The Effect Of Contouring Uncertainty In The Rectum On Clinical Outcome For Patients Treated For Prostate Cancer

Artemis Bouzaki

University of Manchester, School of Physics and Astronomy
and
The Christie Hospital

This project was performed in collaboration with Dylan Green
and supervised by Eliana Vasquez Osorio and Alan McWilliam.

January 2023

Abstract

Manual contouring of organs by clinicians is a standard treatment planning procedure for prostate cancer patients. Inter-observer variation of the target volume has led to efforts for automation of the contouring process to remove uncertainties. The purpose of this study was to quantify the geometric discrepancies between manual and automated rectum contours for a dataset of 1,758 patients. After various filters to ensure validity, 492 patients were kept for analysis. A Student's t permutation test was performed to test the correlation between contouring uncertainties and several rectal symptoms, however, no significance was found. The same test was performed after the removal of 32 potential erroneous patients, yielding the same results. The range of distance discrepancies between manual and automated contours was calculated to be from -1.92 cm to 3.06 cm. A mean of 0.1 cm was found when averaging over all distance discrepancies, suggesting a fundamental error in the methods used.

Contents

1 Introduction

Prostate cancer is the second most common cancer diagnosis made in men and the fifth leading cause of death worldwide. The probability of incidence and death by prostate cancer increases with age, with the average age at the time of diagnosis being 66 years [1]. Treatment planning for prostate cancer is usually done by using 3D Computed Tomography (CT) scans. Those scans are used by oncologists to create delineations of relevant organs, such as the prostate and the rectum, and plan dose delivery. However, inter-observer variation is introduced when contouring target volumes, making the process subjective and inconsistent. Deep learning and machine learning algorithms have been developed that can be applied to CT scans to contour the volumes automatically. Automation of this step can potentially lead to faster treatment planning and consistent treatment results [2]. The treatment of prostate cancer is usually done using external beam radiotherapy, a procedure in which the tumour is irradiated by either proton or photon beams [3]. In this process, healthy tissue of the surrounding organs, such as the rectum, can also be irradiated. Therefore, variations in the contouring procedure can lead to increased severity of rectal-related symptoms post-treatment. The severity of those symptoms is measured by clinical and patient-reported toxicity levels.
In this study, deep learning algorithm delineations were combined with those made by medical professionals for a dataset of 1,758 patients. Those delineations were compared specifically for the rectum, which was the target of this study. The 3D rectal shape was normalised and unfolded into a 2D surface map. The contouring variation between the clinical and software rectum delineations was found at each point of the map. Its correlation to rectal toxicity levels post-treatment was then investigated using a Student's t permutation test.

2 Medical background

2.1 Prostate and prostate cancer

The prostate is a dense fibromuscular gland of the male reproductive system. Its shape can be approximated as a spherical cone and it has the size of a walnut. It consists anatomically of a base, an apex, an interior, a posterior, and two lateral surfaces. The anatomy of the prostate is shown in Figure 1 below. The main purpose of the prostate is to produce a fluid that forms part of the semen [4].

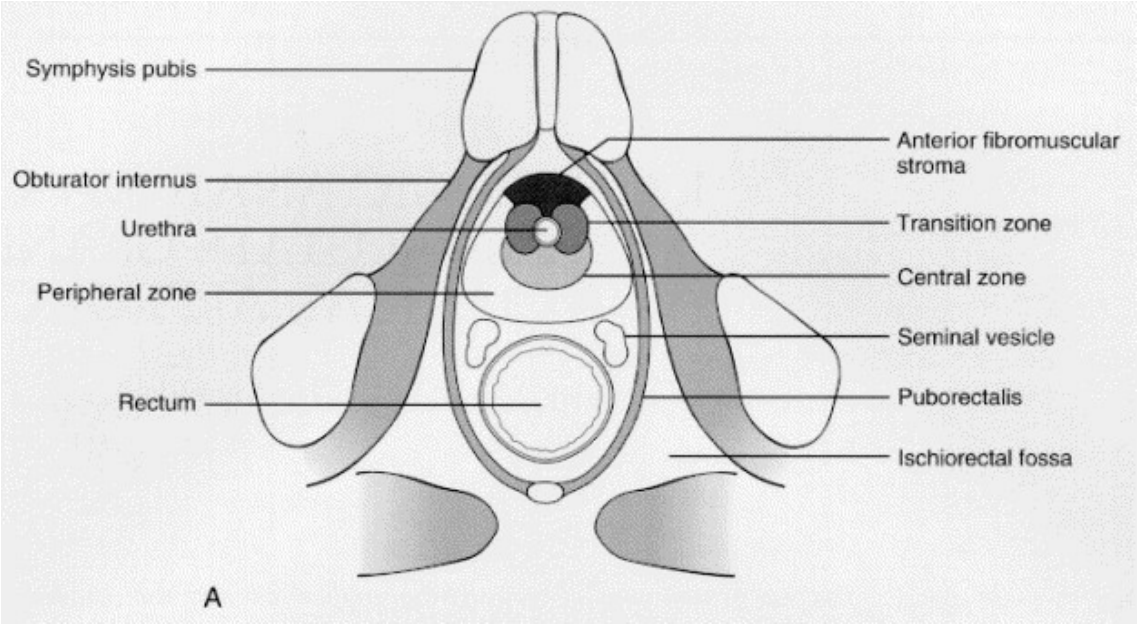

Figure 1: Position and anatomy of the prostate and its surrounding organs [5]. The rectum is approximately cylindrical and lies on the prostate posterior.

The prostate is separated into three different zones: the peripheral zone, the central zone, and the transition zone. The peripheral zone is the largest zone making up 70% of the gland and accounts for 70-80% of prostate cancers. The central zone makes up for 20% of the prostate and accounts for 2.5% of prostate cancers. Prostate cancer in those cases tends to be more aggressive and can metastasise to the seminal vesicles. The transition zone is a small glandular zone that accounts for 10-20% of prostate cancers [6].
As shown in Figure 1, the prostate is located directly below the bladder and surrounds the part of the urethra that is closer to the bladder. It is also close in relation to the rectum on the posterior surface. The rectum is part of the digestive system and is approximately cylindrical in shape [7]. The seminal vesicles, which are responsible for the majority of the fluid that forms the semen, are located between the rectum and the prostate. The ejaculatory ducts that emerge from the seminal vesicles enter the prostate and lead to the urethra [4].

2.2 Symptoms and Diagnosis

Prostate cancer does not usually cause symptoms in the early stages. Most prostate cancers start in the prostate gland and thus to cause symptoms, the cancer tumour needs to press the urethra. If prostate cancer is metastatic, meaning that it has spread beyond the prostate, it can cause symptoms such as back or bone pain, fatigue, and weight loss [8].
The main methods through which prostate cancer is diagnosed are the following: physical tests, imaging tests, blood tests, tissue tests, and genetic tests [8]. An important tool used to identify the risk of prostate cancer is Prostate-Specific Antigen (PSA) level tests [9].

2.3 Treatment planning

Treatment planning for prostate patients in this study was done by utilising computed tomography scans.

2.3.1 Computed tomography (CT) scans

Computed tomography (CT) is an imaging method that uses x-rays to build cross-sectional images of the body. Those cross-sectional images are more commonly called slices. CT is based on the principle that the density of the tissue passed by the x-ray beam can be measured from the calculation of the attenuation coefficient. Using this principle, CT allows the reconstruction of the density of the body [10].
X-rays mainly interact with matter through the photoelectric effect and Compton scattering. In the photoelectric effect, a photon gets absorbed by an electron orbiting an atom, ejecting the electron. X-rays can be produced following atomic de-excitations. In Compton scattering, a photon inelastically collides with an electron which results in a change of direction by an angle θ and energy loss of the incident photon [11]. Both effects are present in the energy range used in CT scans and lead to the reduction of the original x-ray energy [12]. The general form of x-ray attenuation is [12]

$$I_x = I_0 \exp^{-ax}, \tag{1}$$

where I_0 is the x-ray intensity before the object is reached, I_x is the x-ray intensity after penetrating through the object of thickness x, and a is the x-ray linear attenuation coefficient. For inhomogeneous objects like the human body, the attenuation of x-rays is calculated using the Lambert-Beer law [12]

$$I_x = I_0 \exp^{-\int a(x)\,dx}. \tag{2}$$

This equation can be used to deduce the attenuation coefficient and thus the density of the tissue the x-ray went through.

