Imperial College London

MEng Individual Project

Department of Computing

Imperial College of Science, Technology and Medicine

Transfer Learning for bespoke automatic contouring of cervical cancer radiotherapy planning

Author:
Anton Zhitomirsky

Supervisor:
Ben Glocker

Second Marker:
TODO

January 13, 2024

Abstract

☞ First draft of abstract

Clinicians target cancerous tumours by studying 3D contrasting images of cancerous tumours and surrounding soft tissues to plan targets for radiation therapy. The Royal Marsden Hospital is a key contributor of data for this project, which uses this approach to delineate tumours for cervical cancers. Typically after a gross tumour volume (GTV) is extrapolated from the relevant imaging modality, clinicians append tailored safety margins to also account for the microscopic cancerous spreads not visible in the scan to generate the planned target volume (PTV).

The PTV area has to be generous enough to attempt to treat the problem in one-shot, yet conservative enough to not harm surrounding healthy tissue with radiation over the course of the treatment. Compounded with small sample size of labelled data this proposes a significant challenge for developing deep-learning segmentation models to identify an optimal PTV.

Thus we propose a transfer learning strategy to utilize imaging models in similar domains to attempt to learn from the limited input size to provide clinicians with a faster and more accurate segmentation method.

Contents

Chapter 1

Introduction

1.1 Clinical Context

1.2 Motivation

1.3 Current Solutions

1.4 Outline of Report

Chapter 2

Background

2.1 Clinical Context

"To administer radiotherapy we need to identify target structures and structures we want to protect (organs-at-risk). These all need to be delineated on each slice of a CT scan, so the amount of radiation planned to reach each area can be calculated and checked.

The clinical target volume is the area where there is likely microscopic cancer and this area needs to receive sufficient radiation doses to achieve cancer cure. As microscopic cancer cannot be seen on CT, the clinical target volume is not an obvious structure that can be drawn around. The CTV is however constructed based on an Oncologist's knowledge of where each particular cancer is likely to spread to. It is drawn based on guidelines, atlases and clinical information. There are currently no internationally agreed guidelines for exactly how the clinical target volume should be drawn for cervix cancer and practice varies. Practice varies based on how the individual components of the CTV are labelled and this in theory could make it difficult for an AI model to learn patterns (or to generate large quantities of similar quality training data).

For the purpose of reproducibility, all scans in this training data have been labelled consistently, to see if an AI model can learn cervical cancer CTV pattern detection. Clinical decisions also affect how many structures are included in the CTV and an AI model is not likely to be able to make these decisions. An AI model is therefore unlikely to be able to produce a perfect CTV for treatment and a clinician will have to select which components of the CTV are required. Training models capable of producing the substructures needed within the CTV that a clinician can select from could still overall save time and improve consistency within the radiotherapy planning process. "

2.2 Vanilla Image Segmentation Models

2.2.1 Convolutional Neural Networks (CNN)

2.2.2 U-Net

2.2.3 nnU-Net

2.2.4 Traditional Limitations

2.3 Transfer Learning

2.3.1 Intuition

2.3.2 Total Segmentator

2.3.3 UniverSeg

2.3.4 SAM

2.4 Summary

Chapter 3

Data

The data is acquired during a CT scan (Section 3.1) and presented as a set of `NIfTI` (Section 3.2) files provided by the Royal Marsden Hospital. The data is of 100 patients each with a variant of cervical cancer. We have obtained from the hospital a spreadsheet with additional notes about each patient which may be useful in training and debugging (Section 3.3). Finally, this data is labelled into 5 different classes as a binary segmentation problem (Section 3.4). Included is a set of 10 hold-out data items, which are patients with only the raw CT scan with no labels.

3.1   CT scan

Before we consider other aspects of the data it is helpful to consider the context from which it was extracted and therefore what we might expect to see. This data is in CT scan, and so will be the focus, although there exist other imaging modalities such as Magnetic Resonance Imaging (MRI) and others.

