either a primary or boost treatment option. Various applicators are used to hold the sources in an appropriate configuration. Cervical applicators commonly consist of a central tube (tandem) and lateral ovoids (Figure 1-3-1a), however some institutes such as Tygerberg hospital use Ring and Tandem (R-T) applicators (Figure 1-3-1b). Although modern techniques such as Intensity Modulated Radiotherapy (IMRT) or Volumetric Modulated Arc Radiotherapy (VMAT) are commonly used as External Beam (EB) treatment options, HDR brachytherapy remains an irreplaceable curative technique owing to its steep dose gradients that offer excellent local control rates while reducing many normal tissue complications.[7–11] Several studies have shown an increase in survival rates and a decrease in recurrence rates when brachytherapy is used as part of a patient's treatment regimen.[12–15]

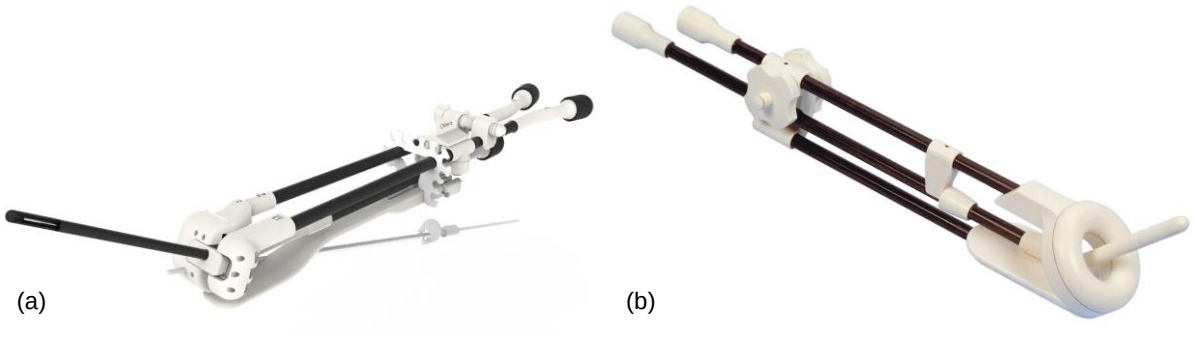


Figure 1-3-1. Elekta (a) tandem-ovoid applicator and (b) tandem-ring applicator.[16]

Over the past two decades, Image-guided (IG) HDR brachytherapy has become the recommended gold standard with several guidelines being published by societies such as Groupe Europeen de Curetherapie (GEC), the European Society for Radiotherapy & Oncology (ESTRO) and the American Brachytherapy Society (ABS).[17–21] Two-dimensional (2D) brachytherapy involves the use of two sets of X-ray images, an anterior-posterior and a lateral image, where treatment doses were commonly prescribed to a point following the Manchester system or ICRU system. However, the lack of spatial/volume information in 2D imaging results in poor localisation of the tumour volume and surrounding organs at risk (OARs). This subsequently leads to poor local control and increased toxicity compared to more modern three-dimensional (3D) IG techniques.[22]

The use of Computed Tomography (CT) imaging or Magnetic Resonance Imaging (MRI) allows health professionals to contour the target and OARs thereby allowing one to plan and deliver highly accurate, patient specific brachytherapy treatments.[23–26] In order to ensure high quality IG-HDR brachytherapy treatments, the accurate delineation of organs at risk (OARs), such as the bladder and rectum, as well as the high-risk clinical target volume (HR-CTV) plays a significant role in the delivery of optimal absorbed dose to the tumour area, while at the same time minimizing the dose to the normal tissue.

Although IG-HDR brachytherapy comes with great advantages, the treatment planning workflow is a labour-intensive, time-consuming process that includes a number of manual steps and involves input from a range of professionals; Radiation Oncologists

(ROs), Radiotherapists (RTs) and Medical Physicists (MPs). Several studies have also highlighted issues due to inter- and intra- observer variability in organ delineation with Hellebust et al and Saarnak et al reporting inter-observer variabilities between 5-8 % and 10-11 % respectively.[27–31] Not only is inter- and intra- observer variability a growing concern, Low- and middle-income countries, such as those in Africa, face several other challenges in an attempt to meet the demands for high quality cancer treatment. These include and are not limited to equipment, maintenance, high workloads as well as huge deficits in experienced ROs, MPs and RTs.[32,33] As a result, there has been a growing demand internationally to try and automate as much of the radiotherapy treatment planning workflow as possible using a range of computer aided algorithms.

1.4 Dose Distribution (AAPM TG-43)

For one to accurately determine the dose to the target and OARs in HDR brachytherapy, it is vital to know what the dose distribution around the source will be. The updated AAPM Task Group 43 (TG-43) report provides the basis for determining the dose distributions around brachytherapy sources by assuming photon interactions only and treating the surrounding medium as water. [34] The dose rate at any point from a single finite source can be calculated from the 2D dose rate equation (equation 1) given below.

$$\dot{D}(r,\theta) = S_K \Lambda \frac{G(r,\theta)}{G(r_0,\theta_0)} g(r) F(r,\theta) \qquad (Eq.\,1.)$$

Where:

r is the distance (cm) from the origin to the point of interest P

θ is the angle between the radius vector r and the long axis of the source

θ_o defines the source transverse plane and is equal to $\pi/2$ radians

S_K is the air kerma strength of the source ($\mu Gy \cdot m^2 \cdot h^{-1}$)

Λ is the dose rate constant in water

G(r,θ) is the geometry function

g(r) is the radial dose function

F(r,θ) is the anisotropy function

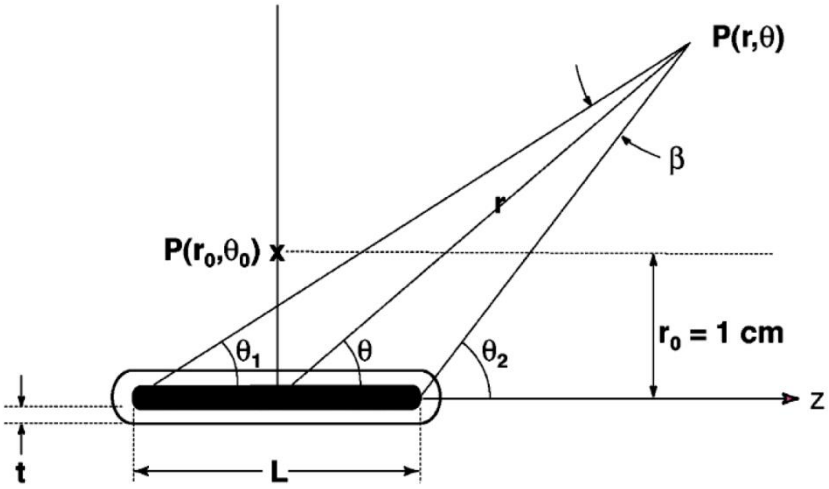


Figure 1-4-1. Coordinate system for 2D brachytherapy dose-rate calculations. [34]

1.4.1 Dose Rate Constant

The dose rate constant L is defined as the dose rate to water at a distance r_0 on the transverse axis, P(r_0 = 1 cm,θ_0 = 90º), per unit air kerma strength source in water (cGy·U^{-1}·h^{-1}):

$$\Lambda = \frac{D(r_0, \theta_0)}{S_K} \qquad (Eq. 2.)$$

The dose rate constant depends on the radionuclide, source model and the internal source design. (1 U = 1 cGy·cm^2·h^{-1}).

1.4.2 Geometry Factor

The geometry function, neglects scattering and attenuation but accounts for the variation of the relative dose due to the spatial distribution of activity within the source by applying an inverse square-law correction based on an approximation of the activity distribution. G(r, θ) reduces to $1/r^2$ for point source approximation and to β/(Lr sinθ) for a line source approximation with β and L shown in figure 1-4-1.

1.4.3 Radial Dose Function

The radial dose function g(r) accounts for the effects of attenuation and scatter in water on the transverse plane of the source (θ = π/2), excluding falloff, which is included by the geometry function G(r, θ).

1.4.4 Anisotropy Function

The anisotropy function F(r, θ) accounts for the directional dependence of the dose rate distribution around the source, including the effects of absorption and scatter within the source core and encapsulation material. F(r, θ) is defined as unity on the transfer plane, however, its value off the transfer plane decreases:

(i) as r decreases;

(ii) as θ approaches 0º or 180º;

(iii) as the source encapsulation thickness increases; and

(iv) as the photon energy decreases.

1.5 Automation in Radiotherapy

Radiotherapy is a rapidly evolving field with new innovations being developed at an ever-increasing rate. A large number of these innovations are being focused on automating various steps or processes in radiation oncology. Automation has a wide range of uses in radiotherapy with many studies looking at automating clinical workflows, quality assurance, chart reviews, treatment planning as well as organ at risk and target delineation.[35–37] Over the past few years, several techniques have been proposed to automate the delineation of OARs and targets in radiotherapy.

1.5.1 Atlas Based Methods

A commonly used approach is Atlas Based (AB) methods which involves deformable image registration between the new image and the atlas or reference image. Once registration is complete, segmentations from the reference image or atlas are transformed onto the new or test image as shown in figure 1-5-1.[38–41] As a result, AB methods depend largely on the type of algorithm applied and therefore tend to suffer from a lack of certainty in deformable registration. Moreover, AB methods are time consuming and are plagued by sub-optimal segmentations especially when dealing with atypical patients, internal organ motion, bladder filling or the presence of gas in anatomical sites such as the rectum.[42–45] Due to the inconsistencies in AB segmentation methods, there has been a recent shift towards deep learning (DL) automated segmentation techniques with promising results.

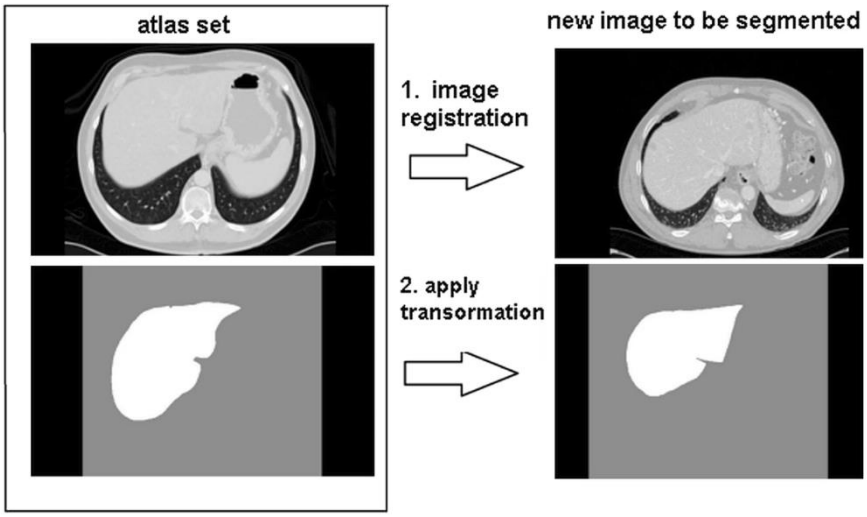


Figure 1-5-1. Example of Atas based segmentation of the liver.[46]

1.5.2 Deep Learning Based Methods

Machine learning (ML) involves the use of algorithms and statistical models to analyse and draw inferences from patterns in data with the ability to learn and adapt to imitate an intelligent human process. ML algorithms can be trained to do a large variety of tasks via either supervised, unsupervised or semi-supervised learning. Supervised learning involves data, where each training input example has a unique corresponding label or output ground truth, as in the case of the cat/dog binary classification task seen in figure 1-5-2. Unsupervised learning on the other hand, attempts to find the desired output from a given unlabelled dataset by identifying patterns in the data through repetitive processes such as clustering or grouping.

1.5.2.1 Linear and Logistic Regression

The simplest form of machine learning is linear regression, where one would like to determine the line of best fit (y =mx +c) for a given set of x values. This is achieved by incrementally adjusting the weights, in this case the gradient (m) and y-intercept (c), until the sum of the squared differences between the actual value and linear fit value (called the cost function), for all given values of x, is minimized.

The same concept can be applied to an image classification task, where each pixel in an image is known as a feature (x_i) and can be used to train a model to classify whether a specific image falls into one of the given classes. For example, the binary classification of whether an image is a dog or cat (figure 1-3.), could have a single layer trial function such as that given in equation 3.

$$y_p = ax_1 + bx_2 + cx_3 + dx_4 + ex_5 + fx_6 + gx_7 + hx_8 + jx_9 + k \qquad (Eq.3.)$$

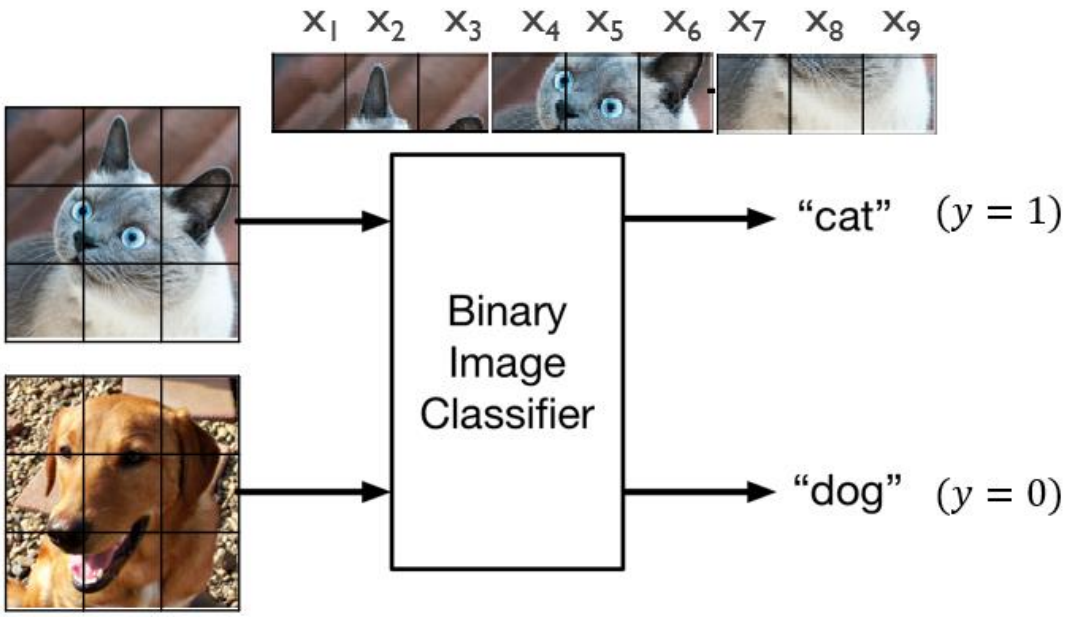


Figure 1-5-2. Example of binary image classifier.

One would then try to find the best weights (a, b, c, d, e, f, g, h, j and k) that can minimize the difference between the predicted value, y_p, and the ground truth value, y. However, for a binary classification task such as this, one would use logistic regression, where the goal is to minimise the difference in the logarithm of the predicted y_p and ground truth y (equation 4) by repetitively adjusting the weights, with y_p taking on a probability between 0 and 1, where a value closer to 1 would indicate a cat and a value close to 0 indicating the image is a dog.

$$Cost\ Function = \frac{1}{m}\sum_{i=l}^{m} -\left[log\left(y_p(x_i)\right) + (y_i)log\left(1 - y_p(x_i)\right)\right] \quad (Eq.\ 4.)$$

1.5.2.2 Deep Learning

DL is a subset of machine learning where models known as neural networks, usually comprised of at least three or more layers, attempt to simulate the behaviour of a human brain. An example of a 5-layer fully-connected neural network is shown in figure1-5-3. The first layer of a neural network is the input layer, followed by a certain number of hidden layers (in this case three) where increasing the number of hidden layers essentially increases the depth of the neural network hence the term deep learning networks. Each hidden layer is comprised of several neurons, where each neuron takes in an input, applies a trial function with weights and returns an output, as

shown in figure 1-5-4. Each trial function also includes a bias term, b, which is analogous to the role of the constant or y-intercept in a linear fit function.