2.3.2 CT data acquisition and image reconstruction

CT is accomplished in three steps: data acquisition by scanning the patient, image reconstruction, and displaying the image. To acquire a CT image, the patient is placed on a bed. In fourth-generation computed tomography, an x-ray source, which is located inside a gantry, rotates around

the patient. X-rays that pass through the tissue of the patient are detected on the opposite side using a ring of detectors.

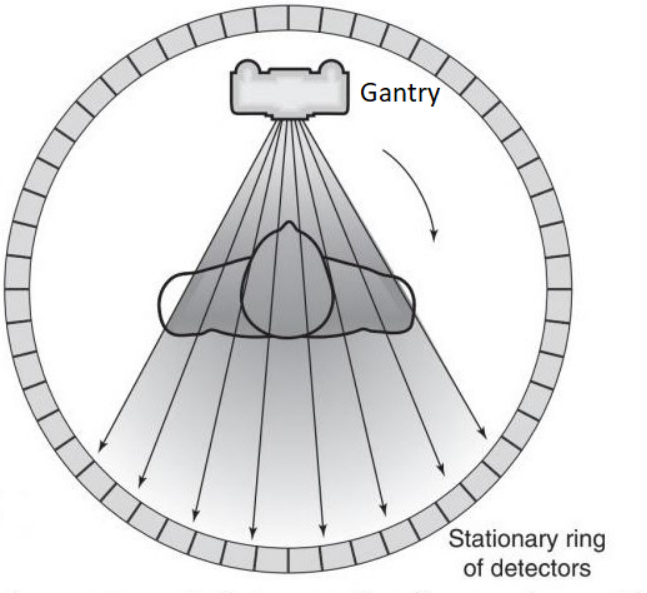

Figure 2: The operation of fourth-generation computed tomography [13].

The scanning process produces the image by the attenuation of the x-ray beam since the patient absorbs the radiation in varying amounts depending on its interactions with the different tissue types. The exit radiation is collected by the detector array and transmitted to the computer for processing. This process is called image reconstruction [10].
The information acquired from the scan is recorded in digital form by the computer. From this information, the computer reconstructs the image using reconstruction algorithms. The algorithms are specific for the type of equipment and software options used. In general, the processing procedure can affect both the resolution and the appearance of the image [14].
Other common types of imaging tests include MRI scans, PET scans, and bone scans. Different types of imaging tests have different uses for prostate cancer. MRI scans are typically used for diagnosing and investigating treatment failure [15]. PET scans are also used to investigate treatment failure but they require specific tracers [16]. Finally, bone scans are used to identify metastatic prostate cancer and specifically bone metastasis, which is the most common metastasis for patients with prostate cancer [17].

2.3.3 Contouring for treatment planning

Oncologists utilise CT scans by creating delineations of organs for each slice of the scan. The process is complex and requires the oncologist to outline the target volume to be irradiated while sparing the organs at risk (OARs) [18]. Organs at risk for prostate cancer typically include the bladder, rectum and seminal vesicles [19].
The target volume is divided into different subsets: the Gross Tumour Volume (GTV), the Clinical Target Volume (CTV) and the Planning Target Volume (PTV). The GTV is defined as the gross palpable or visible size of the tumour. The CTV indicates the suspected microscopic extension of the disease and is made by creating a margin around the tumour. Both CTV and GTV have biological meanings. The PTV is created to ensure that the radiation is indeed delivered to the CTV and it is covered by over 95% of the dose prescription. Contrary to the other two volumes, PTV has geometrical and not biological meaning [20]. In the case of prostate cancer, the target volume mainly includes the prostate, but in some cases, oncologists decide to irradiate part of the seminal vesicles [21].
Manual contouring of the target volume and organs at risk is inconsistent and can vary depending on the observer, a phenomenon which is known as Inter-Observer Variation (IOV). This inter-observer variation has been shown to affect the treatment process [22]. Various studies such as

[23], [24] have shown a significant IOV in the boundary of the prostate with the rectal wall (prostate posterior). The potential effect of IOV on rectal toxicity was the motivation for this study.

2.3.4 Machine learning in treatment planning

In order to manage the effect of IOV and the inconsistency it creates, machine learning and deep learning algorithms have been developed to automate the contouring process. Manual delineation done by oncologists is still widely used in treatment planning for prostate cancer despite the effort put into contouring automation [2].
In this study, each patient had contour delineations that were created using a contouring software: 'Advance Medical Image Registration Engine' (ADMIRE v1.12). This software was trained using deep learning algorithms. This software has the potential advantage of removing the bias created by IOV.

2.4 Radiation therapy

Radiation therapy or radiotherapy is a type of therapy that uses ionising radiation and it is provided as part of cancer treatment to control or kill malignant cancerous cells while sparing normal surrounding tissue. Ionising radiation used in radiotherapy treatment works by damaging the DNA of cancerous tissue leading to cellular death [25]. In the case of prostate cancer, one of the surrounding organs that can be damaged during radiotherapy treatment is the rectum [19].
Radiation exposure is measured in a unit called the gray (Gy). The radiation exposure, also referred to as absorbed dose, is equivalent to the radiation energy deposited per kilogram of substance [25].
There are two types of radiation therapy: external beam radiotherapy and internal radiotherapy [25]. The patients used for analysis in this study were treated using external beam radiotherapy.

2.4.1 External beam radiotherapy

External beam radiotherapy is the type of radiotherapy in which the ionising radiation is delivered with external X-ray or proton beams [3].
Photon and proton beams are the most common in radiotherapy treatment for prostate cancer. Photons can reach tumours deep inside the body. As they travel through the body, photon beams scatter radiation along their path and do not stop once they reach the tumour but go into normal tissue past it. Proton beams can also reach tumours deep in the body. However, proton beams do not scatter radiation on their path through the body and they stop once they reach the tumour [26]. To spare normal tissue, radiation beams are aimed from several angles at the tumour, providing a much larger absorbed dose there than in the surrounding healthy tissue [3].

2.5 Measuring treatment success

2.5.1 Common Terminology Criteria for Adverse Events (CTCAE)

The National Cancer Institute Common Terminology Criteria for Adverse Events is a terminology that is used for reporting and scaling Adverse Events (AE). An Adverse Event (AE) is "any unfavourable and unintended sign, symptom, or disease associated with the use of a medical treatment or procedure". Those symptoms do not need to be directly associated with the treatment [27]. Typical rectal symptoms that can be developed post-treatment for prostate cancer include rectal bleeding, proctitis, and diarrhoea [19]. However, multiple studies show that rectal involvement in prostate cancer is rare despite how close the rectum is to the prostate [28], [29].
A grading scale is provided for each AE term and it indicates its severity. The CTCAE displays grades 1 through 5 which have the following descriptions:

- Grade 1: Asymptomatic or mild symptoms.
- Grade 2: Moderate symptoms for which minimal intervention is needed.
- Grade 3: Severe or medically significant symptoms which are not immediately life-threatening.
- Grade 4: Life-threatening symptoms for which urgent intervention is needed.
- Grade 5 Death related to the symptom.

According to this scale, only symptoms of grade 2 or above require medical intervention. A symptom is graded with 0 when a patient does not display it [27].

2.5.2 Patient reported outcomes

A patient-reported outcome is defined as any report of the status of a patient's health condition that comes directly from the patient. Patient-reported outcomes are not modified or interpreted by a clinician [30].
In this report, patients filled out questionnaires at different time periods since the start of treatment. The symptoms were graded by the patients on a scale from 0 to 4. The grading of the symptoms was considered similar to that of CTCAE scale.

2.6 Anatomical positions

In Figure 3 the standard anatomical positions are shown. This terminology was used consistently throughout this study.