A CT Scan is an X-ray study, where a series of rays are rotated around a specified body part, and computer-generated cross-sectional images are produced [8]. The granularity or image slice thickness is decided by the operator or physician and ranges from 1mm to 10mm. Whilst the scanner rotates the X-ray tube the patient is slowly moved up or down in the table to produce different cross-section images.

We therefore expect to receive a representation of the internal structure or functions of an atomic region in the form of an array of voxels. A voxel represents the value on a grid in three-dimensional space and is decided by the physician once they establish the slice thickness.

3.2   File Format

The files are stored in a `.nii` file format which defines a style of image called the 'Neuroimaging Informatics Technology Initiative' (NIfTI) [8]. It serves as a lightweight alternative to other formats such as DICOM and eliminates ambiguity from spatial orientation information [10].

The file has a fixed-size header which stores information about the data collected. Table 3.1 summarizes some key attributes of the header. All other attributes not listed are handled by the SimpleITK library [2] which we use to read and manipulate the data in this project. The library defines the image as a set of points in a grid occupying a physical region in space as defined by this metadata, and therefore is influenced by the origin, size, spacing and so on.

Name	Meaning	Value example
dim	Image dimension	3 512 512 193 1 1 1
bitpix	Number of bits per voxel	32
pixdim	The grid spacing (voxel size) and optionally time interval	0 1.3 1.3 2.5 0 0 0 0
xyzt_units	indicates units of pixdim and defined in the C header, e.g. `NIFTI_UNITS_MM = 2`	2

Table 3.1: Description of NIfTI header parameters relevant to this project [10, 6, 3]. Example values are taken from patient id:075.

3.3 Notes

The notes contain information about each of the 100 labelled data pairs [4]. This information can be helpful in debugging or troubleshooting. It also provides a good warning regarding the variability of the data. In particular some aspects to note are summarized in Table 3.2.

Patient ID	Comment	Concern
zzAMLART003	"no GTVp"	Some scans contain no visible tumour, but we still draw a CTV
zzAMLART017	"only scanned bottom of kidneys"	We should be cautious of variability of width scope given in our source data
zzAMLART017	"missing left kidney"	Unusual body anatomy might trip up the model, mentioned elsewhere are also
zzAMLART041	"extra slices"	variability in voxels or quantity may require data pre-processing to eliminate data uncertainty
zzAMLART055	"no contrast - hard to see LNs. NG tube in situ. posterior renal vein. small parametrium and low uterus"	An edge case like this will require more thought

Table 3.2: A few captivating notes about each patient and why it might be concerning

There exist note entires for the majority of the 100 patients which weren't shown in Table 3.2. Regardless, these notes are helpful to identify what type of pre-processing we must do in order to fully address some differences between patients. The concern is to not overfit on the 'normal' cases but also generalize and engineer a solution that is also open-minded to extreme or poorly captured cases; there is a vast variability in the anatomy of patients which makes computer vision tasks more challenging.

3.4 Delineation classes

The clinicians at the Royal Marsden Hospital have provided segmentation labels for 5 high-priority classes of interest. These are the Bladder, Anorectum, CTVn, CTVp, and Parametrium.

3.4.1 Organs At Risk

☞ Insert example figure for bladder and anorectum

An organ at risk is an organ which has a substantial probability of being within the PTV despite being healthy. Any areas that are created around the area should actively avoid these organs because by overlapping with them we risk complicating the treatment and compromising the health of functioning organs.

Many anatomies have been provided in the risk categories, however, in particular we have been supplied with contours for the Bladder (Figure **??**) and the Anorectum (Figure **??**). In particular, clinicians have identified that the Bladder may overlap with the CTVn (Section 3.4.3) and the Parametrium (Section 3.4.4).