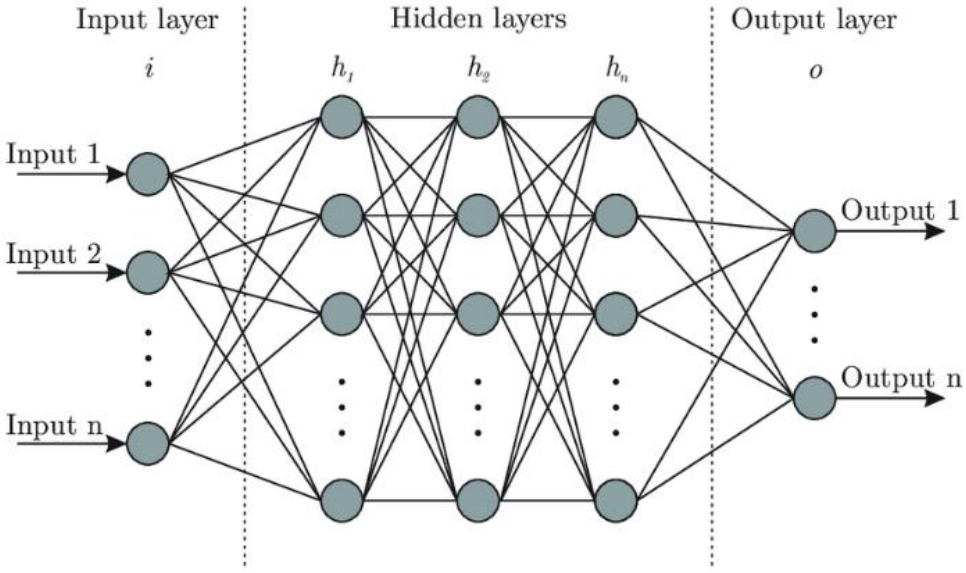


Figure 1-5-3. Example of a 5-layer neural network [47]

The final layer is the output layer, which in the case of the binary classification task mentioned earlier, these output nodes would be the two classes (cat or dog).

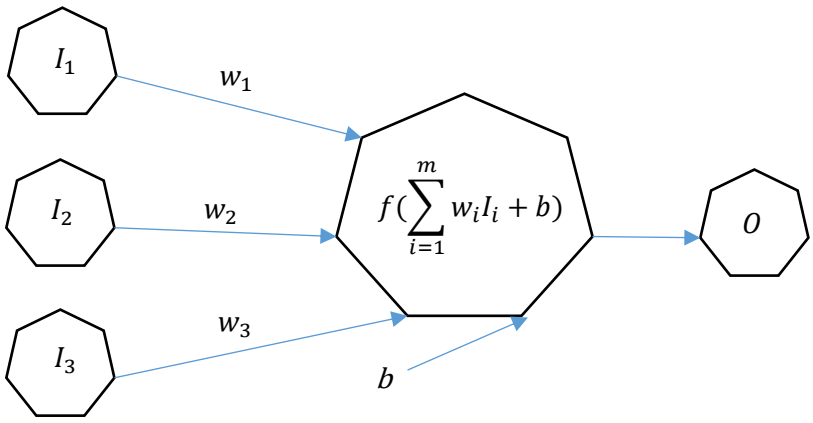


Figure 1-5-4. Example of a single neuron.

1.5.2.3 Training Neural Networks

The aim when training a neural network, is to learn the values that the weights and biases need to take on to minimise the difference between predicted output and ground truth, known as the loss. The training process is iterative and is commonly comprised off four steps: forward propagation, loss calculation, backpropagation and parameter update. Model parameters (weights and bias) are first randomly initialised and then the training input data (analogous to x input of straight line) is passed through the network to obtain a prediction, the forward propagation step.

The predicted output is then compared to the ground truth or label and the loss is calculated. There are a wide variety of loss functions that one can use (binary, cross-entropy, mean squared error, etc) and can be customised to the task at hand. The loss information is then backpropagated from the output layer to all the neurons in the hidden layer, with neurons receiving a fraction of the total loss based on its relative contribution to the original output.

Once the loss has been backpropagated to all the neurons, all the weights are adjusted via a process known as gradient descent. This process uses the derivative or gradient of the loss function to change the weights by small increments in a direction that aims to reduce the total loss function and in doing so improve accuracy of the model. During model training, these steps are continuously repeated until the total loss function has been minimised, i.e. a global minimum for the loss function has been found.

1.5.2.4 Deep Convolution Neural Networks

Deep convolution neural networks (CNN) are the gold standard for tasks such as image recognition, image classification, image segmentation and facial recognition.[48] Two dimensional Images are comprised of arrays with dimensions Height x Width x c, where height and width are the number of pixels or image resolution of the image, while c is known as channel number which takes on the value 3 for colour images (Red, Green, Blue colour channels) and 1 for grayscale images. CNNs contain

convolution layers, whereby filters are applied and move across the image in order to extract or enhance certain features (figure 1-4-5).

These filters or kernels are trainable and replace the concept of weights as seen in the case of neural networks. Stacking several convolution layers creates a deep CNN where one can provide training data comprising of input images and corresponding ground truth outputs for the model to incrementally change and learn the filters needed to extract the relevant information or to classify each pixel in the input image into a given class. This forms the basis of most deep learning CNN based segmentation models.

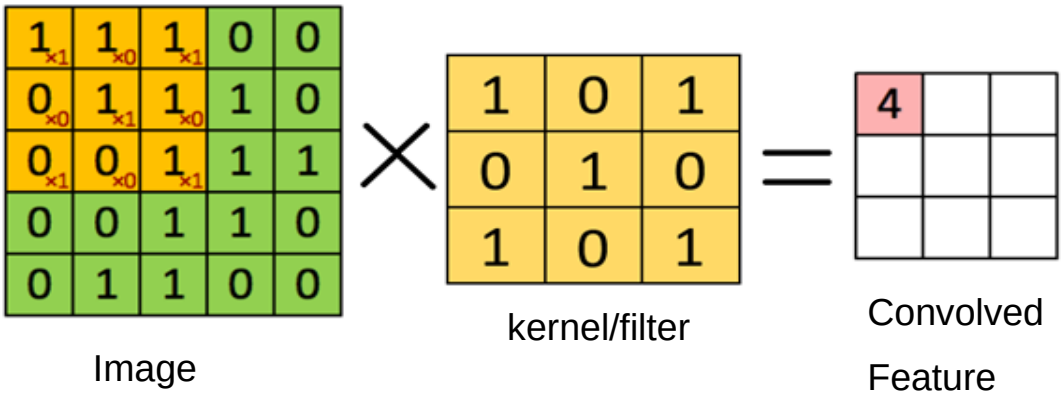


Figure 1-5-5. Convolution of a 5x5x1 image with a 3x3x1 filter [49]

1.5.3 Deep Learning in Radiotherapy

The automated delineation of the organs at risk (OARs) and targets using deep convolutional neural networks in EB radiotherapy have been successfully applied to cases such as head and neck[50], brain metastases[51], breast[52], rectum[53], pelvis[54,55] and cervix.[56,57] DL methods have shown to provide an increase in accuracy, reproducibility, and robustness over conventional AB methods. [52,58–61] DL has also been effectively applied in various areas of brachytherapy from applicator reconstruction, dose calculation, treatment planning as well as organ delineation[62]. In HDR cervical brachytherapy work has been done on automating the reconstruction of applicators using deep learning[63,64], while a few studies have looked at the automatic segmentation of the OARs and targets.[64–66]

The study conducted by Zhang et al, focused on OAR and target delineation as well as applicator reconstruction using a 3D convolution neural network based on the novel 3D-Unet architecture which incorporated residual connections and deep supervision.[64–66] Mohammadi et al, applied a 2D model architecture combining a Residual Network (ResNet) and popular U-Net to create a deeper convolutional neural network for segmenting the OARs but failed to look at the delineation of the high-risk clinical target volume (HR CTV). [64–66] Another study carried out by Yoganathan et al, focused on segmenting the OARs and HR CTV using two different 2D and 2.5D models, with one applying a ResNet50 backbone and the second implemented an Inception Residual network backbone. [64–66]

These studies however did not look at the clinical acceptability of the predicted contours, the effect that an automated system might have on the brachytherapy workflow and involved complex deep learning models that require a high level of programming, as well as machine learning, knowledge to replicate in one's own clinical department.[67] Although there are a number of commercial software tools available to tackle automated segmentation in radiotherapy, most of these systems are extremely expensive and are not task specific, with application mostly centred around general EB radiotherapy treatment sites. [68] The purpose of this study is to train and implement the out of the box deep convolution neural network framework known as No New U-Net (nnU-Net), for the automated delineation of the HR CTV and OARs in HDR cervical brachytherapy. This study will also assess the clinical acceptability of the generated contours as well as the effect that the automated segmentation system will have on the efficiency/timesaving of the brachytherapy workflow in the Department of Radiation Oncology at Tygerberg Hospital.

1.6 Tygerberg Hospital HDR Cervical Brachytherapy Workflow

In the Department of Radiation Oncology at Tygerberg Hospital, cervical brachytherapy patients require a smit tube/sleeve (S-Tube) to be sutured into the cervix. Once the S-tube's position has been verified, the brachytherapy Ring and Tandem (R-T) applicator is then carefully placed into the patient via the S-Tube. Once the applicator is in place, the patient undergoes a CT planning scan. This CT scan is then exported to the Oncentra (Elekta, Stockholm, Sweden, version 4.6) planning system where an RO contours out a HR-CTV along with the bladder and rectum. Once the patient has been imaged, the patient with applicator in place is carefully moved into the brachytherapy suite and the first fraction of the patient's treatment is delivered. Due to the time required for an RO to contour all the necessary organs, every patient's first fraction is treated using a standard plan template. Once target and OAR contouring has been completed, an MP creates a patient specific adapted plan which is then used for treating the patient's remaining fractions.

With the possible implementation of an automated contouring system, the aim would be to substantially improve the efficiency of the brachytherapy workflow. The first few steps are unchanged, however once the applicators have been placed and the planning CT sent through, the automated contouring system would theoretically; pull in the CT, automatically delineate the required structures, and then send the CT with structure set to the planning system. This would allow the MP to create a patient specific plan while the patient is being moved across to the brachytherapy suite. The RO will then need to only verify that the contours are clinically acceptable, as well as approve the treatment plan. The patient will then be able to start their treatment from the first fraction on a plan that has been adapted specifically for them.

2. Problem Statement and Objectives

2.1 Problem Statement

The contouring stage in the IG-HDR cervical brachytherapy treatment planning workflow is a time consuming and labour-intensive process resulting in bottlenecks or sub-optimal treatments, that are amplified in resource constraint environments such as those in low/middle income countries.

2.2 Central Research Objectives

The central research object is to train various nnU-Net configurations for the automatic segmentation of the OARs and HR CTV in IG-HDR cervical brachytherapy. The first stage of the study is to collect training data, anonymize the data, pre-process the data and configure the nnU-Net pipeline.

The second stage objective is aimed at training and testing the three nnU-Net configurations by evaluating each of the different configuration's performance against the ground truth manual contours using the metrics described in section 3.1.

The final objective will be to evaluate the clinical acceptability of the generated contours, as well as the possible time saving when implementing the automatic contouring system into the planning workflow.

3. Research Methodology

Place of research: Tygerberg Hospital, Radiation Oncology Division

Study design: Retrospective study

Selection of participants: 100 locally advanced cervical cancer patients treated with external beam radiotherapy (50 Gy/25 fractions or 46 Gy/23 fractions), followed by HDR brachytherapy (25 Gy/5 fractions or 24 Gy/4 fractions) using the ring and tandem applicators. All patients were scanned on a Phillips Brilliance Big Bore 16 Slice CT scanner. All bladder, rectum and HR CTVs were manually contoured and approved by experienced clinical oncologists.

Exclusion criteria: Patients who were treated with a tandem applicator alone (no ring) or who were treated only with external beam radiotherapy.

Ethics: Ethics approval was obtained from the University of Stellenbosch Health Research Ethics Committee (HREC) with project ID 24418 and ethics reference number S22/01/013.

Data Analysis: The anonymised patient data was used to train three different DL configurations (models) for the task of automatically contouring the OARs and HR CTV. The performance of the different configurations was assessed using the metrics defined in section 3.1. The clinical acceptability of the contours was judged by three independent ROs using the scoring mechanism shown in table 3-1, and the dosimetric impact evaluated based on the cumulative dose volume histogram (DVH) statistics between the manual and predicted contours.

Table 3-1. Scoring mechanism for clinician assessments.

Rank	Evaluation
1	Reject completely
2	Major revisions
3	Minor revisions before clinically acceptable
4	Clinically acceptable

3.1 Performance Metrics

3.1.1 Dice Similarity Coefficient

Dice Similarity Coefficient (DSC) or Sørensen-Dice Coefficient (equation 3) is a commonly used metric in medical image segmentation tasks by comparing the output from a ML model against a reference/ground truth mask. The DSC can also be used to compare two manual volumes drawn by two different users.[69] The DSC is a metric that indicates the spatial overlap between two regions or in this case medical contours. DSC values range between 0 and 1, where 1 indicates complete overlap between the two contours.

$$DSC = 2\frac{(|A \cap B|)}{(|A| + |B|)} \qquad (Eq.5.)$$

3.1.2 Hausdorff Distance and 95th percentile Hausdorff Distance

The Hausdorff distance (HD) or Pompeiu-Hausdorff distance is a measure of how far the surface of two subsets of a metric space are from each other, i.e. how far two contour sets are from each other. The Hausdorff distance is the maximum closest distance between a set of points of two contours and is defined as:

$$HD(A,\ B) = max\{h(A,B),\ h(B,A)\} \qquad (Eq.6.)$$

Where h(A, B) is given by:

$$h(A,B) = \max_{a \in A}\left\{\min_{b \in B}\{d(a,b\}\right\} \qquad (Eq.7.)$$

The Hausdorff distance as it stands can be an adequate indication of a model's performance, however it is prone to outliers which may not be a true indication of how well the model is performing. It has therefore become popular to report the 95th percentile Hausdorff distance (HD95), as it eliminates possible outliers and provides a better indication of a model's performance.

3.1.3 Mean Surface Distance

The Mean Surface Distance (MSD) is calculated using the surface pixels from two contours A and B. For each surface pixel in A, the mean Euclidean distance to the surface voxel of B is calculated. The MSD tells us on average, how much the surface of the segmentation varies from the ground truth. The MSD is defined as in equation 8 below, where the mean distance between every surface voxel in A and closest voxel in B is calculated. The calculation is repeated for B to A and the average taken.

$$MSD(A,\ B) = \frac{\{mean(d(A,B)) + mean(d(B,\ A))\}}{2} \qquad (Eq.8.)$$

3.1.4 Precision Score

The precision score is an indication of the fraction of predictions that are true positives (TP) rather than false positives (FP). In other words, the precision score indicates the ratio of the number of pixels located on the boundary of a predicted contour, which are close enough to the ground truth contour, to the total predicted contour boundary:

$$Precision = \frac{TP}{TP + FP} \qquad (Eq.9.)$$

4. **Submitted Article:** Self-configuring nnU-Net for automatic delineation in high dose-rate cervical brachytherapy, a low/middle income country's experience.