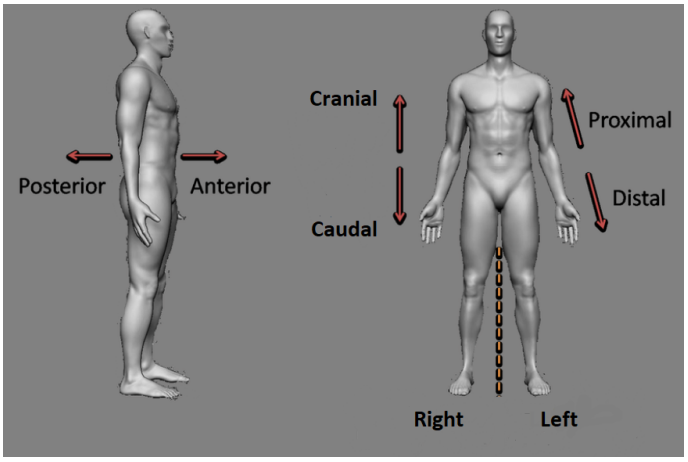

Figure 3: Schematic diagram showing the standard anatomical positions used in medicine [31].

The y axis in this report is referring to the craniocaudal axis. A specific slice of the rectum in the craniocaudal axis (which has a given y coordinate) spans the x-z plane which is referring to the plane defined by the left/right and posterior/anterior axes. This notation will be used later in Section 4.4.

2.7 Patient cohort

The original data sample consisted of 1,758 patients from 26 hospitals in 8 different countries between April 2014 and March 2017 [32]. For those patients there existed a record of different symptoms and their clinical and patient-reported toxicity levels. The toxicity level data were obtained before treatment (at 0 months), as well as at 1 month, 12 months, 24 months, and 36 months since the start of treatment. The clinical toxicity levels were graded according to the CTCAE scale and the patient-reported toxicity levels followed the patient-reported outcome procedure.
Full access was available to the parameters of each patient's 3D CT scan and dose distribution, clinical contours made by the oncologist as well as the automated contours made by ADMIRE. The data was inconsistent in slice length, name of contours, and patient orientation on the bed between different countries and hospitals.
For notation purposes in this report, clinical contours refer to the delineations created by the oncologist whereas software contours refer to delineations created by the deep learning algorithm ADMIRE.

2.8 Worldmatch software

Visualisation and manipulation of the CT scans, as well as the delineations and dose distribution, were performed using an in-house developed software called Worldmatch. Worldmatch is used for analysing medical imaging data at The Christie Hospital due to its wide range of functions. Lua programming language was used to write functions for the software in the Zerobrane studio Integrated Development Environment (IDE). Worldmatch presented the CT data in a series of 2D slices.

3 Statistical background

In this section, a few statistical methods that were used are outlined.

3.1 Sørensen–Dice coefficient

The Sørensen–Dice coefficient or similarity coefficient is a statistical tool used to extract the similarity between two samples. It is also widely used in medical research to measure the spatial overlap between two 3D volumes V_1 and V_2 of organs. The Sørensen–Dice coefficient is defined as [33]

$$SDC(V_1, V_2) = 2\frac{V_1 \cap V_2}{V_1 + V_2}, \tag{3}$$

where $V_1 \cap V_2$ is the intersection between volumes V_1 and V_2.
The Sørensen–Dice coefficient was used in this project to aid the identification of the rectum contours, as seen in Section 4.2.1.

3.2 Hypothesis testing

In order to test the relationship between contouring uncertainties and patient toxicity levels, the statistical method of hypothesis testing is used. In hypothesis testing, there are two types of hypothesis, the null hypothesis, H_0, and the alternative hypothesis, H_1 where

- H_0: Hypothesis of no experimental effect.
- H_1: Hypothesis of experimental effect.

In order to prove H_1, one must have significant evidence to reject H_0. Assuming that H_0 holds, the significance level, α, can be defined as the threshold probability below which H_0 is rejected. Typical values include $\alpha = 0.05$ or $\alpha = 0.01$. Under the same assumption, the significance probability, p, can also be defined as the probability that a test statistic is obtained that is as or more extreme than the observed value. If $p < \alpha$, then one can reject H_0 with confidence 100(1 - p)% [34].
After the null hypothesis is formed, hypothesis testing has the following process. The appropriate test is decided, and thus the relevant test statistic T is deduced. From the distribution of the T statistic, a probability threshold value α can be set below which the null hypothesis can be rejected. In this study, the test that was considered appropriate was a Student's t permutation test.

3.3 Permutation tests

A permutation test involves two or more samples that, under the null hypothesis, come from the same distribution. A permutation test is used to obtain the distribution of the test statistic. This is done by calculating all possible values of the test statistic under possible rearrangements of the data.
A permutation test thus re-samples the labels of the data from two or more different samples under the null hypothesis [35].

3.4 Student's t-test

Student's t-test is a method of testing the null hypothesis for a sample when the population standard deviation is unknown. The test statistic t is assumed to follow a Student's t-distribution under the null hypothesis and the sample means are assumed to be normally distributed. This assumption is allowed because of the Central Limit Theorem, which states that for large sample sizes drawn randomly from a population, the distribution of the means of those samples will approximate a normal distribution [36].
The Student's t-test statistic is defined as

$$t = \frac{\bar{X} - \mu}{s}, \tag{4}$$

where $\bar{X}$ is the sample mean of sample with size n, μ is the population mean and s is the variance of the mean. According to the central limit theorem, $\bar{X}$ follows a normal distribution with mean μ and variance s.
Assuming a certain significance value α, a threshold value of t can be deduced and used to test the null hypothesis [37]. The Student's t-test was considered appropriate in this study since it aids the comparison of two independent groups under the null hypothesis.

4 Method

4.1 Combining delineations

For the analysis to be carried out, the CT scans had to be combined with the clinical delineations, software delineations, and data for the dose in a single file. This was done by loading the scans along with the delineations and the dose on Worldmatch and saving the combined file. After sampling through all 1,758 patients, there were only 1,030 patients for which this process was successful. The reason the sampling failed for a fraction of the patients was mainly due to the naming of the files.

4.2 Identifying clinical rectum contours

The identification of the clinical rectum delineation posed several challenges due to the variety of languages and hospital standards for contour naming. The method for identifying the clinical rectum involved two main steps: the first step was identifying the rectum by the name of the delineation given by the doctor. In the case where the first step had been unsuccessful in finding the contour, the next attempt would be to identify the contour using a method called a distance map.

4.2.1 Distance mapping

It is possible to create a "burn [1]" of a delineation in Worldmatch, which sets the voxel values within that delineation to a value from 0 to 255, which correspond to black and white respectively. The value of 255 is most commonly used as it creates more contrast. This "burned" delineation can then be used to create a distance map with said delineation as the origin. Once the distance map is created, voxel values store information about the distance from the "burned" delineation in cm. It follows that values of 0 would correspond to a voxel that is exactly at the boundary of the delineation and negative values correspond to points inside the delineation. Distance maps were used extensively in this project as they were the main method for finding geometric uncertainties between clinical contours and software contours.
To identify the clinical rectum with this method, a distance map of the software rectum delineation had to be created. This allowed the extraction of the mean distances, d, of all delineations with respect to the software rectum contour. The absolute value of d was calculated to account for contours that existed inside the software rectum contour. The software rectum contour was used as a reference because its naming was consistent. It was expected that the minimum absolute distance from the software rectum contour should come from the clinical rectum contour.
A text file with the patient ID and the delineation that was identified as the rectum was created

[1] The word burn is used here as this is the notation used in Worldmatch.

to validate whether the clinical rectum was identified correctly. Although distance mapping was an adequate method to identify the clinical rectum contour for a lot of patients, there was a small percentage of patients for which the rectum identification was incorrect due to delineations that were very similar to the clinical rectum delineations. Examples of such delineations include the anus and the non-PTV volume of the rectum.
Sørensen–Dice coefficient was implemented in the method as a way to further validate the distance map. All of the delineations were "burned" in order, and the Sørensen–Dice coefficient was calculated for each using Equation 3, where V_1 represented the volume of the "burned" delineation and V_2 represented the volume of the "burned" software rectum delineation. The value of ψ was calculated for each delineation, where

$$\psi = \frac{|d|}{1 + SDC(V_1, V_2)}. \tag{5}$$

According to this equation, smaller values of ψ would correspond to increased similarity.
This validation method eliminated most of the errors imposed by the distance map. There still existed some patients for which the anus was misidentified as the rectum due to the delineations being very similar. Those patients, however, were removed by subsequent filters such as the length of the rectum contours.