3.4.2 CTVp

☞ Find better figure for this

The CTVp stands for the Primary Clinical Target Volume, see the example at Figure **??**. This is the CTV where there may be local microscopic spread (uterus, cervix, upper vagina, primary tumour) [4]. This is the area that contains the tumour.

This isn't by any means an organ in a body, but rather an area comprised of other components formed by joining other structures together. The CTVp is an area defined in Equation 3.4.

3.4.3 CTVn

☞ Find better figure for this

The CTVn stands for Nodal Clinical Target Volume, see the example at Figure **??**. This is the CTV where there may be microscopic spread to lymph nodes. It is drawn based on set margins around pelvic blood vessels and includes pelvic lymph nodes, common iliac lymph nodes and para-aortic lymph nodes [4].

Similarly to CTVp, this is a compound area with three groups of lymph nodes. In clinical practice, the number of these groups included in the CTV varies in each patient, depending on how advanced the disease is. Pathological lymph nodes (GTVn) are also included. The CTVn is an area defined in Equation **??**.

3.4.4 Parametrium

☞ Find better figure for this

The Parametrium (or Paravagina) is the tissue surrounding the cervix/vagina - at risk of local spread, see Figure **??**. Drawn as a complete structure and editing back to the level of vagina to be included [4].

3.5 Establishing Rules for Structures

Notation of Structures

1. Let the Anorectum be denoted as A

2. Let the Bladder be denoted as B

3. Let the Cervix be denoted with C

4. Let the CTVn be denoted with C_n

5. Let the CTVp be denoted with C_p

6. Let the GTVp be denoted with G_p

7. Let the GTVn be denoted with G_n

8. Let the Pelvic Lymph Node be denoted as L_p

9. Let the Common Iliac Lymph Node be denoted as L_i

10. Let the Para-aortic Lymph Node be denoted as L_{pa}

11. Let the Parametrium be denoted with P

12. Let the Uterus be denoted with U

13. Let the Vagina be denoted with V

3.5.1 Relationship between Structures

1. Let O denote the set $O = \{B, A, C_n, C_p, P\}$ for a particular patient. If we want to talk about a specific patient, we should use the super-script notation to differentiate patients, e.g., $O^i = \{B^i, A^i, C_n^i, C_p^i, P^i\}$.

2. Let the overlap of two structures be denoted by the set intersect symbol $\cap$.

3. Let the joint area of two structures be denoted by the set union symbol $\cup$.

3.5.2 Rules

The top 5 priority structures have been selected to identify and plan an area where radiotherapy should be used. With these structures, there are rules that the clinicians have outlined, they are quoted for clarification (these structures only refer to each independent patient):

1. There should be no overlap between the CTVn, CTVp or Anorectum.

$$\forall i, j \in \{C_n, C_p, A\} \text{ with } i \neq j, i \cap j = \emptyset \tag{3.1}$$

2. The Parametrium may overlap with all of the other structures.

$$\forall i \in S, \quad P \cap S_i \neq \emptyset \quad \text{(Possibly)} \tag{3.2}$$

3. The Bladder may overlap with the CTVn.

$$B \cap C_n \neq \emptyset \vee B \cap C_n = \emptyset \tag{3.3}$$

4. The CTVp is defined as a compound structure containing:

$$C_p = \overbrace{C \cup G_p}^{\text{High Risk CTV}} \cup \; U \cup V \tag{3.4}$$

5. The CTVn is defined as a compound structure containing:

$$C_n = G_n \cup L_i \cup L_p + L_{pa} \tag{3.5}$$

3.6 Data Cleaning

☞ TODO

Chapter 4

Evaluation Metrics

To determine if a contour can be used in a clinical context, would be include calculating the difference between the provided labelled data. However, in a delineation context, we have different ways to evaluate this measure.

Suppose we are writing a linear-regression model to match a line onto a set of points. To quantify the performance of our line we would measure the shortest distance between each point and the predicted line. This relies on the assumption of points in a known domain that a model is attempting to approximate. In this case we are fitting a 1-dimensional model onto 0-dimensional points in the grid space.