Manuscript submission number: 22-1580

Short running title: Self-configuring nnU-net in HDR Brachy

Didier Duprez

Division of Medical Physics, Stellenbosch University, Tygerberg Academic Hospital, Cape Town, South Africa

Christoph Trauernicht

Division of Medical Physics, Stellenbosch University, Tygerberg Academic Hospital, Cape Town, South Africa

Hannah Simonds

Department of Oncology, University Hospitals Plymouth NHS trust, Plymouth, United Kingdom

O'Brian Williams

Division of Radiation Oncology, Stellenbosch University, Tygerberg Academic Hospital, Cape Town, South Africa

Corresponding author:

Didier Duprez,

Medical Physicist,

Division of Medical Phyiscs,

Stellenbosch University, Tygerberg Academic Hospital

Cape Town, South Africa

Phone: +27 219386070

E-mail: dduprez@sun.ac.za

ABSTRACT

Background: The High-dose rate (HDR) brachytherapy treatment planning workflow for cervical cancer is a labour intensive, time consuming and expertise driven process. These issues are amplified in low/middle income countries with large deficits in experienced healthcare professionals. Automation has the ability to substantially reduce bottlenecks in the planning process but often require a high level of expertise to develop.

Purpose: To implement the out of the box self-configuring nnU-Net package for the auto-segmentation of the organs at risk (OARs) and high-risk CTV (HR CTV) for Ring-Tandem (R-T) HDR cervical brachytherapy treatment planning.

Methods: The computed tomography (CT) scans of 100 previously treated patients were used to train and test three different nnU-Net configurations (2D, 3DFR and 3DCasc). The performance of the models was evaluated by calculating the Sørensen-Dice similarity coefficient, Hausdorff distance (HD), 95th percentile Hausdorff distance, mean surface distance (MSD) and precision score for 20 test patients. The dosimetric accuracy between the manual and predicted contours was assessed by looking at the various dose volume histogram (DVH) parameters and volume differences. Three different radiation oncologists (ROs) scored the predicted bladder, rectum and HR CTV contours generated by the best performing model. The manual contouring, prediction and editing times were recorded.

Results: The mean DSC, HD, HD95, MSD and precision scores for our best performing model (3DFR) were 0.92/7.5 mm/3.0 mm/ 0.8 mm/0.91 for the bladder, 0.84/13.8 mm/5.2 mm/1.4 mm/0.84 for the rectum and 0.81/8.5 mm/6.0 mm/2.2 mm/0.80 for the HR CTV. Mean dose differences ($D_{2cc/90\%}$) and volume differences were 0.08 Gy/1.3 cm^3 for the bladder, 0.02 Gy/0.7 cm^3 for the rectum and 0.33 Gy/1.5 cm^3 for the HR CTV. On average, 65 % of the generated contours were clinically acceptable, 33 % requiring minor edits, 2 % required major edits and no contours were rejected. Average manual contouring time was 14.0 minutes, while the average prediction and editing times were 1.6 and 2.1 minutes respectively.

Conclusion: Our best performing model (3DFR) provided fast accurate auto generated OARs and HR CTV contours with a large clinical acceptance rate.

1. INTRODUCTION

Cervical cancer is an ever-growing burden in Africa with the 2017 National Centre for Communicable Diseases (NCID) registry recording approximately 6 600 new cervical cancer cases in South Africa alone. This accounted for 15.85 % of all new cancer cases among women.[1] In 2018 the World Health Organisation (WHO) registered 13 004 new cervical cancer cases, accounting for 12.1 % of all cancer cases recorded for men and women combined.[2] These numbers are expected to only increase over the coming years. It is therefore crucial to ensure that patients gain access to the latest treatment techniques and for health professionals to continuously strive to improve all aspects of the radiotherapy treatment process.

Image guided High Dose Rate (IG-HDR) intracavitary brachytherapy is an irreplaceable curative treatment modality for locally advanced cervical cancer. Often provided as a primary or boost treatment option, HDR brachytherapy provides steep dose gradients that offer excellent local control rates while reducing a number of normal tissue complications.[3–7] Several studies have shown an increase in survival rates and a decrease in recurrence rates when brachytherapy is used as part of a patient's treatment regimen.[8–11] Although IG-HDR brachytherapy comes with great advantages, the treatment planning workflow as a whole is a labour intensive process that includes a number of manual, time consuming steps or processes and involves input from a range of professionals; Radiation Oncologists (ROs), Radiotherapists (RTs) and Medical Physicists (MPs). While MRI is the gold standard, CT-based IG brachytherapy is more cost-effective in a resource-constrained environment with limited, or no access to MRI.

Organ at risk (OAR) and target delineation is an extremely important aspect of IG brachytherapy however this step makes up a significant portion of the whole treatment planning time and depends largely on the expertise of the RO on duty. Several studies have highlighted issues due to inter- and intra- observer variability in organ delineation with Hellebust et al and Saarnak et al reporting inter-observer variabilities between 5-8 % and 10-11 % respectively.[12–16] Not only is inter- and intra- observer variability a growing concern, low- and middle-income countries, such as those in Africa, face a variety of challenges in an attempt to meet the demands for high quality cancer treatment. These include and are not limited to equipment, maintenance, high

workloads as well as huge deficits in experienced ROs, MPs and RTs. [17,18] A recent investigation into the current state of cancer in sub-Saharan Africa (SSA), published in the Lancet Oncology journal, found that the shortages of radiotherapy professionals is one of the most crucial barriers hindering access to cancer services in SSA, with an estimated 211 % increase in workforce required to provide equitable access to radiotherapy.[19] Computer aided automation has the ability to mitigate the bottlenecks experienced throughout the radiotherapy process and to alleviate some of the workload on the ROs, MPS and RTs. As a result, there has been a growing demand internationally to automate as much of the radiotherapy treatment planning workflow as possible.

Over the past few years, several techniques have been proposed to automate the delineation of OARs and targets in radiotherapy. A commonly used approach is Atlas Based (AB) methods which involve deformable image registration, where segmentations from the reference image or atlas are transformed onto the new or test image. [20–23] AB methods suffer from lack of certainty in deformable registration and provide sub-optimal segmentations when dealing with atypical patients, organ motion and varying levels of bladder and rectum filling.[24–26] Due to the inconsistencies in AB methods, there has been a worldwide shift towards deep learning (DL) automated segmentation techniques, providing an increase in accuracy, reproducibility, and robustness.[27–30]

The automated delineation of OARs and targets using DL networks in external beam (EB) radiotherapy have been successfully applied to cases such as head and neck[31], brain metastases[32], breast[28], rectum[33], pelvis[34,35] and cervix[36,37]. DL has also been effectively applied in various areas of brachytherapy from applicator reconstruction, dose calculation, treatment planning as well as organ delineation.[38] In IG-HDR cervical brachytherapy, work has been done on automating the reconstruction of applicators using DL[39,40], with a few studies looking at the automatic segmentation of the OARs and targets.[40–43]

These studies however do not look at the clinical acceptability of the generated contours and involve complex deep learning models that require a high level of expertise to reproduce or apply in one's own clinical department. [44] The purpose of this study is to train and implement the self-configuring No New U-Net (nnU-Net),

developed by Isensee et al, for the task of automatically delineating the OARs and HR-CTV in IG-HDR cervical brachytherapy.[45,46] To the best of our knowledge, this is the first CT based study applying the novel nnU-Net to cervical brachytherapy segmentation with ring and tandem applicators. This addition is important as low/middle income departments experience severe bottlenecks in their planning process and may not have the required expertise to design and implement their own machine learning models, a task that nnU-Net was designed to tackle.

2. MATERIALS AND METHODS

2.A. Clinical Dataset

CT images from 100 locally advanced cervical cancer patients were included in the study. All patients were initially treated with external beam radiotherapy (50 Gy/25 fractions or 46 Gy/23 fractions), followed by HDR brachytherapy (25 Gy/5 fractions or 24 Gy/4 fractions) using the ring and tandem applicators. All patients were scanned on a Phillips Brilliance Big Bore 16 Slice CT scanner. Images were reconstructed with a 512 x 512 matrix size and 2- or 3-mm slice thickness. All OARs and HR-CTVs were subsequently contoured and approved by a clinical oncologist, based on the IBS-GEC ESTRO-ABS recommendations.[47]

To maintain patient anonymity, all patient data was completely anonymised. The pixel data from each patient's CT was extracted from the DICOM files and converted into a three-dimensional (3D) NIFTI file format. The contour information was also extracted from the DICOM structure file and used to create 3D multi-class segmentation masks of the OARs and HR-CTVs and saved in NIFTI file format. The 100 patients included in the study, were then split 80 patients for training/validation and 20 patients for testing. The 80 patients were then further split into 64 for training and 16 for validation.

2.B. nnU-Net

Deep learning-based segmentation models tend to be task specific, with any slight adjustment in architecture, training parameters or data, leading to significant drops in performance. Therefore, one often requires a high level of expertise to design a robust network architecture, optimize the data augmentation, determine the appropriate pre-processing or post-processing, and select the ideal training parameters to obtain a

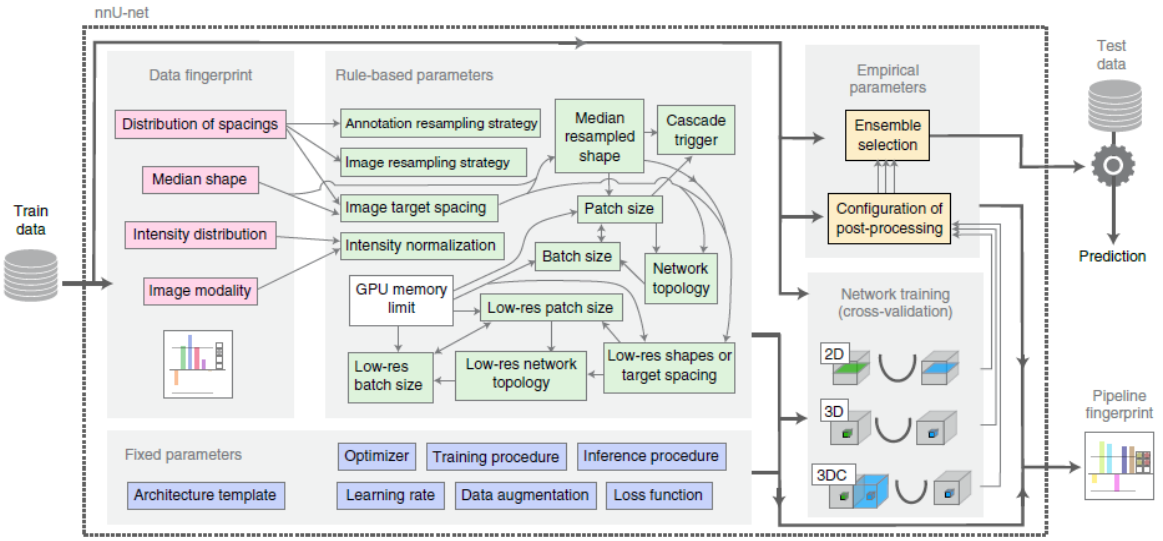
model best suited for the segmentation task at hand.[44] This process is especially cumbersome when dealing with 3D medical images, where image size, imaging modality, voxel size and class imbalance vary substantially resulting in poor transfer of a configuration from one dataset to another.

As a solution, Isensee et al, developed a self-configuring method for deep learning-based image segmentation specific to bio-medical imaging. The result is a completely automated deep learning segmentation pipeline, allowing any user with minimal programming knowledge the ability to set up and train their very own segmentation models.[45,46] At its heart, nnU-Net is based on the U-Net architecture, developed by the computer science department of the University of Freiburg for biomedical image segmentation.[48] Isensee et al, have found that a well-trained U-Net has the potential to beat or at the very least match any task specific segmentation model, leading them to develop their segmentation pipeline known as No New U-Net or nnU-Net. This self-configuring segmentation pipeline was proven to outperform extremely task specific model architectures in several international biomedical auto-segmentation competitions on 23 different publicly available datasets.[46]

Once a dataset has been provided, in the correct format, nnU-Net steps in and carries out several configuration processes based on the dataset fingerprint. These include but are not limited to the pre-processing (sampling, batch size, patch size, etc), network topology, training, and post-processing. A summary depicting how the nnU-Net segmentation pipeline operates as well as certain design and configuration choices, such as learning rate scheduler and loss function is shown in Fig. 1. nnU-Net provides three different trainable U-Net configurations: a two-dimensional (2D) U-Net, 3D full resolution U-Net (3DFR) and a 3D cascade U-Net (3DCasc) where the first part is trained on low resolution images followed by refining the segmentation maps at full resolution.

nnU-Net allows the user to train all configurations using five-fold cross-validation, and once completed, nnU-Net can empirically determine the best performing configuration or ensemble of configurations for the specific task. A main concern with running ensembles is the significantly high prediction times that can lead to impractical clinical application. For this study, all configurations were trained using five-fold cross-validation, where the best single fold configuration was determined independently

using the performance metrics defined in section 2.C, as well as visual inspection of the predicted contours.



Design choice	Required input	Automated (fixed, rule-based or empirical) configuration derived by distilling expert knowledge (more details in online methods)
Learning rate	–	Poly learning rate schedule (initial, 0.01)
Loss function	–	Dice and cross-entropy
Architecture template	–	Encoder–decoder with skip-connection ('U-Net-like') and instance normalization, leaky ReLU, deep supervision (topology-adapted in inferred parameters)
Optimizer	–	SGD with Nesterov momentum ($\mu = 0.99$)
Data augmentation	–	Rotations, scaling, Gaussian noise, Gaussian blur, brightness, contrast, simulation of low resolution, gamma correction and mirroring
Training procedure	–	1,000 epochs × 250 minibatches, foreground oversampling
Inference procedure	–	Sliding window with half-patch size overlap, Gaussian patch center weighting
Intensity normalization	Modality, intensity distribution	If CT, global dataset percentile clipping & z score with global foreground mean and s.d. Otherwise, z score with per image mean and s.d.
Image resampling strategy	Distribution of spacings	If anisotropic, in-plane with third-order spline, out-of-plane with nearest neighbor Otherwise, third-order spline
Annotation resampling strategy	Distribution of spacings	Convert to one-hot encoding → If anisotropic, in-plane with linear interpolation, out-of-plane with nearest neighbor Otherwise, linear interpolation

Image target spacing	Distribution of spacings	If anisotropic, lowest resolution axis tenth percentile, other axes median. Otherwise, median spacing for each axis. (computed based on spacings found in training cases)
Network topology, patch size, batch size	Median resampled shape, target spacing, GPU memory limit	Initialize the patch size to median image shape and iteratively reduce it while adapting the network topology accordingly until the network can be trained with a batch size of at least 2 given GPU memory constraints. for details see online methods.
Trigger of 3D U-Net cascade	Median resampled image size, patch size	Yes, if patch size of the 3D full resolution U-Net covers less than 12.5% of the median resampled image shape
Configuration of low-resolution 3D U-Net	Low-res target spacing or image shapes, GPU memory limit	Iteratively increase target spacing while reconfiguring patch size, network topology and batch size (as described above) until the configured patch size covers 25% of the median image shape. For details, see online methods.
Configuration of post-processing	Full set of training data and annotations	Treating all foreground classes as one; does all-but-largest-component-suppression increase cross-validation performance? Yes, apply; reiterate for individual classes No, do not apply; reiterate for individual foreground classes
Ensemble selection	Full set of training data and annotations	From 2D U-Net, 3D U-Net or 3D cascade, choose the best model (or combination of two) according to cross-validation performance

FIG. 1. nnU-Net automated segmentation pipeline[46]

2.C. Performance Metrics

The performance of the three generated models (2D, 3DFR and 3DCasc) were assessed using the 20 test patients. The predicted and manual contours were compared quantitatively using the Dice Similarity Coefficient (DSC), Hausdorff Distance (HD), 95 percentile Hausdorff Distance (HD95), Mean Surface Distance (MSD) and Precision Score. Where the DSC focuses on the geometric overlap

between the predicted and manual contours, while the HD, HD95 and MSD evaluate the boundary similarity between the manual and predicted segmentations.