4.3 Length of clinical rectum contours

To ensure consistency in the analysis process, a suitable rectum length had to be decided and applied to all the patients. The process followed to decide on an optimal length is described below.

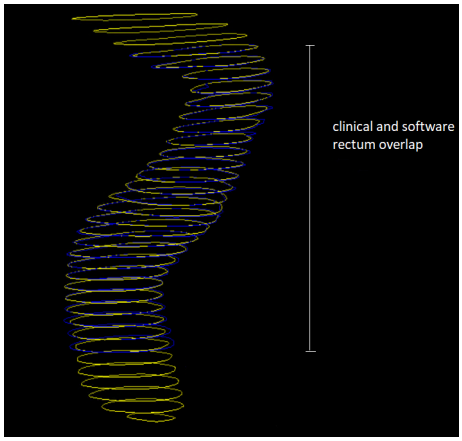

Figure 4: Figure showing an example of the difference in length (slices) between the clinical rectum delineation (yellow) and software rectum delineation (blue) for a patient. The image was taken directly from Worldmatch.

For the patients whose clinical rectum contour had been identified, the overlap between the clinical contour and the software contour in the craniocaudal axis was calculated. This overlap therefore only includes slices of the rectum that exist in both contours. This will be referred to as clinical-software rectum overlap for the remainder of this report. Finding the overlap was important since this study focused on the geometric uncertainties between the clinical and software delineations and therefore both need to exist for the same amount of slices for the analysis to yield results. Figure 4 shows a patient for which the clinical and software rectum delineations were of different lengths.
It was decided that the centre of mass of the prostate would correspond to the origin of the craniocaudal axis. The main way of deciding the length of the rectum was by introducing a limit from the centre of the prostate in which the rectum would be allowed to exist. Any slices of the rectum contour that would fall outside this limit would be cropped to ensure consistency in the analysis for larger and smaller rectums.

4.3.1 Prostate centre identification

The software prostate delineation was used to find the centre of mass of the prostate in the craniocaudal axis. Again, the reason the software prostate delineation was used was because of its consistency in naming. However, since the software contours were produced by the deep learning algorithm ADMIRE, there existed patients for which the software prostate delineation was not anatomically in the correct place. This usually happened in patients whose field of view of the CT scan was longer than what the software was trained on, resulting in the heart being misidentified as the prostate.

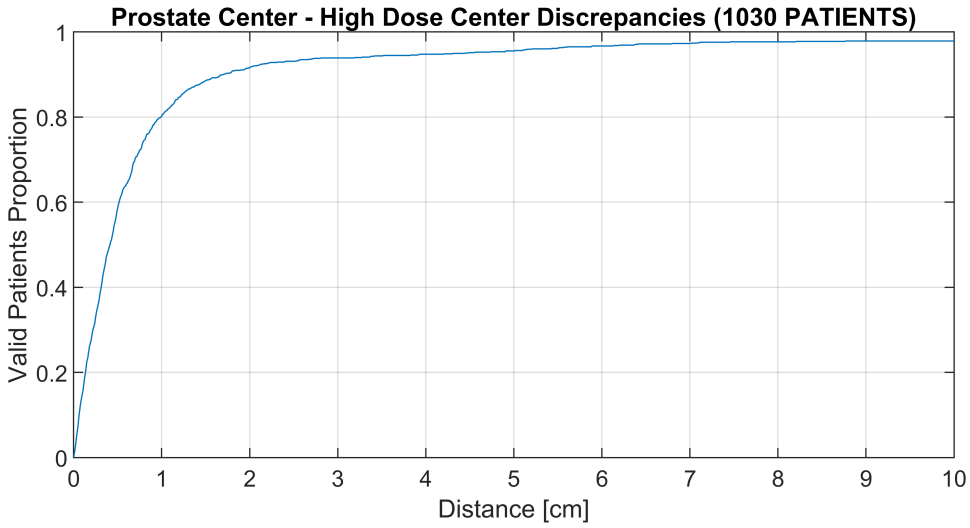

Figure 5: Plot of the proportion of patients (y-axis) against the distance between the high-dose point and the prostate centre in cm. A distance of 2 cm between the prostate centre and the high dose point discards less than 10% of patients.

Use of the data of the radiation dose allowed for the elimination of this error. While the software prostate delineation was not always correct, it was always the case that the highest dose point was located close to the centre of the prostate. A "burn" of the dose was created in the areas where the dose was 95% of its maximum value. PTV sometimes contains part of the seminal vesicles, which is why 95% of the maximum was considered more appropriate than using the maximum dose. The centre of the "burned" dose contour was considered a good approximation for the location of the centre of the prostate. This allowed for the validation of the centre of mass value found from the software prostate delineation.
As shown in Figure 5, a discrepancy of 2 cm between the centre of software prostate delineation and the centre of the high-dose contour allowed 943 patients to proceed to the next step in the analysis. In the cases where the two values were within 2 cm from each other, it was therefore assumed that the software prostate delineation was in the correct place and thus the centre of mass value of this delineation was kept. The patients for which the two values were not close to each other were discarded for consistency purposes.
It is important to mention that although the value of the centre of the high-dose contour was always located inside the prostate, it could not be used on its own as an approximation of the prostate centre of mass value. Different treatment plans would decide the dose distribution in different ways, which would make the method for deciding the length inconsistent between patients.

4.3.2 Length limit from prostate

In order to decide on the optimal length of the rectum relative to the centre of the prostate, statistical analysis was carried out.
The distance in centimetres in the craniocaudal axis was calculated from the top of the clinical-software rectum overlap to the centre of mass of the prostate. The same calculation was done for the bottom of the overlap to the centre of mass of the prostate. As shown in Figure 6 below, the

number of patients against those distances was plotted. This plot allowed for optimisation of the choice in the rectum length, allowing for more patients to be analysed.

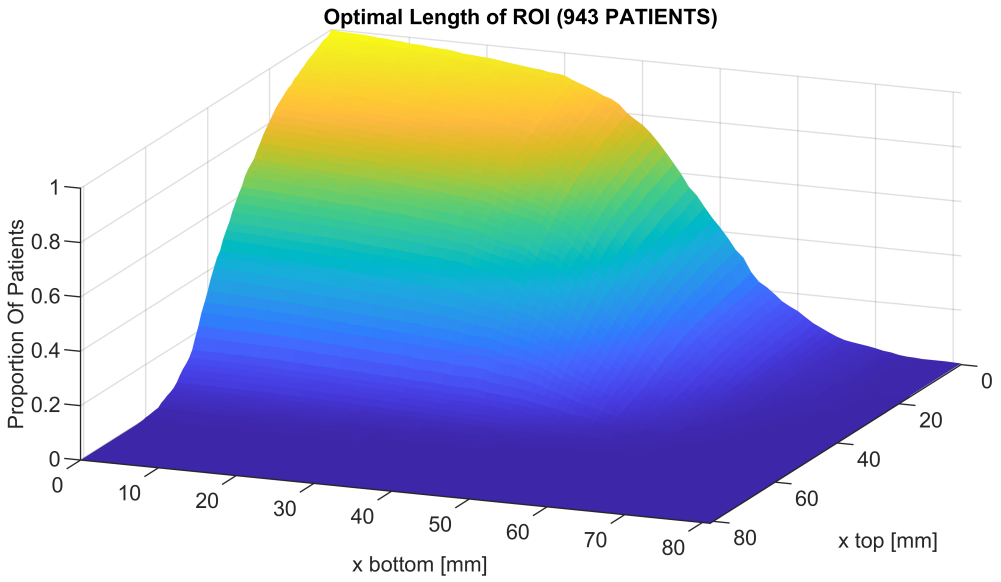

Figure 6: Surface plot showing the proportion out of 943 patients (z-axis) against the distances from the rectum overlap (ROI) to the centre of the prostate. The x and y axes represent the distance from the top and bottom of the rectum overlap to the centre of the prostate, x_{bottom} and x_{top} respectively.

A suitable limit that was suggested from this plot was 3 cm in the craniocaudal axis from the bottom of the rectum and 3 cm from the top of the rectum to the centre of the prostate. The total length of the rectum was thus 6 cm. This allowed keeping 492 patients for analysis. The length from the top and bottom to the centre of the prostate was chosen to be symmetric based on the assumption that dose delivery is symmetric.