However, it is far harder to decide on a scoring system when in a delineation context. Consider a single slice of a CT-scan with a known contour around the perimeter of a tumour[1]. A model like those mentioned in Section 2.2 would attempt to learn a function to closely replicate the contour. Here our domain, prediction and ground truth are all 2-dimensional objects.

Geometric measures are the most popular, a survey has found [12]. These measures compare an auto-contour to a ground-truth contour and return a relative score based on its performance.

4.1 Classification Based

Assesses if voxels within and outside the auto-contour have been correctly labelled [12]. To begin, we define 'positive' to mean that the voxel selected is indeed in need of radiotherapy treatment, and 'negative' to mean that the voxel is classified as healthy.

A standard measure of classification is accuracy. It measures the total amount of correct predictions vs the total predictions it made. However, this measure alone isn't enough to fully capture the bias of a model because it doesn't tell you the full story with class-imbalanced data when there isn't an even number between positive and negative labels.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Better measures are Precision and Recall scores. The Precision (also known as the Positive Predictive Value [17]) measures the proportion of predictions that were successfully correct. The Recall (also known as True Positive Rate [17]), on the other hand, "measures the portion of positive voxels in the ground truth that are also identified as positive by the segmentation being evaluated".

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}$$

---

[1]Here we assume that the contour will hug the GTV tightly with no concern for microscopic spread around the remainder of the system

4.2 Spatial Overlap Based

Similarly to Classification Based metrics in Section 4.1, an Overlap Based metric measures the extent of overlap between an auto-contour and a reference structure [12].

The scores above can be combined into a more general score F_β to give

$$\mathrm{F}_\beta = (1 + \beta^2) \cdot \frac{\mathrm{Precision} \cdot \mathrm{Recall}}{\beta^2 \cdot \mathrm{Precision} + \mathrm{Recall}}$$

A specific case of this equation with $\beta = 1$ is mathematically equivalent to the DICE Similarity Coefficient which was found to be the most popular evaluation metric amongst 2021 studies [12, 17].

$$F_1 = \mathrm{DICE} = \frac{2 \cdot \mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} = \frac{2 \cdot TP}{2TP + FP + FN} \quad = \quad \frac{2|S_g^1 \cap S_t^1|}{|S_g^1| + |S_t^1|}$$

Where S_g^1 is the ground truth and S_t^1 is the automatic segmentation. From this relationship, the DICE score has found popularity in image segmentation for similar reasons that the F_1 score has found its popularity classical machine learning; it is able to provide a fair result for imbalanced datasets. This mentality is applicable in our scenario because a tumour will make up very little of the total volume of the domain space.

Another popular related evaluation method is the Jaccard Index, which measures the intersection over the union of two sets:

$$\mathrm{JAC} = \frac{TP}{TP + FP + FN} = \frac{|S_g^1 \cap S_t^1|}{|S_g^1 \cup S_t^1|} \iff \frac{DICE}{2 - DICE}$$

Since the nuberator for the Jaccard Index is smaller (since we avoid the issue of counting the intersecting sections twice) the JAC is always larger than the DICE score.

4.3 Surface Based

Also commonly known as Boundary-Distance-Based Methods [19] compares the distance between two structure surfaces. These can be either maximum distance, average distance or distance at a set percentile of ordered distances [17].

A common example is the Haussdorf Distance. Here, a directed distance metric is defined as the maximum distance from a point in the first set to a nearest point in the other between two individual voxels [19]. Therefore, the better the HD metric, the smaller the value it returns. Here, the distance is taken by some norm, typically Euclidian distnace.

$$\mathrm{HD}(A, B) = \max(h(A, B), h(B, A)), \quad \text{and directed } \mathrm{h}(A, B) = \max_{a \in A} \min_{b \in B} ||a - b||$$

The HD is generally sensitive to outliers. Because noise and outliers are common in medical segmentations, it is not recommended to use the HD directly [19]. Therefore, we can calculate the average directed Haussdorf Distance.