2.D. Dosimetric Evaluation

All brachytherapy plans were generated using the Oncentra (Elekta, Stockholm, Sweden, version 4.6) treatment planning system according to the OAR dose constraints and HR-CTV prescribed dose recommended by the American Brachytherapy Society (ABS)[49] and GEC-ESTRO[50–52] guidelines. Assuming $\alpha/\beta = 10$, a prescribed dose (EB + brachy) of $D_{90} > 80$ Gy_{EQD2} for the HR-CTV and assuming an $\alpha/\beta = 3$, the maximum dose (D_{2cc}) to the bladder and rectum (EB + brachy) of 90 Gy_{EQD2} and 75 Gy_{EQD2} respectively. To evaluate the dosimetric differences between the manual contours and predicted contours, the predicted segmentations were imported into the planning system and the original dose distribution was overlaid and cumulative dose volume histograms (DVHs) generated. The Dosimetric discrepancies of $D_{90\%}$ for HR-CTV and D_{2cc} for the OARS were evaluated and compared with the DVH statistics generated from the manual contours.

2.E. Clinician Assessment

Performance and dosimetric evaluation provided a good indication of a model's accuracy, however these metrics alone were not sufficient and required input from clinicians to properly assess the clinical acceptability of the generated patient contours. The OARs and HR-CTVs from each of the 20 test patients were evaluated independently by three different ROs, scoring each patient's bladder, rectum, and HR-CTV on a scale of 1 to 4 where; 1 indicates reject completely, 2 major revisions, 3 minor revisions before clinically acceptable and 4 indicating clinically acceptable as is. Contours scored with a 4, also included ones where ROs would have preferred some small adjustments but were not necessary for the contours to be deemed clinically acceptable. For any contours scoring a 3 or lower, the ROs indicated the time required to adjust the contours to the level of clinical acceptability. These times, along with the model prediction times were compared to the average time taken for ROs to manually contour all the OARs and HR-CTV.

3. RESULTS

3.A. Model Comparisons

The mean ± standard deviation of the performance metrics (DSC, HD, HD95, MSD and precision) between the manual contours and predicted contours of the best performing fold for each of the three generated nnU-Net configurations (2D, 3DFR and 3DCasc), using the 20 test patients, are given in Table I. The best performing configurations were 3DFR and 3DCasc with the mean DSC over all OARs and HR-CTVs for both configurations having a value of 0.85. The mean HD, HD95 and MSD values for the 3DFR and 3DCasc configurations over all OARs and HR-CTVs were approximately 9.9/10.1 mm, 4.8/4.9 mm and 1.48/1.54 mm respectively.

TABLE I. Mean DSCs, MSDs (mm), HDs (in mm) and HD95s (in mm) between the manual and predicted contours for all three configurations.

Contour	Model	DSC	HD (mm)	HD95 (mm)	MSD (mm)	Precision
Bladder	2D	0.87 ± 0.05	18.44 ± 0.91	5.71 ± 3.26	1.53 ± 0.77	0.86 ± 0.08
	3DFR	0.92 ± 0.04	7.52 ± 3.13	3.00 ± 1.09	0.84 ± 0.30	0.91 ± 0.05
	3DCasc	0.91 ± 0.04	8.39 ± 3.49	3.26 ± 1.06	0.94 ± 0.33	0.91 ± 0.05
Rectum	2D	0.80 ± 0.05	15.48 ± 5.55	6.17 ± 2.25	1.68 ± 0.52	0.78 ± 0.05
	3DFR	0.84 ± 0.04	13.78 ± 4.62	5.25 ± 1.78	1.36 ± 0.43	0.84 ± 0.04
	3DCasc	0.84 ± 0.04	13.48 ± 3.86	5.36 ±1.65	1.39 ± 0.41	0.83 ± 0.04
HR CTV	2D	0.78 ± 0.06	11.69 ± 3.22	6.66 ± 2.03	2.64 ±0.92	0.77 ± 0.11
	3DFR	0.81 ± 0.05	8.48 ± 1.78	6.03 ± 2.01	2.23 ± 0.75	0.80 ± 0.09
	3DCasc	0.81 ± 0.05	8.71 ± 2.04	6.14 ± 2.01	2.30 ± 0.72	0.80 ± 0.09

The predicted contours from all three configurations were also visually inspected and the 3DFR configuration was found to provide slightly better predictions when it came to the larger bladder volumes.

3.B. Performance Results

The overall boxplots of the DSC, HD, HD95, MSD and precision scores for each structure of the 3DFR configuration is shown in Fig 2. Almost all predicted contours had performance metrics that were distributed within a certain range, with only a single low DSC and high MSD outlier in the boxplots for the bladder contours. Examples comparing the manual contours to those predicted by the 3DFR nnU-Net configuration are shown in Fig 3.

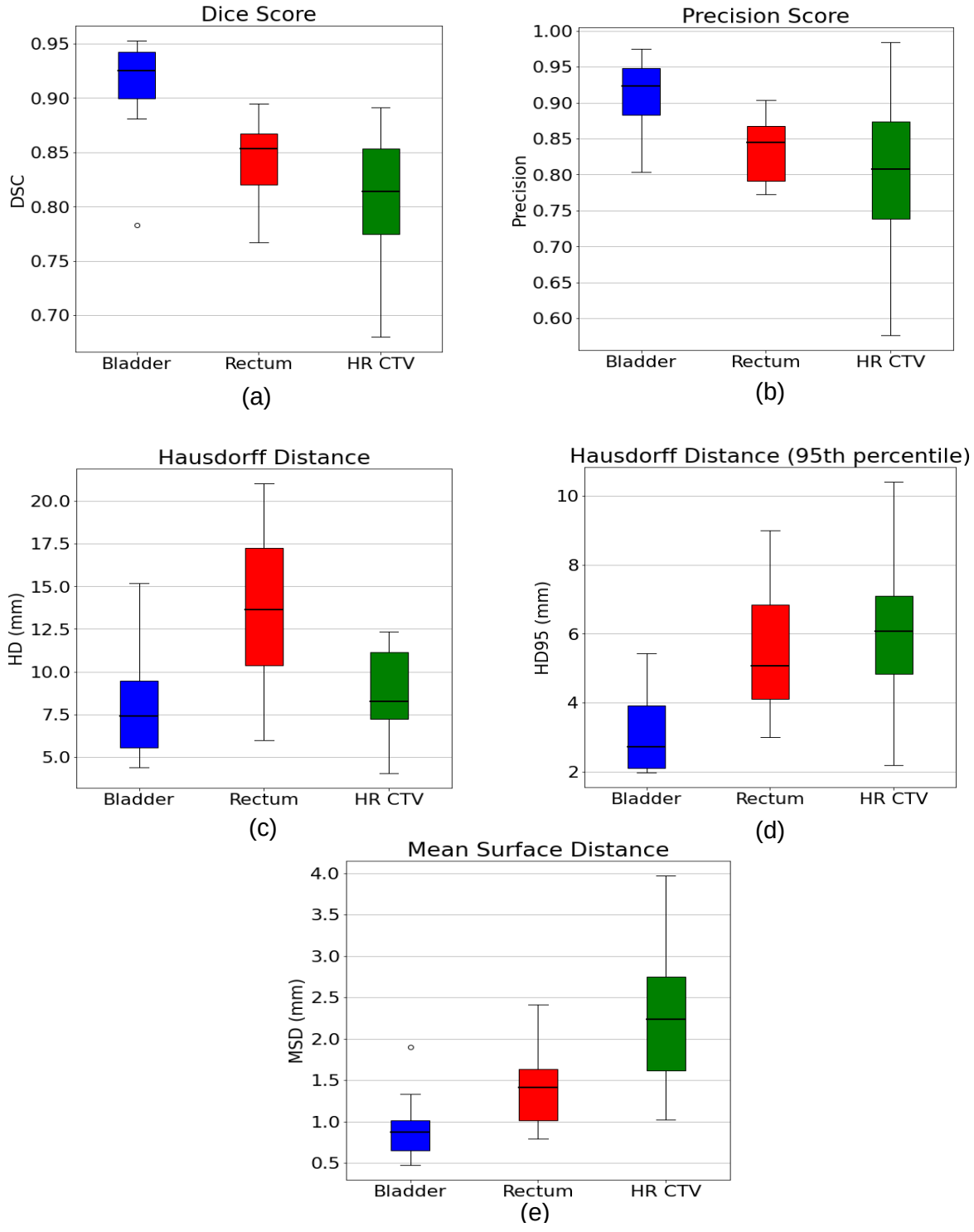


FIG. 2. Box plots showing the performance metrics of the bladder, rectum, and HR CTV for the 3DFR nnU-Net configuration. (a) DSC, (b) precision score, (c) HD, (d) HD95 and (e) the MSD

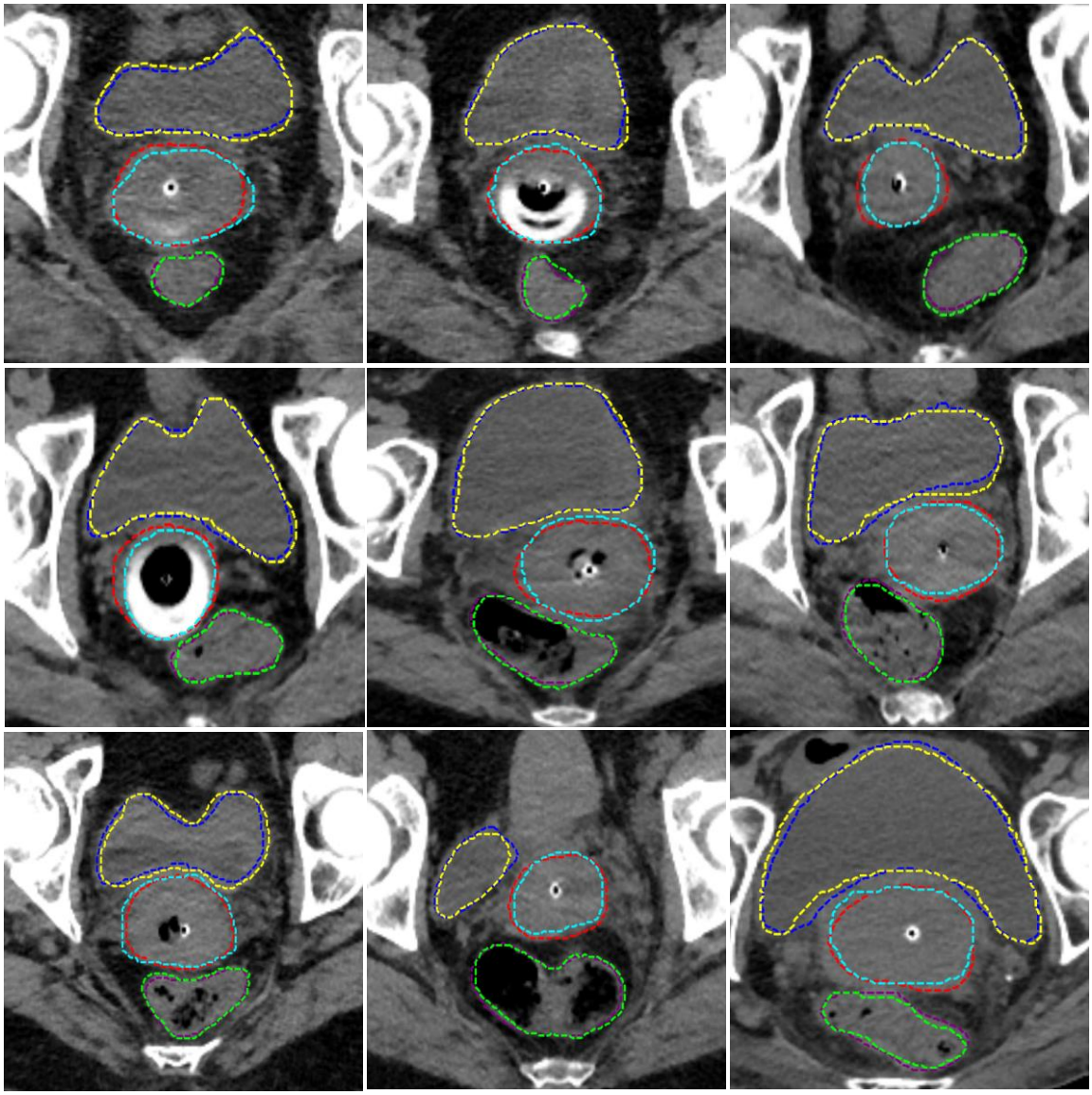


FIG. 3.Examples showing comparisons between manual and predicted contours. Blue = predicted bladder, Yellow = ground truth bladder, Purple = predicted rectum, Green = ground truth rectum, Red = predicted HR CTV, and Light Blue = ground truth HR CTV.

3.C. Dosimetric Evaluation

The dosimetric accuracy of the 3DFR nnU-Net predictions was investigated by looking at the DVH parameter D_{2cc} for the OARs and $D_{90\%}$ for the HR CTV, as well the total volume (in cm^3 or cc) of the OARs and HR CTVs. Boxplots showing the DVH parameters and volumes of all 20 test patients for the manual and predicted contours are given in Fig 4. The dose differences for the DVH parameters (D_{2cc}/D_{90}) for each test patient is shown in Fig 5. The mean dose and volume differences over all OARs and HR CTVs was 0.1 ± 1.2 Gy and 1.1 ± 7.1 cm^3 respectively.

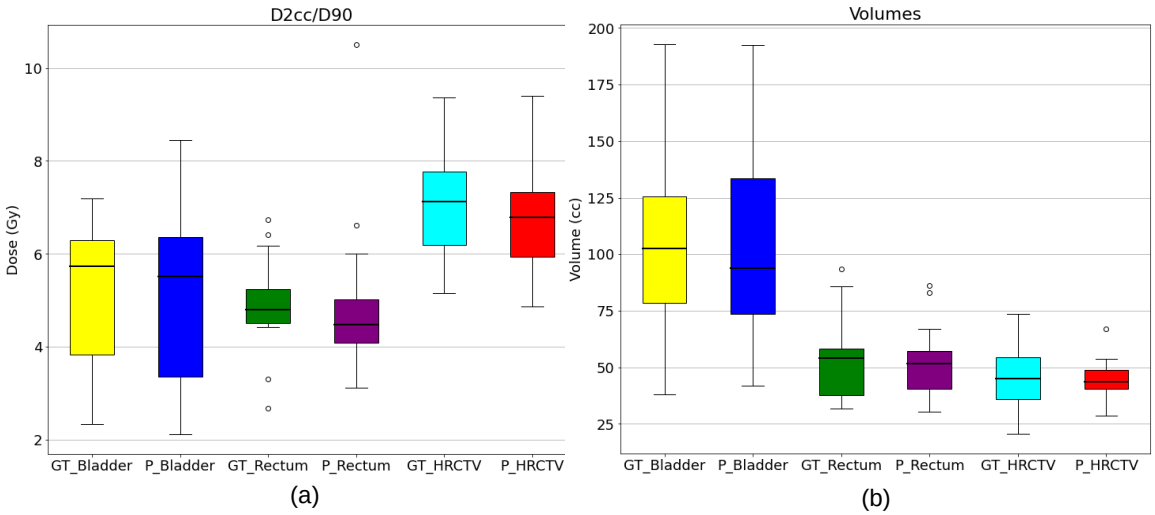


FIG. 4. (a) DVH parameters and (b) volumes for the ground truth and predicted contours. GT = Ground Truth and P = Predicted.

A summary of the mean ± standard deviation of the DVH parameters and volumes for the manual and predicted contours are shown in Table II, along with the mean ± standard deviation of the DVH dose parameters and volume differences.

TABLE II. Results of the dosimetric and volume comparison of the bladder, rectum, and HR CTV for the manual and predicted contours. All values reported as mean ± standard deviation.