4.4 Cylindrical coordinate transformation

The rectum was modelled as a perfect cylinder to simplify analysis and data mining. Converting the rectum into a perfect cylinder allowed moving to surface-based analysis. This was important since the data mining procedure, as shown in Section 4.5, required 2D surface maps in the form of XDR files.
In the method followed, the boundaries of the delineation of the clinical rectum were converted into cylindrical coordinates. This was carried out by the following process:

1. Firstly, the CT scans were cropped so that they only contained the clinical rectum with the pre-decided length. The cropped scans were then separated into individual slices. It was ensured that the slice length was consistent among patients by setting each voxel side to a length of 0.1 cm.

2. For each slice (which has a set y coordinate), the centre $\{x_{\text{centre}}, z_{\text{centre}}\}$ coordinates were found. Sampling over 100 angle steps from those centre coordinates, the coordinates of different points of the boundary were calculated. This was done by converting the centre coordinates $\{x_{\text{centre}}, z_{\text{centre}}\}$ into pixel [2] coordinates and taking steps of 1 pixel until crossing the boundary of the delineation. The distance r between the centre coordinates and the point in the boundary was thus calculated, as well as the quasi $\{x_{\text{bound}}, z_{\text{bound}}\}$ coordinates of that point.

[2]Switching notation to pixels which is the 2D equivalent of a voxel.

3. When the positions of the intersection points were found for different angle samples, the next step was to calculate the distance from those points to the software rectum delineation. As described earlier, this can be done using a distance map of the software rectum delineation. Using the distance map, the distances in cm from those points to the software rectum delineation were obtained and were used to show the discrepancy between the clinical and software contour delineations.

4. Those distance discrepancies were saved for analysis in the form of matrices, where the i-th component corresponded to the slice number ($i = \{0, 1, ..., 60\}$) and the j-th component corresponded to the angle ($j = \{0, 1, ..., 100\}$). By creating those matrices it was automatically assumed that r was the same everywhere. Those matrices were visualised in the form of XDR files and their pixels stored information about the distance discrepancies.

In Figure 7 below, the approximations made to convert the rectum coordinates into a 2D surface map are shown in more detail. By finding the coordinates of the boundaries of the clinical rectum delineation, we could assume that all slices had a common $\{x_{\text{centre}}, z_{\text{centre}}\}$ and thus only differed in the slice number. Furthermore, by assuming the distance r was the same for all slices and angles, the rectum could thus become a smooth cylinder which was then converted to a 2D surface map. The assumption that the rectum was circular could have introduced some errors in the sampling, as will be discussed in Section 6.3.3. In Figure 7, the dose was used in the surface map since it was easier to visualise the location of the prostate relative to the rectum. The surface maps used in the analysis stored information about the distance discrepancies instead of the dose.

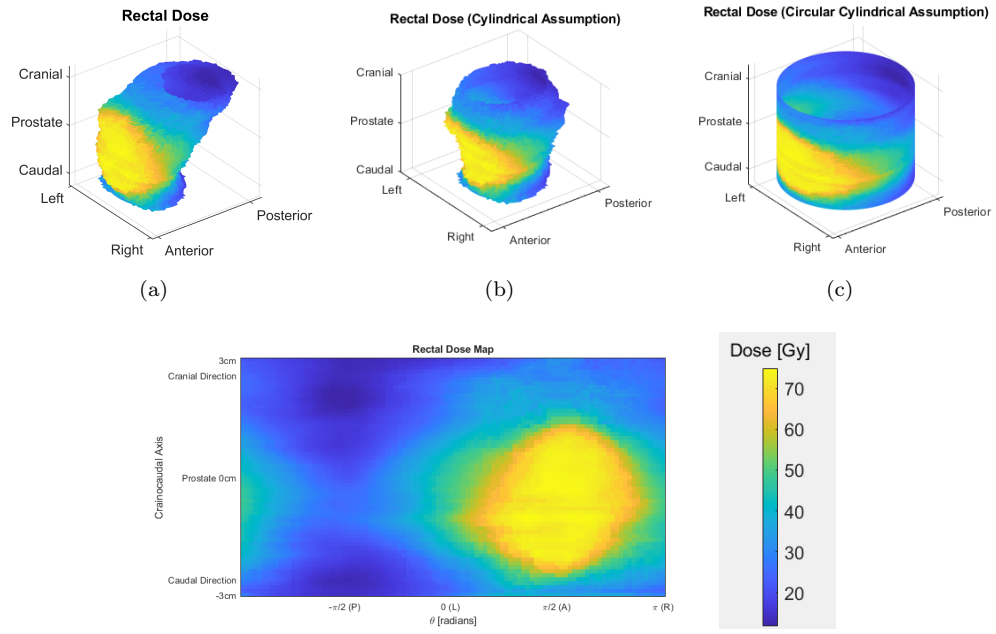

Figure 7: Plot (a) shows the raw data of the rectum. Plot (b) shows the cropped rectum after the assumption was made that all slices were concentric in the x-z plane. Plot (c) shows the cropped rectum of (b) with the additional assumption that r was the same for all slices. Plot (d) is (b) unfolded which allowed converting the data into a 2D surface map. Note that in the actual analysis, the 2D map held information about the distance discrepancies and not the dose. The dose was used in those plots to aid visualisation of the anatomy since it corresponded approximately to the location of the prostate.

4.5 Data Mining

The analysis was performed using an in-house developed package called Manchester Toolkit (MTK). This package contained multiple functions and tests such as Student's t permutation test.

4.5.1 Transformation of data into binary form

The toxicity levels per symptom per patient before treatment (recorded in month 0) were used as a baseline value. In the cases where a baseline toxicity value was not recorded, it was assumed to be 0. In addition, the highest toxicity value per symptom per patient was found. To ensure enough time had passed since treatment started, the highest toxicity values were found by considering data recorded at least three months after the start of treatment.
The data needed to be in binary form for binary analysis to be performed, however, the CTCAE scale ranges from 0 to 5. In order to transform the data into binary form, the difference between the highest toxicity value and the baseline toxicity value was calculated per symptom per patient. For the cases where this value was smaller than 2 then the toxicity was set to 0. When the value exceeded 2 then the toxicity was set to 1. The value of 2 on the CTCAE scale was chosen to dichotomize the data due to the fact that toxicity above the value of 2 requires medical intervention as discussed in Section 2.5.1.

4.5.2 Student's t permutation test

The following method was inspired by the method suggested in [38]. After the toxicity data was made into binary form, the sample of patients had two outcomes: non-event (N) and event (E). Non-event corresponded to the group of patients whose recorded toxicity value for a specific symptom was 0 and event corresponded to the patients whose recorded value was 1. A null hypothesis was formed, which assumed the geometric uncertainties between the clinical and software delineations did not play a role in the increased toxicity levels post-treatment. Because of the null hypothesis, there was no difference between patients with recorded toxicity of 1 for a symptom and those with recorded toxicity of 0. Thus, a new randomised sample was created with i permutations.
For each sample set i and pixel k, the test statistic was given by

$$T_{i,k} = \frac{\mu_{\mathrm{E},i,k} - \mu_{\mathrm{N},i,k}}{\sigma_k}. \tag{6}$$

In Equation 6, $\mu_{\mathrm{E},i,k}$ corresponds to the mean of the distance discrepancy values for group E for a given sample i and pixel k. Equivalently, $\mu_{\mathrm{N},i,k}$ corresponds to the mean of the distance discrepancy values for group E for a given sample i and pixel k. σ_k is the standard deviation of $\mu_{\mathrm{E},i,k} - \mu_{\mathrm{N},i,k}$ over all permutations for a given pixel k. Equivalently, we can define $\tilde{T}_k$ as the test statistic for the non-permuted data, and it is independent of the permutations, i.
Because the number of patients analysed was large, the number of permutations chosen was 10,000. It is suggested in [38] that for large datasets, even 1,000 permutations are sufficient. The permutation test contains the following steps: For every permuted sample i, a $\mu_{\mathrm{E},i,k} - \mu_{\mathrm{N},i,k}$ was computed for every pixel k. After 10,000 permutations, the standard deviation σ_k was computed which allowed for the calculation of the test statistic. The maximal test statistic for a permutation was defined as $T_{max,i}$ and it represented the most significant test statistic for a permutation. Equivalently, $\tilde{T}_{max}$ is the maximum test statistic of the non-permuted data. The p-value is then defined as the fraction of permutations below $\tilde{T}_{max}$.
The Student's t-test was performed with $\alpha = 0.05$. Therefore, the 5th and 95th percentile of the $T_{max,i}$ distribution for a sample i gives a threshold value $T^*_{min,i}$ and $T^*_{max,i}$ respectively. The MTK analysis performed was an extreme pixel analysis and therefore the maximum and minimum of all $\tilde{T}_k$ were outputted as well as the distribution of $T^*_{min,i}$ and $T^*_{max,i}$. By comparing the extreme $\tilde{T}_k$ values to the threshold distribution, the confidence level of statistical significance was found.