4.4 Estimated Editing Based

Because of the clinical context of evaluating the segmentation by a machine, it may sometimes be helpful to define a performance metric as the "fraction of the surface that needs to be redrawn" [14] since models at this point require manual review to avoid automation bias (Section 6.2). For larger

structures, this method is useful it doesn't assign a lot of weight on the large trivial internal volume which accounts for a much larger proportion of the score.

We therefore combine aspects of Haussdorf Distance from Section 4.3 and DICE from Section 4.2 into a Surface DICE. This assesses the specified tolerance instead of the overlap of the two volumes.
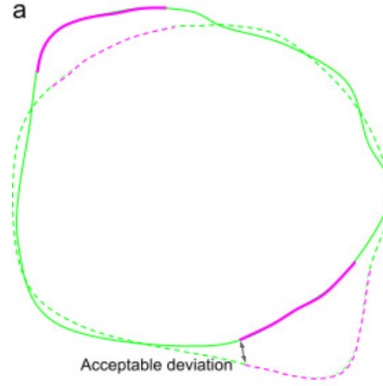
Figure 4.1: Taken from [14]. Illustrates the computation of the surface DICE, where the continuous line is the predicted surface and the dashed line is the ground truth. The black arrows show the maximum deviation tolerated without penalty; therefore, in pink is the unacceptable deviations and green otherwise.

4.5 Volume Based

Volume-based metrics consider only the volume of the segmentation [13, 12, 19]. However, due to its poor spatial descriptions it is more commonly used jointly with other metrics.

$$\text{Relative Volume Difference (RVD)} = \left| \frac{|S_g^1| - |S_t^1|}{|S_g^1|} \right|$$

4.6 Summary

All these methods can be advantageous in some places rather than other. We can begin to list off some challenging scenarios to decide which segmentation is the best.
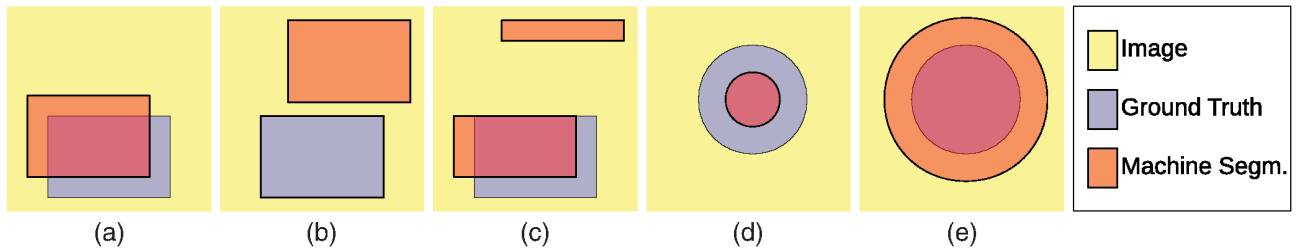
Figure 4.2: Figure from [19] illustrating cases of segmentation to aid with explanation of set-backs of certain evaluation metrics

Classification Based (Section 4.1) and Spatial Overlap Based (Section 4.2) are similar; they are concerned with the number of correctly classified or misclassified voxels without taking into account their spatial distribution. Here, Figure 4.2(a) and Figure 4.2(c) would achieve similar results despite Figure 4.2(a) being locally bound to a better area.

With Haussdorf Distance (Section 4.3) output segmentations generated by Figure 4.2(d) and Figure 4.2(e) will result in the same score, which is not favorable in a radiotherapy planning environment where an organ-at-risk is involved.

This is why we settle at the Surface DICE (Section /refsect:surface-dice) which prioritizes deviation along boundary to a certain degree while measuring the fraction of the surface that needs to be redrawn, thus favouring a more conservative prediction of Figure 4.2(d) instead of (e).