	Manual		Predicted		Mean Change	
Contour	D2cc/D90 (Gy)	vol (cm^3)	D2cc/D90 (Gy)	vol (cm^3)	ΔD2cc/90% (Gy)	Δvol (cm^3)
Bladder	5.2 ± 1.5	108.6 ± 45.7	5.1 ± 1.8	107.4 ± 47.5	0.08 ± 1.14	1.3 ± 7.3
Rectum	4.9 ± 1.0	53.0 ± 16.8	4.9 ± 1.6	52.3 ± 14.5	0.02 ± 1.17	0.7 ± 6.2
HR CTV	7.1 ± 1.2	45.6 ± 12.8	6.8 ± 1.2	44.1 ± 8.9	0.33 ± 1.33	1.5 ± 8.0

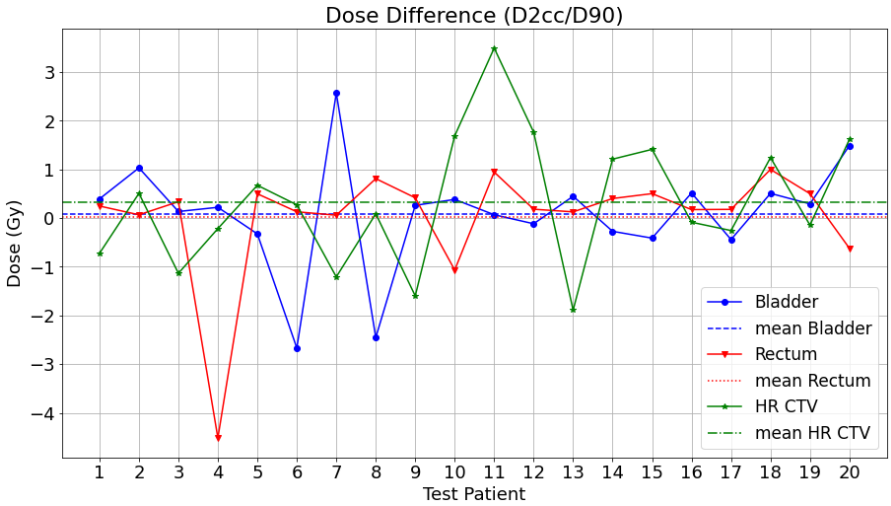


FIG. 5. Plot of the dose difference in the DVH parameters. D2cc for OARs and D90 for the HR CTV.

3.D. Clinician Reviews

The average scores from all three of the ROs reviews of the predicted bladder, rectum and HR CTVs for all 20 test patients are shown in Fig 6. Overall, OARs and HR CTV combined, 65 % of the predicted contours were suitable for clinical use while 33 % required minor adjustments before being clinically accepted, the remaining 2 % required major revisions and 0 % were rejected completely. A summary of the average review scores for the bladder rectum and HR CTV is given in table III.

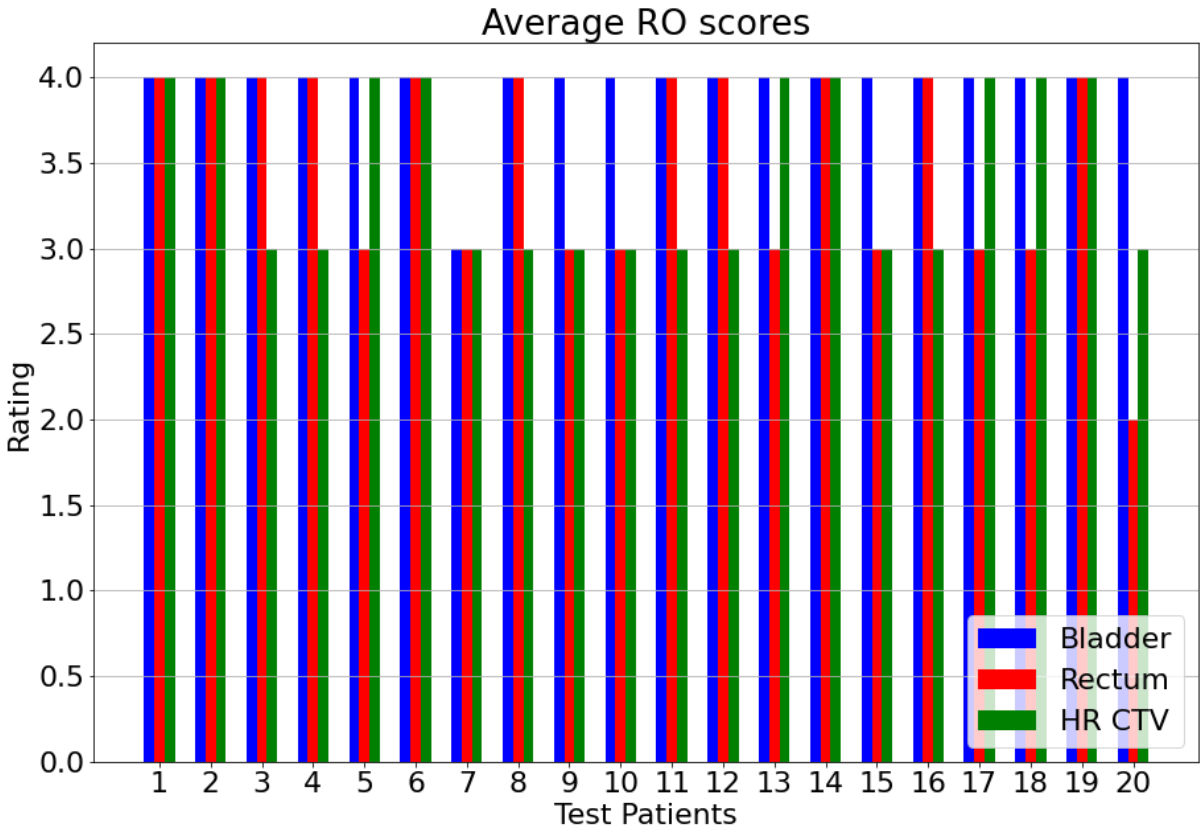


FIG. 6. Average RO reviews of the predicted bladder, rectum, and HR CTV contours for all 20 test patients.

TABLE III. Summary of the average RO review scores for the predicted contours of the bladder, rectum, and HR CTV for the 20 test patients.

Contour	Accept	Minor edits	Major edits	Reject
Bladder	95 % (19)	5 % (1)	0 % (0)	0 % (0)
Rectum	55 % (11)	40 % (8)	5 % (1)	0 % (0)
HR CTV	45 % (9)	55 % (11)	0 % (0)	0 % (0)

For patient contours that scored a 3 or lower, the ROs recorded the time needed to make the necessary corrections to the contours. The mean editing time, along with the mean manual contouring time and model prediction time (for a single 3DFR fold) are given in table IV. The mean time taken for the auto-contouring system to delineate the OARs and HR CTV is substantially less than the mean manual contouring time.

TABLE IV. Results of manual contouring time, model prediction time and editing times. All values reported as mean ± standard deviation.

	Manual	Prediction	Edits
Times (min)	14.0 ± 3.3	1.607 ± 0.004	2.1 ± 1.0

4. DISCUSSION

With only a few lines of code, we successfully implemented the self-configuring nnU-Net for the automatic segmentation of the OARs and HR CTV in HDR cervical brachytherapy. Three nnU-Net configurations were trained (2D, 3DFR and 3DCasc) with a total of 64 training CTs and 16 validation CTs. All three configurations were tested on the remaining 20 CT scans, where the 3DFR and 3DCasc models provided substantially better performance with the 3DFR providing better generalisation over the larger bladder volumes when inspected visually. We demonstrated that the nnU-Net can produce clinically acceptable contours for both the OARs and HR CTV, with significant time saving possible when applying the automatic contouring system.

4.A. Quantitative Results

All three nnU-Net configurations were assessed using the performance metrics described in section 2.C with 3DFR and 3DCasc models significantly outperforming the 2D model ($p < 0.002$, using student T-test) over all OAR and HR CTV contours combined (Table I). There were no statistical differences between the 3DFR and 3DCasc models ($p > 0.863$, using student T-test), however after visual inspection the 3DFR model provided slightly better generalisation when it came to larger bladder volumes. The 3DFR model predicted bladder, rectum, and HR CTV contours with mean DSC/HD95/MSDs of 0.92/3.0 mm/0.8 mm, 0.84/5.4 mm/1.4 mm and 0.81/6.0 mm/2.2 mm respectively. Among all the contours, it was found that the bladder

provided the highest DSC and precision scores, as well as the lowest distance metrics (HD, HD95 and MSD). A comparison of the 3DFR model performance metrics to those found in previous work is shown in table V.

TABLE V. Comparison of performance metric between the current study and previous work.

Studies	HR CTV	Bladder	Rectum
Zhang et al.[40] Method: DSD-UNet	DSC: 83 ± 4 % HD: 8.1 ± 2.3 mm	DSC: 87 ± 3 % HD: 12.1 ± 4.0 mm	DSC: 82 ± 5 % HD: 9.2 ± 4.6 mm
Wong et al.[41]	DSC: 71 % HD95: 9.32 mm	DSC: 92 % HD95: 3.5 mm	DSC: 76 % HD95: 18.4 mm (rectum + sigmoid)
Mohammadi et al.[42] Method: 2D ResU-Net		DSC: 96 ± 4 % HD95: 2.3 ± 3.4 mm	DSC: 97 ± 2 % HD95: 1.4 ± 1.4 mm
Yoganathan et al.[43] 2.5D-InRN	DSC: 85 ± 6 % HD95: 4.9 ± 2.2 mm	DSC: 90 ± 5 % HD95: 6.3 ± 3.4 mm	DSC: 76 ± 7 % HD95: 8.2 ± 4.1 mm
Current Study 3DFR nnU-Net	DSC: 81 ± 5 % HD: 8.5 ± 2.5 mm HD95: 6.0 ± 2.0 mm	DSC: 92 ± 4 % HD: 7.5 ± 3.1 mm HD95: 3.0 ± 1.1 mm	DSC: 84 ± 4 % HD: 13.8 ± 4.6 mm HD95: 5.2 ± 1.8 mm

The performance of the 3DFR nnU-Net configuration compared well to work done in previous studies, where complex model architectures were designed specific to the task of HDR cervical brachytherapy contouring. 3DFR managed to outperform two of the studies when it came to bladder and rectum contouring, beaten only by the 2D ResU-Net by Mohammadi et al.[42] However, this study did not look at the segmentation of the HR CTV, included a larger dataset and contained CT scans where strict bladder and rectal filling protocols were followed, a limitation of this study. Comparing the results of this study with various other works is challenging as each study used different datasets to train and generate the DSC/HD values seen in Table V. The current work provided HR CTV contour performance comparable to those found in other studies.

4.B. Dosimetric Accuracy

Dosimetric evaluation between the predicted and manual contours was carried out for all 20 test patients. There were no statistically significant differences in the mean D_{2cc}/$D_{90\%}$ or volumes for the bladder, rectum, and HR CTV (all $p > 0.8$). The largest

mean change in DVH parameter was for the HR CTV, with a mean change in $D_{90\%}$ of 0.33 Gy and mean change in volume of 1.5 cm^3. The mean change in D_{2cc} and volume for the bladder and rectum (Table II) showed better accuracy than those found in other studies, where Mohammadi et al[42] had mean changes in D_{2cc}/volume of -0.5 Gy/34 cm^3 and 0.3 Gy/17 cm^3 for the bladder and rectum respectively. The largest difference in D_{2cc} of -4.5 Gy (Fig 5.) was observed for a rectum contour of a single test patient, this test patient coincides with the lowest DSC value (0.77) and one of the highest HD values (20 mm) obtained for the predicted rectum contours.

4.C. Clinical acceptability

There was an average acceptance rate 65 % over all the OARs and HR CTVs for the 20 test patients, with an average acceptance rate of 95 % (19 patients) and 55 %(11 patients) for the bladder and rectum respectively. Forty percent (8 patients) of the rectum and 55 % (11 patients) of the HR CTV cases required very minor revisions before being clinically acceptable. Most of the minor revisions for the rectum were the superior border of the rectum, where the model struggled slightly with distinguishing the rectum from sections of the bowel or sigmoid (see section A4 of supplementary material). The minor revisions for the HR CTV were a combination of either incomplete contours at the superior borders or at the inferior border, where too much of the ring was included in the contour. Based on the average scores for the bladder, minor revisions were only required for one patient, where the model failed to accurately determine the bladder wall edge.

The average time taken for ROs to manually contour out the OARs and HR CTV was around 14 minutes, while the average model prediction time was 1.6 minutes and average editing time of approximately 2.1 minutes (for those that required minor revisions before being clinically acceptable). The results show that application of the auto-contouring system, on average, has the capability to substantially reduce the overall required contouring time thereby allowing for a more efficient planning workflow by alleviating the bottleneck experienced at the contouring stage.

This study faced several limitations such as small dataset size, no consistent bladder or rectum preparation and no contrast agent. Over and above these limitations, this study was also hindered by inter observer variations in the manual contours, used for training, as well as inter observer variations between the three ROs scoring the

generated contours. In addition, it is acknowledged that CT-based brachytherapy contours are inferior in accuracy to MRI-based contours.

5. CONCLUSION

We have demonstrated that with only a few lines of code and relatively small patient dataset, one can develop an accurate and robust auto-segmentation model that provides clinically acceptable contours of the OARs and HR CTV in T-R HDR cervical brachytherapy, with no significant dose difference between manual and predicted contours. We have shown that an auto-contouring system can significantly reduce the required contouring time thereby providing a more efficient planning workflow. This is ideally suited to a resource-constrained oncology department using CT-based image guided brachytherapy.

ACKNOWLEDGEMENTS

No funding to declare.

CONFLICTS OF INTEREST

Ethics approval was obtained from the University of Stellenbosch Health Research Ethics Committee (HREC) with project ID 24418 and ethics reference number S22/01/013. The authors have no conflict of interest to disclose.

ADDITIONAL

The nnU-Net code developed by Isensee et al[45] can be found at: https://github.com/MIC-DKFZ/nnUNet.