5 Results

5.1 Statistical results

An example of how the data was stored for analysis is shown in Figure 8.
The pixel values of the XDR image store information about the distance discrepancies between the clinical and software contour delineations. Darker values in the XDR image show decreasing discrepancies.
Negative distance differences were allowed to account for the cases where the clinical rectum delineation was inside the software rectum delineation. As a result, very dark areas in the XDR image

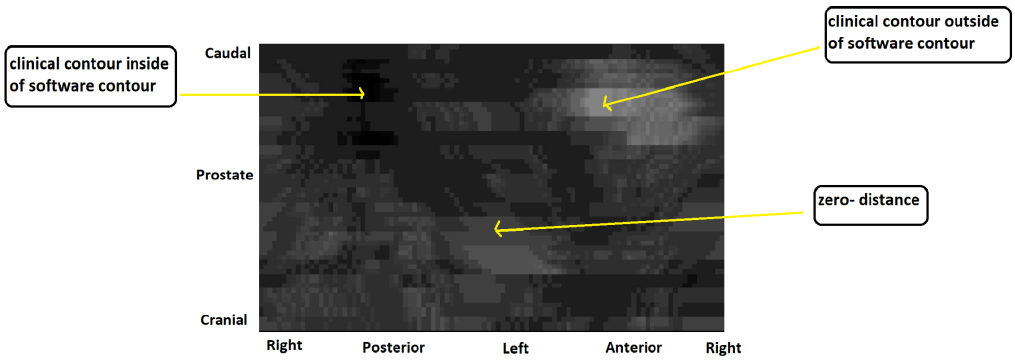

Figure 8: XDR image of the distance differences between the clinical and software rectum delineations. The XDR file has a height of 60 pixels corresponding to the number of slices and a width of 100 pixels corresponding to the angles iterated per slice. Darker values correspond to decreasing distance. Different examples of the distance between the software and clinical delineations are shown.

show areas with negative distance differences.
The MTK Student's t-test was performed for a variety of clinical and patient-reported symptoms such as rectal bleeding, proctitis etc. Emphasis was given on symptoms that are directly related to the rectum and its functions. For a two-tailed Student's t-test with a significance level, α, of 5%, the null hypothesis could not be rejected; thus, no significance was found in the results. Therefore, the analysis performed could not link the geometric uncertainties between the clinical and software rectum delineations to rectal toxicity levels post-treatment.

5.1.1 Distance distribution and distribution of means

Figure 9 shows a surface map in which every point corresponds to the average distance at that point for all patients. There are two clusters in the map where the average distance is not 0.1 cm. Comparing Figure 9 with Figure 7, we can see that the parts where the average distance is highest correspond to the point of the rectum where the rectal wall meets the posterior of the prostate. Therefore, the points in the rectum where the software contours and clinical contours disagree the most, also tend to be the parts of the rectum that are the most irradiated. This could be clinically relevant if the statistical results were significant.
The average of all distance discrepancies for all patients was found to be $\mu = 0.11$ cm. Since the distance discrepancies should not follow a specific trend, we would expect a homogeneous mean of 0.0 cm. This value of 0.11 cm is likely because of an error in the method followed and it will be discussed further in Section 6.3.1.

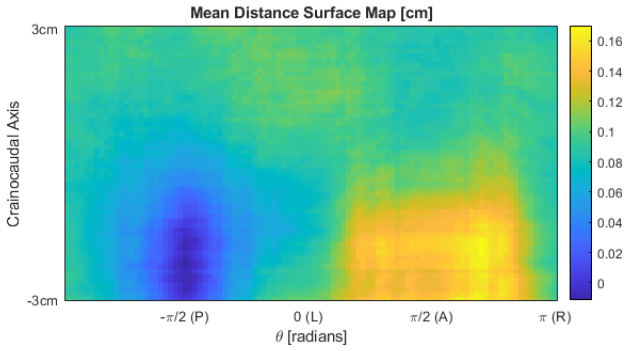

Figure 9: Surface map showing the mean of all distance maps. Each pixel on this map corresponds to the average distance discrepancy between the clinical and software delineations at that point. The average distance is 0.1 cm almost everywhere apart from two clusters at the posterior and anterior of the rectum at the caudal end.

5.1.2 Range of distances and outliers

The maximum distance observed between the software rectum contour and the clinical rectum contour was 3.06 cm. The minimum distance observed was -1.92 cm. Those extreme distance values come from the same patient showing that machine learning-generated contours can produce contours with very high discrepancies to the clinical contours. It was found that 32 patients existed whose software rectum contour and clinical rectum contour deviated more than 1 cm. Another Student's t permutation test was performed after removing those potential outliers to test their effect on the results. No significance was found in the reduced data set of 460 patients.

6 Discussion

6.1 Limitations and Biases

The results produced in this study were only preliminary and were limited by the methods used as well as other parameters.
The initial data sample consisted of 1,758 patients out of which, only 492 were used in the analysis. The steps where several patients were eliminated were as follows:

1. 728 patients were eliminated in the process of combining the CT scans with the clinical and software delineations. The main reason for elimination was due to the naming of the files, which caused issues in differentiating between the delineation, dose, and CT scan files. This led to the potential exclusion of certain countries and hospitals from the dataset. More patients could be kept at this stage if those files were differentiated based on the file size rather than the file name.

2. In the next stage in the process, which was the identification of the clinical rectum, no patients were eliminated. However, as mentioned in Section 4.2.1, the clinical rectum was misidentified in some cases. The percentage of misidentified rectums was 5% out of 1,030 patients. However, after the length limit for the rectum was imposed, there were no patients with misidentified rectums in the final dataset that was analysed. More sophisticated methods other than distance mapping could be introduced to ensure all clinical rectums are correctly identified.

3. Applying a 2 cm limit between the software prostate delineation and the high dose contour eliminated a further 8% out of the 1,030 patients left at this stage.

4. The final stage of the elimination process was choosing a suitable length for the rectum to apply to all patients. Applying the same length to all the patients led to more consistency and it also allowed the XDR files to be of the same size for the data mining process. However, bias was introduced since smaller rectum contours were not included in the analysis. The length of the rectum was chosen with the prostate as the centre, which is where the highest dose in the rectum is located. Accepting a smaller rectum length would significantly increase the number of patients analysed, however, it could potentially eliminate areas in the rectum that are radiated with a high dose. As shown in Figure 7, the part of the rectum below the prostate (caudal end) was generally irradiated more than the part above the prostate (cranial end). To maximise the number of patients analysed without cropping high-dose areas, the limit from the centre of the prostate to the top of the rectum can be chosen to be $0.5 - 1$ cm smaller.

6.2 Variations in rectal volumes

Another limitation of this study was the variation in rectal volumes as related to CTV design in treatment planning. The rectal wall location can vary significantly for patients depending on how full the rectum was when the CT scan was taken compared to when the radiation was delivered [39]. Therefore, the initial contouring can be different from the actual rectum position during treatment. There are various studies which show that changes to the posterior/anterior axis have a mean of 1 mm [40] and others show that the movement of the rectum and prostate in posterior/anterior axis can be more than 5 mm [41]. This shows that the resolution of 1 mm chosen might be oversampling noise without providing any clinically relevant results.