Figure 4.2(b) would score flawlessly when using volumetric score estimation, however, it doesn't take into acount spatial placement, which makes this measurement rather poor when used individually.

For the purpose of this project, we shall select a evaluation measurement which is more bias towards conservative boundary estimates to not touch the organs at risk. This choice was in-part influenced by the clinician's review pipeline; it would easier to correct Figure 4.2(d) instead of (e) because correcting the latter would likely take a considerable amount of time as it would require redrawing almost all of the boundary, whereas the former could be corrected much faster [14].

Chapter 5

Proposal

5.1 Baseline Results

☞ TODO for all

5.1.1 nnU-Net

5.1.2 Total Segmentator

5.1.3 UinverSeg

5.2 Summary

☞ Should we keep labels individual and segment them separately or should we segment it all in one shot?

☞ Solution to resolution problem, down-sample all samples or think of average?

Chapter 6

Ethics

This project involves very intimate and personal information of many female patients. Researchers may collaborate with third-parties by providing anonymized data which may not be reverse engineered back to the patient. The lack of this effort may result in "stigma, embarrassment, and discrimination" [15] if the data is misused.

6.1 Patient disclosures

The Royal Marsden Hospital doesn't require "explicit consent" for sharing collected clinical data with outside entities as long as the patient is made aware of the ways their "de-identified/anonymized" data may be used [18]. Formalities are also arranged with Imperial Collage's Medical Imaging team such as acting as "ethical data stewards" [9]. Without such disclosure and anonymisation of data, patients may be reluctant to provide candid and complete disclosures of their sensitive information, even to physicians, which may prevent a full diagnosis if their data isn't maintained in an anonymous fashion.

The MIRA team acts as responsible data stewards by storing anonymized data within a folder on the college network. All provided data was anonymized by the Royal Marsden Hospital and sent to team MIRA in the `NIfTI` file format which discloses no personal identifiable information, as defined by GOV website [7]. This folder contains security measures which limit the availability of data only to those with specific access rights. Furthermore, operating on the preamble of de-identified data further reduces individual patient risk in the event that data is ever brought outside the confines of this folder.

6.2 Using the tool

The applications of this tool bode well in the healthcare ecosystem as the community slowly realizes the importance of AI-powered tools for the next generation of medical technology. Radiology has been one application that has been most welcoming of the new advances in technology as there is potential for substantial aid by reducing manual labor, increasing precision and freeing up the primary care physician's time [1].

Yet, it is too early to take result the medical tool as gospel. For current cervical radiotherapy delineation tools, only 90% of the output is considered as acceptable for clinical use [11]. The remainder therefore has the potential to cause more harm than good if not checked properly. For example, overlap of a PTV onto an organ-at-risk may invoke a cascade of negative effects for the patient. A potential cause may be the lack of multivariate analysis, where an oncologist would need to consider a variety of data, whereas this model only considers a single point of evidence (results of an imaging modality).

Clinicians can fall into the trap of automation-bias as AI becomes more common place in clinical environments [16]. However, many models of this age codify the existing bias in common cases,

which often will fail those patients who do not fit the expectations of the majority. Therefore, a degree of supervision required from physicians has to be established if this tool is to be used in practice. Oncologists will be required to reverse-engineer results of the 'black-box' to verify why a decision has been made. Secondly, the responsible party for incorrect decisions made by DL tools should also be determined [5].

Bibliography

[1]   Amisha et al. "Overview of artificial intelligence in medicine". In: *Journal of Family Medicine and Primary Care* 8.7 (2019).

[2]   Richard Beare, Bradley Lowekamp, and Ziv Yaniv. "Image Segmentation, Registration and Characterization in R with SimpleITK". In: *Journal of Statistical Software* 86.8 (2018), pp. 1–35. DOI: 10.18637/jss.v086.i08. URL: https://www.jstatsoft.org/article/view/v086i08.