REFERENCES

1. Summary statistics of cancer diagnosed histologically in 2017 female-all population groups. Accessed January 31, 2022. https://www.nicd.ac.za/wp-content/uploads/2020/12/NCR_2017_Final_02dec2020.pdf
2. Total number of cancer cases (2018). Accessed January 31, 2022. https://www.who.int/cancer/country-profiles/ZAF_2020.pdf
3. Tanderup K, Eifel PJ, Yashar CM, Pötter R, Grigsby PW. Curative radiation therapy for locally advanced cervical cancer: Brachytherapy is NOT optional. *Int J Radiat Oncol Biol Phys*. 2014;88(3):537-539. doi:10.1016/j.ijrobp.2013.11.011

4. Logsdon, M. D., & Eifel, P. J. (1999). FIGO IIIB squamous cell carcinoma of the cervix: an analysis of prognostic factors emphasizing the balance between external beam and intracavitary radiation therapy. *Int J Radiat Oncol Biol Phys*, *43*(4), 763–775. https://doi.org/10.1016/S0360-3016(98)00482-9

5. al Feghali KA, Elshaikh MA. Why brachytherapy boost is the treatment of choice for most women with locally advanced cervical carcinoma? *Brachytherapy*. 2016;15(2):191-199. doi:10.1016/j.brachy.2015.12.003

6. Han K, Milosevic M, Fyles A, Pintilie M, Viswanathan AN. Trends in the utilization of brachytherapy in cervical cancer in the United States. *Int J Radiat Oncol Biol Phys*. 2013;87(1):111-119. doi:10.1016/j.ijrobp.2013.05.033

7. Castelnau-Marchand P, Chargari C, Maroun P, et al. Clinical outcomes of definitive chemoradiation followed by intracavitary pulsed-dose rate image-guided adaptive brachytherapy in locally advanced cervical cancer. *Gynecol Oncol*. 2015;139(2):288-294. doi:10.1016/j.ygyno.2015.09.008

8. Montana GS, Martz KL, Hanks GE. Patterns and sites of failure in cervix cancer treated in the U.S.A. in 1978. *Int J Radiat Oncol Biol Phys*. 1991;20(1):87-93. doi:10.1016/0360-3016(91)90142-Q

9. Hanks GE, Herring DF, Kramer S. Patterns of care outcome studies results of the national practice in cancer of the cervix. *Cancer*. 1983;51(5):959-967. doi:10.1002/1097-0142(19830301)51:5<959::AID-CNCR2820510533>3.0.CO;2-K

10. Coia L, Won M, Lanciano R, Marcial VA, Martz K, Hanks G. The patterns of care outcome study for cancer of the uterine cervix results of the second national practice survey. *Cancer*. 1990;66(12):2451-2456. doi:10.1002/1097-0142(19901215)66:12<2451::AID-CNCR2820661202>3.0.CO;2-5

11. Lanciano RM, Won M, Coia LR, Hanks GE. Pretreatment and treatment factors associated with improved outcome in squamous cell carcinoma of the uterine cervix: A final report of the 1973 and 1978 patterns of care studies. *Int J Radiat Oncol Biol Phys*. 1991;20(4):667-676. doi:10.1016/0360-3016(91)90007-Q

12. Nesvacil N, Tanderup K, Hellebust TP, et al. A multicentre comparison of the dosimetric impact of inter- and intra-fractional anatomical variations in fractionated cervix cancer brachytherapy. *Radiotherapy and Oncology*. 2013;107(1):20-25. doi:10.1016/j.radonc.2013.01.012

13. Saarnak AE, Boersma M, van Bunningen BNFM, Wolterink RÂ, Steggerda MJ. *Inter-Observer Variation in Delineation of Bladder and Rectum Contours for Brachytherapy of Cervical Cancer*. *Radiotherapy and Oncology*, *56*(1), 37–42. https://doi.org/10.1016/S0167-8140(00)00185-7

14. Hellebust, T. P., Dale, E., Skjønsberg, A., & Olsen, D. R. (2001). Inter fraction variations in rectum and bladder volumes and dose distributions during high dose rate brachytherapy treatment of the uterine cervix investigated by repetitive CT-examinations. *Radiotherapy and Oncology*, *60*(3), 273–280. https://doi.org/10.1016/S0167-8140(01)00386-3

15. Duane FK, Langan B, Gillham C, et al. Impact of delineation uncertainties on dose to organs at risk in CT-guided intracavitary brachytherapy. *Brachytherapy*. 2014;13(2):210-218. doi:10.1016/j.brachy.2013.08.010

16. Tanderup K, Nesvacil N, Pötter R, Kirisits C. Uncertainties in image guided adaptive cervix cancer brachytherapy: Impact on planning and prescription. *Radiotherapy and Oncology*. 2013;107(1):1-5. doi:10.1016/j.radonc.2013.02.014

17. Vanderpuye V, Hammad N, Martei Y, et al. Cancer care workforce in Africa: Perspectives from a global survey. *Infect Agent Cancer*. 2019;14(1). doi:10.1186/s13027-019-0227-8

18. Balogun O, Rodin D, Ngwa W, Grover S, Longo J. Challenges and Prospects for Providing Radiation Oncology Services in Africa. *Semin Radiat Oncol*. 2017;27(2):184-188. doi:10.1016/j.semradonc.2016.11.011

19. Ngwa W, Addai BW, Adewole I, et al. Cancer in sub-Saharan Africa: a Lancet Oncology Commission. *Lancet Oncol*. 2022;23(6):e251-e312. doi:10.1016/S1470-2045(21)00720-8

20. Arabi H, Koutsouvelis N, Rouzaud M, Miralbell R, Zaidi H. Atlas-guided generation of pseudo-CT images for MRI-only and hybrid PET-MRI-guided radiotherapy treatment planning. *Phys Med Biol*. 2016;61(17):6531-6552. doi:10.1088/0031-9155/61/17/6531

21. Teguh DN, Levendag PC, Voet PWJ, et al. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. *Int J Radiat Oncol Biol Phys*. 2011;81(4):950-957. doi:10.1016/j.ijrobp.2010.07.009

22. Anders LC, Stieler F, Siebenlist K, Schäfer J, Lohr F, Wenz F. Performance of an atlas-based autosegmentation software for delineation of target volumes for radiotherapy of breast and anorectal cancer. *Radiotherapy and Oncology*. 2012;102(1):68-73. doi:10.1016/j.radonc.2011.08.043

23. Greenham S, Dean J, Fu CKK, et al. Evaluation of atlas-based auto-segmentation software in prostate cancer patients. *J Med Radiat Sci*. 2014;61(3):151-158. doi:10.1002/jmrs.64

24. Voet PWJ, Dirkx MLP, Teguh DN, Hoogeman MS, Levendag PC, Heijmen BJM. Does atlas-based autosegmentation of neck levels require subsequent manual contour editing to avoid risk of severe target underdosage? A dosimetric analysis. *Radiotherapy and Oncology*. 2011;98(3):373-377. doi:10.1016/j.radonc.2010.11.017

25. Zhong H, Kim J, Chetty IJ. Analysis of deformable image registration accuracy using computational modeling. *Med Phys*. 2010;37(3):970-979. doi:10.1118/1.3302141

26. Mohammadi R, Mahdavi SR, Jaberi R, et al. Evaluation of deformable image registration algorithm for determination of accumulated dose for brachytherapy of cervical cancer patients. *J Contemp Brachytherapy*. 2019;11(5):469-478. doi:10.5114/jcb.2019.88762

27. Zhu J, Zhang J, Qiu B, Liu Y, Liu X, Chen L. Comparison of the automatic segmentation of multiple organs at risk in CT images of lung cancer between deep convolutional neural network-based and atlas-based techniques. *Acta Oncol (Madr)*. 2019;58(2):257-264. doi:10.1080/0284186X.2018.1529421

28. Men K, Zhang T, Chen X, et al. Fully automatic and robust segmentation of the clinical target volume for radiotherapy of breast cancer using big data and deep learning. *Physica Medica*. 2018;50:13-19. doi:10.1016/j.ejmp.2018.05.006

29. Sharp G, Fritscher KD, Pekar V, et al. Vision 20/20: Perspectives on automated image segmentation for radiotherapy. *Med Phys*. 2014;41(5). doi:10.1118/1.4871620

30. Wong J, Fong A, McVicar N, et al. Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning. *Radiotherapy and Oncology*. 2020;144:152-158. doi:10.1016/j.radonc.2019.10.019

31. Men K, Chen X, Zhang Y, et al. Deep deconvolutional neural network for target segmentation of nasopharyngeal cancer in planning computed tomography images. *Front Oncol*. 2017;7(DEC). doi:10.3389/fonc.2017.00315

32. Liu Y, Stojadinovic S, Hrycushko B, et al. A deep convolutional neural network-based automatic delineation strategy for multiple brain metastases stereotactic radiosurgery. *PLoS One*. 2017;12(10). doi:10.1371/journal.pone.0185844

33. Men K, Dai J, Li Y. Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks. *Med Phys*. 2017;44(12):6377-6389. doi:10.1002/mp.12602

34. Balagopal A, Kazemifar S, Nguyen D, et al. Fully automated organ segmentation in male pelvic CT images. *Phys Med Biol*. 2018;63(24). doi:10.1088/1361-6560/aaf11c

35. Wang S, He K, Nie D, Zhou S, Gao Y, Shen D. CT male pelvic organ segmentation using fully convolutional networks with boundary sensitive representation. *Med Image Anal*. 2019;54:168-178. doi:10.1016/j.media.2019.03.003

36. Rhee DJ, Jhingran A, Rigaud B, et al. Automatic contouring system for cervical cancer using convolutional neural networks. *Med Phys*. 2020;47(11):5648-5658. doi:10.1002/mp.14467

37. Rigaud B, Anderson BM, Yu ZH, et al. Automatic Segmentation Using Deep Learning to Enable Online Dose Optimization During Adaptive Radiation Therapy of Cervical Cancer. *Int J Radiat Oncol Biol Phys*. 2021;109(4):1096-1110. doi:10.1016/j.ijrobp.2020.10.038

38. Hu H, Shao Y, Hu S. A Review of the Application of Deep Learning in Brachytherapy. *OAlib*. 2020;07(07):1-9. doi:10.4236/oalib.1106589

39. Hu H, Yang Q, Li J, et al. Deep learning applications in automatic segmentation and reconstruction in CT-based cervix brachytherapy. *J Contemp Brachytherapy*. 2021;13(3):325-330. doi:10.5114/jcb.2021.106118

40. Zhang D, Yang Z, Jiang S, Zhou Z, Meng M, Wang W. Automatic segmentation and applicator reconstruction for CT-based brachytherapy of cervical cancer using 3D convolutional neural networks. *J Appl Clin Med Phys*. 2020;21(10):158-169. doi:10.1002/acm2.13024

41. Wong J, Kolbeck C, Giambattista J, Giambattista JA, Huang V, Jaswal JK. Deep Learning-based Auto-Segmentation for Pelvic Organs at Risk and Clinical Target Volumes in Intracavitary High Dose Rate Brachytherapy. *Int J Radiat Oncol Biol Phys*. 2020;108(3):e284. doi:10.1016/j.ijrobp.2020.07.681

42. Mohammadi R, Shokatian I, Salehi M, Arabi H, Shiri I, Zaidi H. Deep learning-based auto-segmentation of organs at risk in high-dose rate brachytherapy of cervical cancer. *Radiotherapy and Oncology*. 2021;159:231-240. doi:10.1016/j.radonc.2021.03.030

43. Yoganathan SA, Paul SN, Paloor S, et al. Automatic segmentation of magnetic resonance images for high-dose-rate cervical cancer brachytherapy using deep learning. *Med Phys*. 2022;49(3):1571-1584. doi:10.1002/mp.15506

44. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88. doi:10.1016/j.media.2017.07.005

45. Isensee F, Kickingereder P, Wick W, Bendszus M, Maier-Hein KH. No new-net. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol 11384 LNCS. Springer Verlag; 2019:234-244. doi:10.1007/978-3-030-11726-9_21

46. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18(2):203-211. doi:10.1038/s41592-020-01008-z

47. Mahantshetty U, Poetter R, Beriwal S, et al. IBS-GEC ESTRO-ABS recommendations for CT based contouring in image guided adaptive brachytherapy for cervical cancer. *Radiotherapy and Oncology*. 2021;160:273-284. doi:10.1016/j.radonc.2021.05.010

48. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Published online May 18, 2015. http://arxiv.org/abs/1505.04597

49. Viswanathan AN, Thomadsen B. American Brachytherapy Society consensus guidelines for locally advanced carcinoma of the cervix. Part I: General principles. *Brachytherapy*. 2012;11(1):33-46. doi:10.1016/j.brachy.2011.07.003

50. Dimopoulos JCA, Petrow P, Tanderup K, et al. Recommendations from Gynaecological (GYN) GEC-ESTRO Working Group (IV): Basic principles and parameters for MR imaging within the frame of image based adaptive cervix cancer brachytherapy. *Radiotherapy and Oncology*. 2012;103(1):113-122. doi:10.1016/j.radonc.2011.12.024

51. Hellebust TP, Kirisits C, Berger D, et al. Recommendations from Gynaecological (GYN) GEC-ESTRO working group: Considerations and pitfalls in commissioning and applicator reconstruction in 3D image-based treatment planning of cervix cancer brachytherapy. *Radiotherapy and Oncology*. 2010;96(2):153-160. doi:10.1016/j.radonc.2010.06.004

52. Pötter R, Haie-Meder C, van Limbergen E, et al. Recommendations from gynaecological (GYN) GEC ESTRO working group (II): Concepts and terms in 3D image-based treatment planning in cervix cancer brachytherapy - 3D dose volume parameters and aspects of 3D image-based anatomy, radiation physics, radiobiology. *Radiotherapy and Oncology*. 2006;78(1):67-77. doi:10.1016/j.radonc.2005.11.014

5. Conclusions

The primary overall objective of this study was to develop, train and implement a deep convolution neural network for the automatic delineation of the OARs and HR CTV in IG-HDR cervical brachytherapy, in an attempt to acquire fast accurate contours that can substantially reduce the overall contouring time in the treatment planning process.

5.1 Summary of findings

Prior to this study, work had been done on developing highly specialised deep learning algorithms for the automatic segmentation of the various structures in HDR cervical brachytherapy. These previous studies however did not analyse the clinical acceptability of the generated contours or possible time saving when implementing an auto contouring system. On top of this, previous studies involved complex model architectures that would require a high level of programming and machine learning knowledge in order to replicate in one's own department.

This study managed to train and evaluate three nnU-Net configurations, 3DFR, 3DCasc and 2D, that were all capable of producing quality OAR and HR CTV contours. Using the metrics described in section 3.1 as well as visual inspection, 3DFR was found to be the best performing model. Comparison with previous studies showed that the 3DFR model managed to outperform a number of the previous studies when it came to delineation of the bladder and rectum, with comparable results for the HR CTV.

The dosimetric evaluation between the generated contours and those manually contoured by the ROs, showed no significant statistical differences. The clinical acceptability of the predicted contours was assessed by three independent ROs, where a large percentage of the test cases were found to be clinically acceptable, with only a small fraction requiring minor edits and only one test patient rectum contour requiring major edits (see Appendix D). The prediction and editing times, when compared to the manual contouring times, showed that with the addition of an automatic contouring system one will be able to substantially reduce the required

contouring time thereby reducing the number of bottlenecks experienced in the IG-HDR cervical brachytherapy workflow.

5.2 Future Work

This study faced several limitations such as small dataset size, no consistent bladder or rectum preparation and no contrast agent. Over and above these pitfalls, this study also experienced limitations due to inter observer variability in the ground truth contours, as well as in the clinical assessments of the generated contours. With focus on reducing these limitations, one should re-train the nnU-Net configurations with a substantially larger dataset that would help to reduce the impact of inconsistencies in OAR preparation and manual contouring.

This study focused primarily on a single time-consuming step in the cervical brachytherapy treatment planning workflow. Another significant contribution to bottlenecks in the planning process is the generation of patient specific treatment plans, where planning times can vary substantially depending on the experience of the MP doing the planning. Automatic plan generation, using DL, along with an auto contouring system, has the potential to substantially reduce the overall planning time.