6.3 Errors

6.3.1 Mean of distances

As shown in Figure 9 we observe the mean of the distance discrepancies to be approximately 0.1 cm. Since the resolution for the y-axis was chosen to be 0.1 cm, this suggests that the clinical contours lie 1 pixel outside the software rectum contours on average.
There is a possibility that the initial assumption that the software contours are correct is not valid. The resolution to which ADMIRE has been trained is unknown and therefore the resolution used in this project might be too high, causing the software delineation to round to the nearest pixel. This has the potential of introducing an error of 0.1 cm, which is the size of one pixel.
In order to test this, we followed the same method in Section 4, to compare the distance discrepancies of the clinical rectum to itself. We would expect the distance map to give values of 0 everywhere when the clinical rectum delineation was compared to itself instead of the software rectum delineation. When this simple method test was performed, the distance map returned non-zero values. Those values were either 0.1 cm or $\sqrt{2} \times 0.1$ cm $= 0.141$ cm. The error of 0.141 cm has a physical meaning, since it appeared at the points most diagonal to the centre, so it corresponds to the diagonal of a square of size 1 pixel. This consistent error suggests that there is an error in the distance mapping itself. This is demonstrated more clearly in Figure 10. It was decided that due to time constraints, those preliminary results including this error should be presented in this report.

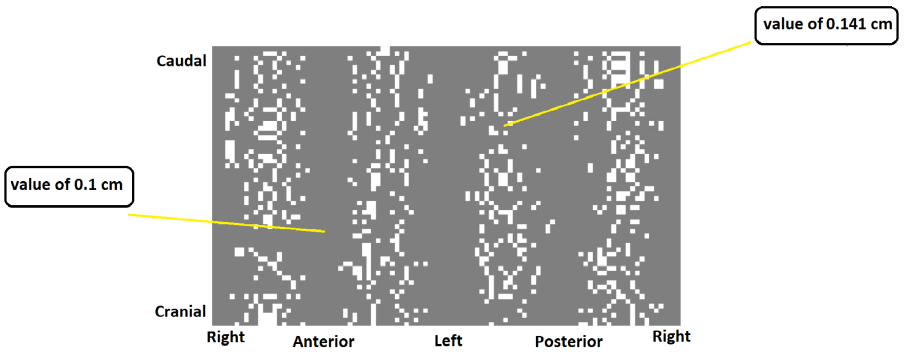

Figure 10: XDR file showing the distance discrepancies in the clinical rectum contour when compared to itself following the method suggested in this report. The most diagonal points from the centre of each slice have a value of $\sqrt{2} \times 0.1$ cm (white) the rest have a value of 0.1 cm (grey), demonstrating that the error is of size 1 pixel.

6.3.2 Resolution

As mentioned in Section 4.4, the resolution was set to a voxel size length of 0.1 cm for consistency. It was concluded that increasing the resolution was not necessary since changes of less than 0.1 cm cannot massively alter the contouring procedure and thus affect the toxicity outcomes. Furthermore, increasing the resolution allows for the sampling of more noise.

6.3.3 Geometry-based errors

There are a number of possible errors arising from the geometric assumptions throughout the project that could lead to the lack of significance of the results.
As mentioned, the assumption that the rectum is circular can cause the distance sampling to lose its physical meaning. There were cases in which parts of a slice of the rectum were massively distorted/folded and therefore the angular separation was unable to capture the true structure of the rectum. A potential way to improve this would be to choose a different angular step. The choice of 100 angles was made arbitrarily and it could potentially lead to the sampling of more noise while, at the same time, not showing the structure of the rectum.

7 Future suggestions

In this section, a few possible future directions for this research project will be discussed.

7.1 Dataset expansion

As discussed previously, the fraction of patients from the original dataset that was eventually analysed was 28%. Increasing the number of patients is important to draw more meaningful and complete conclusions, and it has the potential of altering the statistical significance of the analysis. There are several ways in which the number of patients analysed can be increased, such as improving the method for reading the original data files and combining the software and clinical contours. Another way of ensuring more patients are analysed is by adjusting the accepted length of the rectum. In this project, a length limit of 3 cm was allowed from the centre of the prostate. Choosing a smaller limit would include patients whose rectum delineation existed for fewer slices.

7.2 Advanced analysis techniques

Another possible direction for this project would be to implement machine learning and deep learning algorithms to avoid simplistic analysis of the data and yield more meaningful results. For instance, decision tree classifiers are a popular machine learning model that can be used as a predictive model to draw conclusions about the data [42]. More advanced statistical processes such as cluster-based permutation tests should also be considered [38].

7.3 Extension to other organs

Since the geometric uncertainties between the software and clinical rectum delineations did not yield any statistical significance when compared to the toxicity levels, an interesting extension to this research project would be to perform a similar analysis on other organs surrounding the prostate. In the set of patients used in this study, clinically reported toxicities are higher in bladder-related symptoms, which makes the bladder an interesting candidate for this extension.

7.4 Improving geometric errors

Some errors in the analysis potentially rise from the circular cylindrical assumptions made to unfold the rectum to a 2D plane. Those errors could possibly be improved if the analysis was carried out in 3D space rather than 2D.
Another potential improvement would be to include the dose to remove spatial dependence. The dose contains more information about how the rectum was irradiated and thus can be linked to the toxicity levels post-treatment. On the other hand, the distance discrepancies between the software and clinical rectum contours are only relative and do not provide any information on the radiation received at a certain point in space. The distance map can vary for patients that have received the exact same dose at all points of the rectum. There are multiple studies that use dose-volume histograms to deduce conclusions about late toxicity levels for the rectum, specifically for complications such as rectal bleeding [43], [44].

8 Conclusion

The analysis carried out yielded that the minimum and maximum values of distance discrepancies between the clinical and software contour delineations were -1.92 cm and 3.06 cm respectively. The Student's t permutation test found no significance when analysing a sample of 492 patients. The same test was performed after removing 32 potential outliers and again, no significance was found. Therefore, the discrepancies in rectal contouring cannot be linked to increased rectal toxicity levels post-treatment. The mean of all distances found for all patients was 0.1 cm, which suggests an error in the distance mapping procedure and could be improved by altering the resolution used. The percentage of patients analysed from the original sample was 28% after multiple filters to ensure validity of the results. This suggests that the method followed can be improved in the future to maximise the percentage of patients analysed. More sophisticated analysis models that

are more suitable to the type of data used can also be implemented to improve the data mining procedure.

References

[1] P. RAWLA, "Epidemiology of prostate cancer," *WORLD J ONCOL*, vol. 10, no. 2, pp. 63–89, 2019.

[2] G. Almeida and J. M. R. Tavares, "Deep learning in radiation oncology treatment planning for prostate cancer: A systematic review," *Journal of medical systems*, vol. 44, no. 10, pp. 1–15, 2020. DOI: 10.1007/s10916-020-01641-3.

[3] P. Hoskin, *External Beam Therapy*. Oxford University Press, 2019, p. 1.

[4] K. H. Hammerich, G. E. Ayala, and T. M. Wheeler, "Anatomy of the prostate gland and surgical pathology of prostate cancer," Contemporary Issues in Cancer Imaging, pp. 1–10, 2009. DOI: 10.1017/CBO9780511551994.003.

[5] F. V. Coakley and H. Hricak, "Radiologic anatomy of the prostate gland: A clinical approach," pp. 15–30, 2000. DOI: https://doi.org/10.1016/S0033-8389(05)70147-0.

[6] J. E. McNeal, "The zonal anatomy of the prostate," pp. 35–49, 1981. DOI: https://doi.org/10.1002/pros.2990020105.

[7] R. Heald and B. Moran, "Embryology and anatomy of the rectum," in *Seminars in surgical oncology*, Wiley Online Library, vol. 15, 1998, pp. 66–71.

[8] P. T. Scardino, R. Weaver, and A. H. M'Liss, "Early detection of prostate cancer," *Human pathology*, vol. 23, no. 3, pp. 211–222, 1992.

[9] C. W. Hudson MA Bahnson RR, "Clinical use of prostate specific antigen in patients with prostate cancer.," pp. 1011–1017, 1989. DOI: 10.1016/s0022-5347(17)38972-3.

[10] J. H., "Progress in medical physics," pp. 1–17, 2021. DOI: https://doi.org/10.14316/pmp.2021.32.1.1.

[11] G. Gilmore, "Practical gamma-ray spectrometry: Second edition," pp. 25–43, 2008.