[3]   Hester Breman. *NIfTI-1 Data Format*. URL: https://nifti.nimh.nih.gov/nifti-1/documentation/nifti1diagrams/.

[4]   Institute of Cancer Research and The Royal Marsden Hospital. *AMLART data*.

[5]   Zi-Hang Chen et al. "Artificial intelligence for assisting cancer diagnosis and treatment in the era of precision medicine". In: *Cancer Communications* 41.11 (2021). URL: https://onlinelibrary.wiley.com/doi/10.1002/cac2.12215.

[6]   Bob Cox. *nifti-1 header field-by-field documentation*. URL: https://nifti.nimh.nih.gov/pub/dist/src/niftilib/nifti1.h.

[7]   *Data Protection Act 2018*. URL: https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted.

[8]   Michele Larobina and Loredana Murino. "Medical Image File Formats". In: *Journal of Digital Imaging* 27 (2013). URL: https://link.springer.com/article/10.1007/s10278-013-9657-9.

[9]   David B. Larson et al. "Ethics of Using and Sharing Clinical Imaging Data for Artificial Intelligence: A Proposed Framework". In: (2020). URL: https://pubs.rsna.org/doi/full/10.1148/radiol.2020192536.

[10]   Xiangrui Li et al. "The first step for neuroimaging data analysis: DICOM to NIfTI conversion". In: *Journal of Neuroscience Methods* 264 (2016). URL: https://pubmed.ncbi.nlm.nih.gov/26945974/.

[11]   Zhikai Liu et al. "Development and validation of a deep learning algorithm for auto-delineation of clinical target volume and organs at risk in cervical cancer radiotherapy". In: *Radiotherapy and Oncology* 153 (2020).

[12]   K. Mackay et al. *A Review of the Metrics Used to Assess Auto-Contouring Systems in Radiotherapy*. 2023. URL: https://www.clinicaloncologyonline.net/action/showPdf?pii=S0936-6555(23)00021-3.

[13]   Ying-Hwey Nai et al. "Comparison of metrics for the evaluation of medical segmentations using prostate MRI dataset". In: *Computers in Biology and Medicine* 134 (2021), p. 104497. ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2021.104497. URL: https://www.sciencedirect.com/science/article/pii/S0010482521002912.

[14]   Stanislav Nikolov et al. "Clinically Applicable Segmentation of Head and Neck Anatomy for Radiotherapy: Deep Learning Algorithm Development and Validation Study". en. In: *J Med Internet Res* 23.7 (July 2021), e26151.

[15]   Nass SJ, Levit LA, and Gostin LO. "The Value and Importance of Health Information Privacy". In: (2009). URL: https://www.ncbi.nlm.nih.gov/books/NBK9579/.

[16] Isabel Straw. "The automation of bias in medical Artificial Intelligence (AI): Decoding the past to create a better future". In: *Artificial Intelligence in Medicine* 110 (2020), p. 101965. ISSN: 0933-3657. DOI: https://doi.org/10.1016/j.artmed.2020.101965. URL: https://www.sciencedirect.com/science/article/pii/S0933365720312306.

[17] Abdel Aziz Taha and Allan Hanbury. "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool". In: *BMC Medical Imaging* 15.1 (Aug. 2015), p. 29. ISSN: 1471-2342. DOI: 10.1186/s12880-015-0068-x. URL: https://doi.org/10.1186/s12880-015-0068-x.

[18] The Royal Marsden NHS Foundation Trust. "Privacy Note". In: (2023). URL: https://rm-d8-live.s3.eu-west-1.amazonaws.com/d8live.royalmarsden.nhs.uk/s3fs-public/2023-10/T22020ac_Revised%20privacy%20policy_V1_AW_WEB.pdf.

[19] Varduhi Yeghiazaryan and Irina Voiculescu. "Family of boundary overlap metrics for the evaluation of medical image segmentation". In: *Journal of Medical Imaging* 5.1 (2018), pp. 015006–015006.