6. References

1. Cancer Tomorrow. Accessed January 31, 2022. https://gco.iarc.fr/tomorrow/home.
2. Denny L. Control of Cancer of the Cervix in Low- and Middle-Income Countries. *Ann Surg Oncol*. 2015;22(3):728-733. doi:10.1245/s10434-014-4344-8
3. Coleman JS, Cespedes MS, Cu-Uvin S, et al. An Insight into Cervical Cancer Screening and Treatment Capacity in Sub Saharan Africa. *J Low Genit Tract Dis*. 2016;20(1):31-37. doi:10.1097/LGT.0000000000000165
4. SUMMARY STATISTICS OF CANCER DIAGNOSED HISTOLOGICALLY IN 2017 FEMALE-ALL POPULATION GROUPS COMBINED SITE N(OBS) N(ADJ) % CRUDE ASR 95% LCL 95% UCL CUMRISK 0-74 LR 0-74. Accessed January 31, 2022. https://www.nicd.ac.za/wp-content/uploads/2020/12/NCR_2017_Final_02dec2020.pdf
5. Total # cancer cases (2018). Accessed January 31, 2022. https://www.who.int/cancer/country-profiles/ZAF_2020.pdf
6. Thomadsen R, Miller. *From: Practical Radiation Oncology Physics*.; 2016.
7. Tanderup K, Eifel PJ, Yashar CM, Pötter R, Grigsby PW. Curative radiation therapy for locally advanced cervical cancer: Brachytherapy is NOT optional. *Int J Radiat Oncol Biol Phys*. 2014;88(3):537-539. doi:10.1016/j.ijrobp.2013.11.011
8. Logsdon MD, Eifel PJ. *FIGO IIIB SQUAMOUS CELL CARCINOMA OF THE CERVIX: AN ANALYSIS OF PROGNOSTIC FACTORS EMPHASIZING THE BALANCE BETWEEN EXTERNAL BEAM AND INTRACAVITARY RADIATION THERAPY*.; 1999.
9. al Feghali KA, Elshaikh MA. Why brachytherapy boost is the treatment of choice for most women with locally advanced cervical carcinoma? *Brachytherapy*. 2016;15(2):191-199. doi:10.1016/j.brachy.2015.12.003
10. Han K, Milosevic M, Fyles A, Pintilie M, Viswanathan AN. Trends in the utilization of brachytherapy in cervical cancer in the United States. *Int J Radiat Oncol Biol Phys*. 2013;87(1):111-119. doi:10.1016/j.ijrobp.2013.05.033
11. Castelnau-Marchand P, Chargari C, Maroun P, et al. Clinical outcomes of definitive chemoradiation followed by intracavitary pulsed-dose rate image-guided adaptive brachytherapy in locally advanced cervical cancer. *Gynecol Oncol*. 2015;139(2):288-294. doi:10.1016/j.ygyno.2015.09.008
12. Montana GS, Martz KL, Hanks GE. Patterns and sites of failure in cervix cancer treated in the U.S.A. in 1978. *International Journal of Radiation Oncology*Biology*Physics*. 1991;20(1):87-93. doi:10.1016/0360-3016(91)90142-Q
13. Coia L, Won M, Lanciano R, Marcial VA, Martz K, Hanks G. The patterns of care outcome study for cancer of the uterine cervix results of the second national practice survey. *Cancer*. 1990;66(12):2451-2456. doi:10.1002/1097-0142(19901215)66:12<2451::AID-CNCR2820661202>3.0.CO;2-5

14. Hanks GE, Herring DF, Kramer S. Patterns of care outcome studies results of the national practice in cancer of the cervix. *Cancer*. 1983;51(5):959-967. doi:10.1002/1097-0142(19830301)51:5<959::AID-CNCR2820510533>3.0.CO;2-K

15. Lanciano RM, Won M, Coia LR, Hanks GE. Pretreatment and treatment factors associated with improved outcome in squamous cell carcinoma of the uterine cervix: A final report of the 1973 and 1978 patterns of care studies. *International Journal of Radiation Oncology*Biology*Physics*. 1991;20(4):667-676. doi:10.1016/0360-3016(91)90007-Q

16. Applicator Guide, ver 2.0, 2020, Elekta Group, Stockholm, Sweden. Accessed: July 1, 2022. [Online]. Available: https://www.elekta.com/products/brachytherapy/gynecology/

17. Hellebust TP, Kirisits C, Berger D, et al. Recommendations from Gynaecological (GYN) GEC-ESTRO working group: Considerations and pitfalls in commissioning and applicator reconstruction in 3D image-based treatment planning of cervix cancer brachytherapy. *Radiotherapy and Oncology*. 2010;96(2):153-160. doi:10.1016/j.radonc.2010.06.004

18. Pötter R, Haie-Meder C, van Limbergen E, et al. Recommendations from gynaecological (GYN) GEC ESTRO working group (II): Concepts and terms in 3D image-based treatment planning in cervix cancer brachytherapy - 3D dose volume parameters and aspects of 3D image-based anatomy, radiation physics, radiobiology. *Radiotherapy and Oncology*. 2006;78(1):67-77. doi:10.1016/j.radonc.2005.11.014

19. Dimopoulos JCA, Petrow P, Tanderup K, et al. Recommendations from Gynaecological (GYN) GEC-ESTRO Working Group (IV): Basic principles and parameters for MR imaging within the frame of image based adaptive cervix cancer brachytherapy. *Radiotherapy and Oncology*. 2012;103(1):113-122. doi:10.1016/j.radonc.2011.12.024

20. Viswanathan AN, Thomadsen B. American Brachytherapy Society consensus guidelines for locally advanced carcinoma of the cervix. Part I: General principles. *Brachytherapy*. 2012;11(1):33-46. doi:10.1016/j.brachy.2011.07.003

21. Prescribing, Recording, and Reporting Brachytherapy for Cancer of the Cervix: *J ICRU*. 2013;13(1-2):NP.1-NP. doi:10.1093/jicru/ndw027

22. Derks K, Steenhuijsen JLG, van den Berg HA, et al. Impact of brachytherapy technique (2D versus 3D) on outcome following radiotherapy of cervical cancer. *J Contemp Brachytherapy*. 2018;10(1):17-25. doi:10.5114/jcb.2018.73955

23. Simpson DR, Scanderbeg DJ, Carmona R, et al. Clinical outcomes of computed tomography-based volumetric brachytherapy planning for cervical cancer. *Int J Radiat Oncol Biol Phys*. 2015;93(1):150-157. doi:10.1016/j.ijrobp.2015.04.043

24. Harkenrider MM, Alite F, Silva SR, Small W. Image-based brachytherapy for the treatment of cervical cancer. *Int J Radiat Oncol Biol Phys*. 2015;92(4):921-934. doi:10.1016/j.ijrobp.2015.03.010

25. Charra-Brunaud C, Harter V, Delannes M, et al. Impact of 3D image-based PDR brachytherapy on outcome of patients treated for cervix carcinoma in France: Results of the French STIC prospective study. *Radiotherapy and Oncology*. 2012;103(3):305-313. doi:10.1016/j.radonc.2012.04.007

26. Tanderup K, Nielsen SK, Nyvang GB, et al. From point A to the sculpted pear: MR image guidance significantly improves tumour dose and sparing of organs at risk in brachytherapy of cervical cancer. *Radiotherapy and Oncology*. 2010;94(2):173-180. doi:10.1016/j.radonc.2010.01.001

27. Nesvacil N, Tanderup K, Hellebust TP, et al. A multicentre comparison of the dosimetric impact of inter- and intra-fractional anatomical variations in fractionated cervix cancer brachytherapy. *Radiotherapy and Oncology*. 2013;107(1):20-25. doi:10.1016/j.radonc.2013.01.012

28. Saarnak AE, Boersma M, van Bunningen BNFM, Wolterink RÂ, Steggerda MJ. *Inter-Observer Variation in Delineation of Bladder and Rectum Contours for Brachytherapy of Cervical Cancer*. www.elsevier.com/locate/radonline

29. Tanderup K, Nesvacil N, Pötter R, Kirisits C. Uncertainties in image guided adaptive cervix cancer brachytherapy: Impact on planning and prescription. *Radiotherapy and Oncology*. 2013;107(1):1-5. doi:10.1016/j.radonc.2013.02.014

30. Duane FK, Langan B, Gillham C, et al. Impact of delineation uncertainties on dose to organs at risk in CT-guided intracavitary brachytherapy. *Brachytherapy*. 2014;13(2):210-218. doi:10.1016/j.brachy.2013.08.010

31. Paulsen Hellebust T, Dale E, Skjùnsberg A, Olsen DR. *Inter Fraction Variations in Rectum and Bladder Volumes and Dose Distributions during High Dose Rate Brachytherapy Treatment of the Uterine Cervix Investigated by Repetitive CT-Examinations*. www.elsevier.com/locate/radonline

32. Vanderpuye V, Hammad N, Martei Y, et al. Cancer care workforce in Africa: Perspectives from a global survey. *Infect Agent Cancer*. 2019;14(1). doi:10.1186/s13027-019-0227-8

33. Balogun O, Rodin D, Ngwa W, Grover S, Longo J. Challenges and Prospects for Providing Radiation Oncology Services in Africa. *Semin Radiat Oncol*. 2017;27(2):184-188. doi:10.1016/j.semradonc.2016.11.011

34. Rivard MJ, Coursey BM, DeWerd LA, et al. Update of AAPM Task Group No. 43 Report: A revised AAPM protocol for brachytherapy dose calculations. *Med Phys*. 2004;31(3):633-674. doi:10.1118/1.1646040

35. Cokelek M, Holt E, Kelly F, et al. Automation: The Future of Radiotherapy. *International Journal of Radiation Oncology*Biology*Physics*. 2020;108(3):e314. doi:10.1016/j.ijrobp.2020.07.750

36. Aliotta E, Nourzadeh H, Choi W, Leandro Alves VG, Siebers J v. An Automated Workflow to Improve Efficiency in Radiation Therapy Treatment Planning by Prioritizing Organs at Risk. *Adv Radiat Oncol*. 2020;5(6):1324-1333. doi:10.1016/j.adro.2020.06.012

37. Bijman R, Rossi L, Sharfo AW, et al. Automated Radiotherapy Planning for Patient-Specific Exploration of the Trade-Off Between Tumor Dose Coverage and Predicted Radiation-Induced Toxicity—A Proof of Principle Study for Prostate Cancer. *Front Oncol*. 2020;10. doi:10.3389/fonc.2020.00943

38. Arabi H, Koutsouvelis N, Rouzaud M, Miralbell R, Zaidi H. Atlas-guided generation of pseudo-CT images for MRI-only and hybrid PET-MRI-guided radiotherapy treatment planning. *Phys Med Biol*. 2016;61(17):6531-6552. doi:10.1088/0031-9155/61/17/6531

39. Teguh DN, Levendag PC, Voet PWJ, et al. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. *Int J Radiat Oncol Biol Phys*. 2011;81(4):950-957. doi:10.1016/j.ijrobp.2010.07.009

40. Anders LC, Stieler F, Siebenlist K, Schäfer J, Lohr F, Wenz F. Performance of an atlas-based autosegmentation software for delineation of target volumes for radiotherapy of breast and anorectal cancer. *Radiotherapy and Oncology*. 2012;102(1):68-73. doi:10.1016/j.radonc.2011.08.043

41. Greenham S, Dean J, Fu CKK, et al. Evaluation of atlas-based auto-segmentation software in prostate cancer patients. *J Med Radiat Sci*. 2014;61(3):151-158. doi:10.1002/jmrs.64

42. Voet PWJ, Dirkx MLP, Teguh DN, Hoogeman MS, Levendag PC, Heijmen BJM. Does atlas-based autosegmentation of neck levels require subsequent manual contour editing to avoid risk of severe target underdosage? A dosimetric analysis. *Radiotherapy and Oncology*. 2011;98(3):373-377. doi:10.1016/j.radonc.2010.11.017

43. Zhong H, Kim J, Chetty IJ. Analysis of deformable image registration accuracy using computational modeling. *Med Phys*. 2010;37(3):970-979. doi:10.1118/1.3302141

44. Mohammadi R, Mahdavi SR, Jaberi R, et al. Evaluation of deformable image registration algorithm for determination of accumulated dose for brachytherapy of cervical cancer patients. *J Contemp Brachytherapy*. 2019;11(5):469-478. doi:10.5114/jcb.2019.88762

45. van de Velde J, Wouters J, Vercauteren T, et al. Optimal number of atlases and label fusion for automatic multi-atlas-based brachial plexus contouring in radiotherapy treatment planning. *Radiation Oncology*. 2016;11(1). doi:10.1186/s13014-015-0579-1

46. Spinczyk D KA. Simple atlas selection strategies for liver segmentation in ct images. In: Pietka E BPKJWW, ed. *Information Technologies in Medicine*. Springer; 2016:137-148.

47. Bhardwaj A. What are neural networks? An Introduction to machine learning algorithms. https://iaviral.medium.com/what-are-neural-networks-an-introduction-to-machine-learning-algorithms-6b73383c9089.

48. Nousi P, Tefas A. *Deep Convolutional Neural Networks Lightweight Deep Learning*.

49. Saha S. A comprehensive guide to Convolutional Neural Networks. Towards Data Science.

50. Men K, Chen X, Zhang Y, et al. Deep deconvolutional neural network for target segmentation of nasopharyngeal cancer in planning computed tomography images. *Front Oncol*. 2017;7(DEC). doi:10.3389/fonc.2017.00315

51. Liu Y, Stojadinovic S, Hrycushko B, et al. A deep convolutional neural network-based automatic delineation strategy for multiple brain metastases stereotactic radiosurgery. *PLoS One*. 2017;12(10). doi:10.1371/journal.pone.0185844

52. Men K, Zhang T, Chen X, et al. Fully automatic and robust segmentation of the clinical target volume for radiotherapy of breast cancer using big data and deep learning. *Physica Medica*. 2018;50:13-19. doi:10.1016/j.ejmp.2018.05.006

53. Men K, Dai J, Li Y. Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks. *Med Phys*. 2017;44(12):6377-6389. doi:10.1002/mp.12602

54. Wang S, He K, Nie D, Zhou S, Gao Y, Shen D. CT male pelvic organ segmentation using fully convolutional networks with boundary sensitive representation. *Med Image Anal*. 2019;54:168-178. doi:10.1016/j.media.2019.03.003

55. Balagopal A, Kazemifar S, Nguyen D, et al. Fully automated organ segmentation in male pelvic CT images. *Phys Med Biol*. 2018;63(24). doi:10.1088/1361-6560/aaf11c

56. Rigaud B, Anderson BM, Yu ZH, et al. Automatic Segmentation Using Deep Learning to Enable Online Dose Optimization During Adaptive Radiation Therapy of Cervical Cancer. *Int J Radiat Oncol Biol Phys*. 2021;109(4):1096-1110. doi:10.1016/j.ijrobp.2020.10.038

57. Rhee DJ, Jhingran A, Rigaud B, et al. Automatic contouring system for cervical cancer using convolutional neural networks. *Med Phys*. 2020;47(11):5648-5658. doi:10.1002/mp.14467

58. Zhu J, Zhang J, Qiu B, Liu Y, Liu X, Chen L. Comparison of the automatic segmentation of multiple organs at risk in CT images of lung cancer between deep convolutional neural network-based and atlas-based techniques. *Acta Oncol (Madr)*. 2019;58(2):257-264. doi:10.1080/0284186X.2018.1529421

59. Sharp G, Fritscher KD, Pekar V, et al. Vision 20/20: Perspectives on automated image segmentation for radiotherapy. *Med Phys*. 2014;41(5). doi:10.1118/1.4871620

60. Wong J, Fong A, McVicar N, et al. Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning. *Radiotherapy and Oncology*. 2020;144:152-158. doi:10.1016/j.radonc.2019.10.019

61. Moradi S, Oghli MG, Alizadehasl A, et al. MFP-Unet: A novel deep learning based approach for left ventricle segmentation in echocardiography. *Physica Medica*. 2019;67:58-69. doi:10.1016/j.ejmp.2019.10.001

62. Hu H, Shao Y, Hu S. A Review of the Application of Deep Learning in Brachytherapy. *OAlib*. 2020;07(07):1-9. doi:10.4236/oalib.1106589

63. Hu H, Yang Q, Li J, et al. Deep learning applications in automatic segmentation and reconstruction in CT-based cervix brachytherapy. *J Contemp Brachytherapy*. 2021;13(3):325-330. doi:10.5114/jcb.2021.106118

64. Zhang D, Yang Z, Jiang S, Zhou Z, Meng M, Wang W. Automatic segmentation and applicator reconstruction for CT-based brachytherapy of cervical cancer using 3D convolutional neural networks. *J Appl Clin Med Phys*. 2020;21(10):158-169. doi:10.1002/acm2.13024

65. Mohammadi R, Shokatian I, Salehi M, Arabi H, Shiri I, Zaidi H. Deep learning-based auto-segmentation of organs at risk in high-dose rate brachytherapy of cervical cancer. *Radiotherapy and Oncology*. 2021;159:231-240. doi:10.1016/j.radonc.2021.03.030

66. Yoganathan SA, Paul SN, Paloor S, et al. Automatic segmentation of magnetic resonance images for high-dose-rate cervical cancer brachytherapy using deep learning. *Med Phys*. 2022;49(3):1571-1584. doi:10.1002/mp.15506

67. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88. doi:10.1016/j.media.2017.07.005

68. la Macchia M, Fellin F, Amichetti M, et al. Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer. *Radiation Oncology*. 2012;7(1). doi:10.1186/1748-717X-7-160

69. Dice LR. Measures of the Amount of Ecologic Association Between Species. *Ecology*. 1945;26(3):297-302. doi:10.2307/1932409

7. Appendices

7.1 Appendix A: Performance 2D nnU-Net configuration

Table 7-1. Performance Metrics for the bladder of all 20 test patients for the best performing 2D U-Net configuration.