[12] N. R. C. (US), I. of Medicine (US) Committee on the Mathematics, and P. of Emerging Dynamic Biomedical Imaging., *Mathematics and Physics of Emerging Biomedical Imaging*. National Academies Press (US), 1996, Chapter 3, X–Ray Computed Tomography.

[13] E. Seeram, *Computed Tomography-E-Book: Physical Principles, Clinical Applications, and Quality Control*. Elsevier Health Sciences, 2015.

[14] R. A. Brooks and G. Di Chiro, "Theory of image reconstruction in computed tomography," *Radiology*, vol. 117, no. 3, pp. 561–572, 1975. DOI: https://doi.org/10.1148/117.3.561.

[15] A. Shukla-Dave and H. Hricak, "Role of mri in prostate cancer detection," *NMR in Biomedicine*, vol. 27, no. 1, pp. 16–24, 2014.

[16] R. Li, G. C. Ravizzini, M. A. Gorin, *et al.*, "The use of PET/CT in prostate cancer," *Prostate Cancer and Prostatic Diseases*, vol. 21, no. 1, pp. 4–21, 2018.

[17] M. S. Soloway, S. W. Hardeman, D. Hickey, *et al.*, "Stratification of patients with metastatic prostate cancer based on extent of disease on initial bone scan," *Cancer*, vol. 61, no. 1, pp. 195–202, 1988.

[18] A.-L. Grosu, L. D. Sprague, and M. Molls, "Definition of target volume and organs at risk. biological target volume," in Springer Berlin Heidelberg, 2006, pp. 167–177. DOI: 10.1007/3-540-29999-8_13.

[19] R. Krol, R. J. Smeenk, E. N. van Lin, E. E. Yeoh, and W. P. Hopman, "Systematic review: Anal and rectal changes after radiotherapy for prostate cancer," pp. 273–283, 2014.

[20] J. W. H. Leer, "What the clinician wants to know: Radiation oncology perspective," *Cancer Imaging*, S1–2, 2005.

[21] G. Li, Y. Li, J. Wang, *et al.*, "Guidelines for radiotherapy of prostate cancer (2020 edition)," *Precision Radiation Oncology*, vol. 5, no. 3, pp. 160–182, 2021. DOI: https://doi.org/10.1002/pro6.1129.

[22] C. Rasch, I. Barillot, P. Remeijer, A. Touw, M. van Herk, and J. V. Lebesque, "Definition of the prostate in ct and mri: A multi-observer study," *International Journal of Radiation Oncology* Biology* Physics*, vol. 43, no. 1, pp. 57–66, 1999.

[23] X. Gual-Arnau, M. Ibáñez-Gual, F. Lliso, and S. Roldán, "Organ contouring for prostate cancer: Interobserver and internal organ motion variability," *Computerized Medical Imaging and Graphics*, vol. 29, no. 8, pp. 639–647, 2005. DOI: https://doi.org/10.1016/j.compmedimag.2005.06.002.

[24] M. Debois, R. Oyen, F. Maes, *et al.*, "The contribution of magnetic resonance imaging to the three-dimensional treatment planning of localized prostate cancer," *International Journal of Radiation Oncology*Biology*Physics*, vol. 45, no. 4, pp. 857–865, 1999. DOI: https://doi.org/10.1016/S0360-3016(99)00288-6.

[25] C. M. Washington and D. T. Leaver, *Principles and Practice of Radiation Therapy*. Elsevier Health Sciences, 2015, p. 1.

[26] H. Suit, S. Goldberg, A. Niemierko, *et al.*, "Proton beams to replace photon beams in radical dose treatments," *Acta Oncologica*, vol. 42, no. 8, pp. 800–808, 2003.

[27] N. C. Institute, *Common terminology criteria for adverse events (ctcae)*, 2017. [Online]. Available: https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/ctcae_v5_quick_reference_8.5x11.pdf.

[28] R. Llarena Ibarguren, J. A. Zabala Egurrola, A. Arruza Echevarrıa, *et al.*, "Rectal involvement secondary to prostatic adenocarcinoma," *Arch Esp Urol*, vol. 46, no. 2, pp. 112–115, 1993.

[29] M. S. Haq, P. M. Thomas, M. Almonte, and V. Mohan, "Prostate cancer presenting with an unusual presentation of rectal pain and bleeding," *Cureus*, vol. 13, no. 10, e18766, 2021.

[30] P. D. Johnston BC, "Cochrane handbook for systematic reviews of interventions," 2022.

[31] R. Cisneros, "Impulsive pedipulation: Using the humanoid's leg for impulsive "manipulation" of objects," Ph.D. dissertation, Feb. 2015.

[32] P. Seibold, A. Webb, M. E. Aguado-Barrera, *et al.*, "REQUITE: A prospective multicentre cohort study of patients undergoing radiotherapy for breast, lung or prostate cancer," *Radiother Oncol*, vol. 138, pp. 59–67, 2019.

[33] C.-C. Teng, L. Shapiro, and I. Kalet, "Head and neck lymph node region delineation with image registration," pp. 12–13, 2010. DOI: 10.1186/1475-925X-9-30.

[34] R. L. Mason, R. F. Gunst, and J. L. Hess, *Statistical design and analysis of experiments: with applications to engineering and science*. John Wiley & Sons, 2003, pp. 53–57.

[35] P. Good, *Permutation tests: a practical guide to resampling methods for testing hypotheses*. Springer Science & Business Media, 2013, pp. 6–8.

[36] S. G. Kwak and J. H. Kim, "Central limit theorem: The cornerstone of modern statistics," *Korean journal of anesthesiology*, vol. 70, no. 2, pp. 144–156, 2017.

[37] S. M. Pereira and G. Leslie, "Hypothesis testing," *Australian Critical Care*, vol. 22, no. 4, pp. 187–191, 2009. DOI: https://doi.org/10.1016/j.aucc.2009.08.003.

[38] C. Chen, M. Witte, W. Heemsbergen, and M. v. Herk, "Multiple comparisons permutation test for image based data mining in radiotherapy," *Radiation Oncology*, vol. 8, no. 1, p. 293, 2013. DOI: 10.1186/1748-717X-8-293.

[39] M. E. Daly, J. D. Murphy, E. Mok, C. Christman-Skieller, A. C. Koong, and D. T. Chang, "Rectal and bladder deformation and displacement during preoperative radiotherapy for rectal cancer: Are current margin guidelines adequate for conformal therapy?" *Practical Radiation Oncology*, vol. 1, no. 2, pp. 85–94, 2011, ISSN: 1879-8500. DOI: https://doi.org/10.1016/j.prro.2010.11.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1879850010000184.

[40] H. Yamashita, R. Takenaka, A. Sakumi, A. Haga, K. Otomo, and K. Nakagawa, "Analysis of motion of the rectum during preoperative intensity modulated radiation therapy for rectal cancer using cone-beam computed tomography," *Radiat Oncol*, vol. 10, p. 2, Jan. 2015.

[41] A. R. Padhani, V. S. Khoo, J. Suckling, J. E. Husband, M. O. Leach, and D. P. Dearnaley, "Evaluating the effect of rectal distension and rectal movement on prostate gland position using cine MRI," *Int J Radiat Oncol Biol Phys*, vol. 44, no. 3, pp. 525–533, Jun. 1999.

[42] B. Zupan, J. Demšar, M. W. Kattan, J. R. Beck, and I. Bratko, "Machine learning for survival analysis: A case study on recurrence of prostate cancer," *Artificial intelligence in medicine*, vol. 20, no. 1, pp. 59–75, 2000.

[43] M. Someya, M. Hori, K. Tateoka, *et al.*, "Results and DVH analysis of late rectal bleeding in patients treated with 3D-CRT or IMRT for localized prostate cancer," *Journal of Radiation Research*, vol. 56, no. 1, pp. 122–127, 2014. DOI: 10.1093/jrr/rru080.

[44] S. L. Tucker, L. Dong, R. Cheung, *et al.*, "Comparison of rectal dose–wall histogram versus dose–volume histogram for modeling the incidence of late rectal bleeding after radiotherapy," *International Journal of Radiation Oncology*Biology*Physics*, vol. 60, no. 5, pp. 1589–1601, 2004. DOI: https://doi.org/10.1016/j.ijrobp.2004.07.712.