Test Pt	DSC	HD (mm)	HD95 (mm)	MSD (mm)	Precision
1	0.91	6.27	2.88	0.97	0.96
2	0.79	23.11	16.24	3.71	0.88
3	0.93	20.12	2.49	0.76	0.93
4	0.92	6.00	2.49	0.80	0.97
5	0.87	15.43	5.69	1.64	0.92
6	0.92	5.59	2.61	0.80	0.96
7	0.87	13.42	4.67	1.41	0.95
8	0.74	22.46	9.76	2.87	0.90
9	0.91	25.13	6.62	1.48	0.95
10	0.86	30.94	5.61	1.58	0.92
11	0.84	40.48	6.92	1.90	0.93
12	0.91	10.23	4.88	1.26	0.94
13	0.85	10.00	5.06	1.40	0.93
14	0.94	6.85	3.24	0.75	0.95
15	0.91	35.65	6.94	1.65	0.93
16	0.90	9.22	4.53	1.08	0.95
17	0.91	25.29	2.49	0.83	0.95
18	0.87	35.27	5.43	1.48	0.93
19	0.80	14.34	9.10	2.67	0.91
20	0.82	13.02	6.61	1.58	0.92

Table 7-2. Performance Metrics for the rectum of all 20 test patients for the best performing 2D U-Net configuration.

Test Pt	DSC	HD (mm)	HD95 (mm)	MSD (mm)	Precision
1	0.80	10.05	5.44	1.58	0.87
2	0.84	18.79	3.90	1.25	0.89
3	0.80	16.71	5.19	1.34	0.81
4	0.72	27.00	8.66	2.26	0.84
5	0.87	9.25	4.59	1.41	0.91
6	0.82	11.71	6.00	1.64	0.86
7	0.84	14.83	4.09	1.26	0.90
8	0.81	11.51	4.82	1.62	0.88
9	0.85	13.12	4.64	1.40	0.88
10	0.81	10.47	5.36	1.48	0.83
11	0.70	22.56	12.62	2.99	0.79

12	0.79	9.15	4.84	1.38	0.81
13	0.86	9.95	4.22	1.12	0.89
14	0.82	18.75	5.62	1.50	0.85
15	0.76	19.15	9.19	2.45	0.71
16	0.84	11.18	5.00	1.25	0.87
17	0.84	17.94	6.98	1.50	0.76
18	0.72	26.84	8.29	2.48	0.80
19	0.84	12.84	5.06	1.35	0.86
20	0.74	17.83	8.96	2.33	0.73

Table 7-3. Performance Metrics for the HR CTV of all 20 test patients for the best performing 2D U-Net configuration.

Test Pt	DSC	HD (mm)	HD95 (mm)	MSD (mm)	Precision
1	0.78	12.88	7.60	3.16	0.93
2	0.72	12.95	9.64	3.54	0.90
3	0.83	6.83	4.44	1.59	0.93
4	0.82	12.39	5.00	2.00	0.94
5	0.81	9.37	6.08	2.47	0.94
6	0.83	6.96	3.46	1.46	0.95
7	0.83	10.74	4.12	1.79	0.93
8	0.66	12.62	10.22	4.02	0.91
9	0.79	11.04	5.63	2.48	0.94
10	0.71	12.39	8.90	3.70	0.90
11	0.67	18.75	9.26	4.38	0.90
12	0.69	15.30	9.48	4.11	0.86
13	0.83	11.46	5.53	1.90	0.93
14	0.74	13.30	8.03	3.00	0.89
15	0.77	16.88	7.01	2.67	0.89
16	0.78	10.92	6.59	2.49	0.94
17	0.85	6.08	4.59	1.64	0.94
18	0.76	11.46	6.24	2.14	0.94
19	0.81	13.28	6.42	2.57	0.94
20	0.84	8.24	5.01	1.60	0.92

7.2 Appendix B: Performance of 3DCasc nnU-Net configuration

Table 7-4. Performance Metrics for the bladder of all 20 test patients for the best performing 3DCasc U-Net configuration.

Test Pt	DSC	HD (mm)	HD95 (mm)	MSD (mm)	Precision
1	0.93	4.88	2.11	0.70	0.98
2	0.94	11.30	3.23	0.97	0.98
3	0.94	6.03	2.70	0.75	0.97
4	0.94	4.96	2.09	0.68	0.97
5	0.91	15.31	4.07	1.06	0.96
6	0.95	4.39	1.98	0.47	0.97
7	0.89	12.62	4.15	1.24	0.96
8	0.78	12.51	5.64	1.96	0.94
9	0.93	11.35	4.23	1.00	0.96
10	0.90	7.87	3.11	1.01	0.94
11	0.89	6.61	3.17	0.97	0.97
12	0.95	6.61	2.86	0.78	0.97
13	0.88	7.23	3.70	0.97	0.96
14	0.95	6.31	2.49	0.56	0.97
15	0.92	15.05	4.96	1.34	0.92
16	0.89	10.82	4.84	1.23	0.96
17	0.92	6.16	2.40	0.75	0.95
18	0.91	5.33	2.78	0.89	0.95
19	0.94	6.00	2.38	0.77	0.97
20	0.92	6.55	2.38	0.66	0.95

Table 7-5. Performance Metrics for the rectum of all 20 test patients for the best performing 3DCasc configuration.

Test Pt	DSC	HD (mm)	HD95 (mm)	MSD (mm)	Precision
1	0.85	12.27	5.95	1.40	0.90
2	0.86	8.34	4.25	1.21	0.89
3	0.85	10.41	3.68	0.98	0.85
4	0.76	17.51	6.77	1.70	0.88
5	0.90	12.87	3.67	1.14	0.93
6	0.87	12.54	5.53	1.26	0.88
7	0.85	14.83	4.40	1.23	0.89
8	0.80	11.30	4.97	1.63	0.85
9	0.86	18.36	6.03	1.51	0.90
10	0.81	11.18	6.28	1.68	0.84
11	0.84	15.61	7.09	1.38	0.89
12	0.86	7.67	3.78	0.95	0.86
13	0.89	8.90	3.02	0.84	0.92
14	0.83	15.70	5.68	1.48	0.85

15	0.76	20.12	9.96	2.60	0.74
16	0.87	8.00	3.09	0.83	0.90
17	0.89	17.38	5.12	1.15	0.87
18	0.79	19.62	6.76	1.91	0.86
19	0.83	12.84	6.19	1.55	0.87
20	0.84	14.12	5.06	1.43	0.83

Table 7-6. Performance Metrics for the HR CTV of all 20 test patients for the best performing 3DCasc configuration.

Test Pt	DSC	HD (mm)	HD95 (mm)	MSD (mm)	Precision
1	0.80	8.00	8.00	3.11	0.93
2	0.79	10.78	9.03	2.72	0.93
3	0.79	8.52	6.07	2.08	0.91
4	0.88	5.67	3.48	1.23	0.97
5	0.81	8.00	5.80	2.32	0.94
6	0.81	6.03	4.06	1.74	0.95
7	0.84	6.00	4.00	1.86	0.96
8	0.67	13.21	11.27	4.03	0.93
9	0.84	9.85	5.66	1.98	0.95
10	0.75	8.74	8.00	3.38	0.91
11	0.77	8.44	6.00	2.95	0.93
12	0.77	11.22	8.90	3.16	0.88
13	0.83	8.15	4.37	2.14	0.96
14	0.85	9.37	6.19	1.73	0.92
15	0.76	9.85	6.06	2.40	0.92
16	0.87	9.85	4.44	1.34	0.95
17	0.84	6.08	5.19	1.86	0.93
18	0.75	11.56	5.58	2.27	0.94
19	0.86	6.85	4.23	1.83	0.95
20	0.82	8.12	6.44	1.94	0.91

7.3 Appendix C: Performance of 3DFR nnU-Net configuration

Table 7-7. Performance Metrics for the bladder of all 20 test patients for the best performing 3DFR U-Net configuration.

Test Pt	DSC	HD (mm)	HD95 (mm)	MSD (mm)	Precision
1	0.93	4.49	2.11	0.69	0.98
2	0.94	11.63	3.36	1.06	0.97
3	0.95	5.67	2.11	0.56	0.98
4	0.94	4.86	2.00	0.65	0.97
5	0.93	7.67	2.51	0.82	0.97
6	0.95	4.39	1.98	0.47	0.97
7	0.89	13.39	4.59	1.33	0.96
8	0.78	9.09	5.43	1.90	0.94
9	0.93	10.57	3.89	0.93	0.96
10	0.90	8.52	3.68	1.01	0.94
11	0.89	7.92	3.60	0.98	0.96
12	0.95	6.31	2.78	0.74	0.96
13	0.88	7.09	4.01	1.00	0.97
14	0.95	6.31	2.47	0.57	0.96
15	0.92	15.21	4.84	1.25	0.93
16	0.89	10.57	4.80	1.20	0.96
17	0.93	8.00	2.11	0.65	0.97
18	0.91	5.26	2.68	0.93	0.94
19	0.95	4.39	2.09	0.55	0.98
20	0.92	7.11	2.38	0.71	0.95

Table 7-8. Performance Metrics for the rectum of all 20 test patients for the best performing 3DFR configuration.

Test Pt	DSC	HD (mm)	HD95 (mm)	MSD (mm)	Precision
1	0.85	9.76	4.84	1.29	0.91
2	0.86	7.88	4.12	1.25	0.89
3	0.87	8.83	3.00	0.85	0.87
4	0.77	20.14	6.86	1.65	0.88
5	0.87	17.22	7.22	1.62	0.91
6	0.90	13.97	4.63	0.97	0.90
7	0.86	11.38	3.48	1.02	0.89
8	0.81	11.60	5.96	1.66	0.86
9	0.86	17.43	6.01	1.49	0.88
10	0.80	10.59	6.24	1.74	0.84
11	0.83	18.89	9.00	1.60	0.87
12	0.86	9.15	4.11	0.94	0.85
13	0.89	12.51	3.93	0.93	0.90
14	0.83	14.51	5.09	1.45	0.84

15	0.77	21.05	7.84	2.42	0.75
16	0.88	6.00	3.46	0.80	0.91
17	0.89	16.01	4.67	1.08	0.85
18	0.78	20.41	7.02	1.99	0.85
19	0.83	13.28	6.84	1.59	0.86
20	0.84	15.28	5.04	1.37	0.84

Table 7-9. Performance Metrics for the HR CTV of all 20 test patients for the best performing 3DFR configuration.

Test Pt	DSC	HD (mm)	HD95 (mm)	MSD (mm)	Precision
1	0.82	10.00	7.87	2.71	0.94
2	0.78	11.75	9.26	2.86	0.91
3	0.81	8.12	5.02	1.78	0.92
4	0.89	4.05	2.18	1.02	0.98
5	0.80	6.68	6.00	2.68	0.95
6	0.84	5.36	3.76	1.46	0.96
7	0.87	4.56	3.48	1.42	0.97
8	0.68	12.04	10.41	3.97	0.93
9	0.81	11.29	6.85	2.36	0.94
10	0.75	8.24	8.00	3.22	0.91
11	0.76	9.09	6.16	2.91	0.94
12	0.75	12.35	9.79	3.44	0.87
13	0.83	7.67	4.25	2.25	0.96
14	0.85	8.24	6.16	1.61	0.91
15	0.78	11.12	6.25	2.22	0.93
16	0.87	9.42	4.19	1.34	0.96
17	0.86	6.31	5.22	1.62	0.94
18	0.74	11.83	5.59	2.40	0.93
19	0.85	7.42	6.00	2.01	0.95
20	0.81	8.29	6.80	2.00	0.91

7.4 Appendix D: Example of Major revisions

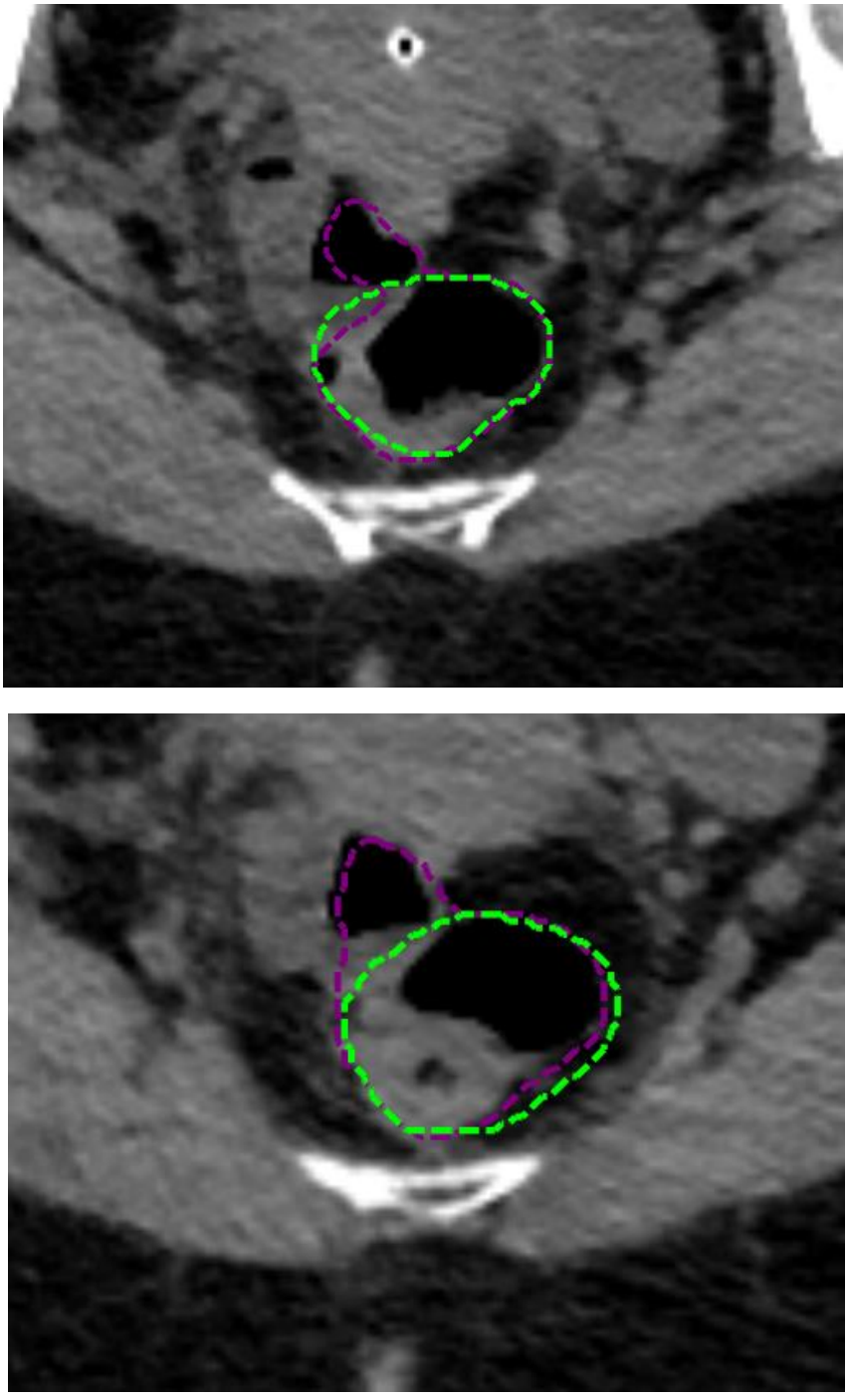


Figure 7-4-1. The single test patient that required major revisions on the rectum contour, where green = ground truth and purple = predicted.